Dubbing non-human voices

Measuring the vocal characteristics of monsters

Dillen Louis Smit

Supervisor: Dr. Mechtild Tronnier

Centre for Language and Literature, Lund University
MA in Language and Linguistics, Phonetics
SPVR01 Language and Linguistics: Degree Project – Master's (Two Years) Thesis, 30 credits

Spring term 2023

Abstract

What do monsters sound like? The monsters or otherwise non-human characters in film, animation, and video games will often have vocal traits that signify or accentuate their physical properties, distinguishing them for humans even though they speak human language. Of course, the voice actors dubbing these voices are human, and have to find ways to alter their voice quality in order to represent the monster in question. Voice quality, is described as a phonetic descriptor of accent (Esling et al., 2019) and comprises a combination of different components, including phonation type, laryngeal and supralaryngeal aspects. These aspects were analysed using pitch measurements, vowel formants, and measurements relating to aperiodicity and noise. Four participants were recorded in order to examine the ways and extend to which they would alter their natural voice when dubbing non-human characters. The acoustic measurements were also related to character descriptions provided by the participants.

For the characters with deeper voices the participants all went below the minimum pitch of their modal voice, showing that creaky and harsh voice types can extend the vocal range below the range of modal phonation. For the higher pitch characters the speakers raised their pitch to a relatively higher degree, but did so without approaching their upper limit. The results indicate there is a tendency to increase the pitch range for character voices, although there was one exception where the character voice was more monotone relative to the normal voice. In terms of phonation type, different forms of creaky, harsh, tense, and falsetto were used. One participant also employed ingressive phonation, a relatively ineffective way to produce vocal fold vibration. It was found that vowel systems can shift as a whole, becoming more fronted or lowered, but can also be compressed and expanded individually along either the vertical or horizontal axis. The concept of monster is very culturally dependent, this study was therefore consciously restricted to Swedish participants. A future study could involve voice actors with different cultural and linguistic backgrounds to study the different vocal effects they might use to phonetically manifest their monsters.

Keywords: voice, vocal characteristics, voice quality, pitch, vowel quality, phonation type, dubbing, voice acting

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Mechtild Tronnier for her support and encouragement.

My gratitude also goes out to Prof. Jordan Zlatev & Åsa Wikström for their help and guidance through the master programme in Linguistics at Lund University. I would also like to thank Dr. Shinichiro Ishihara & Dr. Arthur Holmer for their inspiration and advice.

A special thanks to the participants and to the Humanities Lab at Lund University for providing a suitable recording environment.

Thanks to my family for always supporting me, and for inviting me into their home to let me work on the thesis without any distractions.

Thanks to my fellow students and friends, Jule Nabrotzky, Laura Timm, Daniel Johansson, Erika Sombeck, Marcus Holmström, Henry Nicholson, Josef Ekelund, Veronica Lattanzi, Tom Tóth, Ho Wan Jeremy Leung, Gabriele Ninivaggi, Juraj Pikuliak, Tadeo Hepperle, Liam Cole, Monika Hegerová.

Table of contents

List of figures

List of tables

Abbreviations

ATR	Advanced Tongue Root
EP	Egressive Phonation
F0	Fundamental frequency
F1	First formant
F2	Second formant
HNR	Harmonics-to-Noise Ratio
IP	Ingressive Phonation
IPA	International Phonetic Alphabet
LAM	Laryngeal Articulator Model

Chapter 1 Introduction

1.1 Voices & dubbing

This study explores the ways and extend of how one's vocal characteristics can be changed. Voice actors can alter their vocal characteristics to impersonate a specific character. Changing the vocal characteristics can emphasise specific character traits. Depending on the character, be it in a movie or a video game, the vocal characteristics can reflect more non-human traits. Perhaps the character is snake-like and to create a fitting vocal performance, the impersonator produces a whispery voice. There are many ways in which one can change their voice. In this study those changes are acoustically measured and analysed.

Dubbing is described as the post-production process of adding sound to film and video, or as the replacing of the original voices with a foreign language. In this paper dubbing will refer to the voicing of characters in film and cartoon. The term ”voicing” can also refer to the (de-)voicing of consonants and is deemed too ambiguous, especially in the field of linguistics, therefore the term dubbing will be used.

Voice is broad concept, with many different aspects which will be further explored, but here it will be introduced with a quick summary. Voice is a culmination of articulatory aspects, your vocal cords vibrate, the vibrating air goes through your larynx and then tongue position and mouth position further add constriction or obstruction to articulate specific consonants and amplify certain frequencies. The rate and amplitude of vibration changes the pitch and loudness of your voice.

In the most narrow sense voice can be described as the vibrations of the vocal folds (Garellek in Katz & Assmann, 2019). In this study a broader definition is used in order to include other vocal tract articulators. If a more specific definition is required it will be referred to accordingly. For example, in case the fundamental frequency needs to be addressed separately, or the phonation type.

To an extend one's physiology is a decisive factor in what the voice sounds like. Longer or thicker vocal cords will vibrate more slowly creating a lower pitch for example. However, by changing the way you use your vocal tract to create speech sounds, you can change your voice

quality. How much do we change our voice quality when dubbing a non-human character is the main topic of this research. In the next section this concept of voice quality will be further discussed.

1.2 Voice & vocal features

One of the key concepts in this study is voice quality, however, how do we define this seemingly straightforward concept? Voice quality can be described as the quasi-permanent characteristics that are present in one's speech more or less all the time that one is talking (Abercrombie, 1967).

Kreiman & Sidtis (2011) describe voice quality as one of the primary means by which speakers project their identity to the world. This identity entails physical features, but also psychological and social characteristics. They also mention the impression listeners gain from someone's voice are not necessarily accurate. Even if you have only ever heard someone's voice without meeting them, you are still likely to form a mental image based on their voice quality. However, once you meet you might discover there is a mismatch between this previously established mental image and this person's actual appearance (Kreiman & Sidtis, 2011).

An important aspect of voice quality is phonation type. Akita (2021) studied the relationship between phonation type and sound symbolism (defined as the direct linkage between sound and meaning). Akita conducted a perception test and found that phonational symbolism has relatively clear acoustic grounds. They argue that phonation types make a considerable contribution to sound-symbolic ratings of size and shape in monolingual Japanese speakers (Akita, 2021).

Apart from phonation type there are other aspects of voice quality. Firstly, pitch, F0, or the rate of vocal cord vibration is what dictates the tonal aspect of voice. Intonation can change the meaning of words and phrases. It is measured in Hz, and there is generally a difference between adult male speakers and female speakers. For example, in this study the male participants had an average pitch of 108 Hz whereas the female participants had an average pitch of 202 Hz (see table 2 for more details).

After the glottis, where the vocal fold vibration serves as the sound source, the sound, including its harmonics are amplified in the different parts of the vocal tract. Laryngeal aspects can affect the signal in different ways, laryngeal constriction for example can lower particularly the F2 and

F3 formants (Teshigawara, 2003). Vowel formants are amplified harmonics that are mostly affected by tongue position and mouth shape, resulting in different vowel phonemes.

A distinction is made between intrinsic and extrinsic features of voice quality. Intrinsic features are the result of the anatomy of the speaker and as such cannot be controlled. Extrinsic features are the results of the way a speaker uses their vocal tract and larynx, and thus can be controlled volitionally (Teshigawara, 2003). In this study we will be looking at the extrinsic features, as we want to see how speakers can volitionally change their vocal features, in other words, the extrinsic features.

With all the different aspects of voice quality there are many parameters that could be looked at within this research. In order to create a self-contained study more specific approach will be utilised. By operationalising the research questions we can establish which parameters are relevant and what data will be necessary to verify the hypotheses.

1.3 Research questions

The main research question is: “In what way do voice actors change their voice when dubbing a non-human character?” This is quite an open question so in this section it will be broken down into smaller, more specific questions that can be operationalised.

1.3.1 The main research questions

> *In what way do voice actors change their voice when dubbing a non-human character?*

Starting from the most general question, we can further specify what we are looking for. The information we are after is not what is the absolute maximum alteration physically possible to the voice, but rather the following:

> *What is the furthest away from their normal voice an impersonator comes up with when asked to dub a non-human character?*

In order to operationalise the research question, we have to specify the parameters on which we are going to base this vocal change.

How does the voice quality change in terms of pitch, phonation type and vowel quality?

This question can consequently be broken down into fully operationalised sub-questions that relate more or less directly to the measurements and results.

1.3.2 The operationalised research questions

1. Pitch & phonation type

    1.1 Compared to the average pitch in one's normal voice (in speech), how much higher or lower is the average pitch for the non-human characters? (change in semitones)

    1.2 Compared to the average pitch variation in one's normal voice, is there smaller or larger pitch variation in the non-human characters?

    1.3 In the non-human characters, is the pitch closer to the normal pitch (humming tone) or to the minimum and maximum pitch of their voice? (compared to min & max F0)

    1.4 Do participants use different phonation types, and can these be measured in terms of jitter, shimmer and HNR?

2. What happens to vowels when dubbing a non-human/monster character that still speaks human language?

    2.1 Are vowels centralised? (do they become less distinct, which could affect comprehensibility)

    2.2 Is the whole vowel system shifted? (is it stretched or are vowels individually dislocated)

    2.3 Do vowel formants shift along with raised-larynx voice, or other forms of laryngeal constriction?

3. Do the different speakers use a similar approach, or is there for example a systematic difference between male and female voice actors when dubbing monsters/aliens/non-human characters?

4. What do these vocal techniques mean for the character?

4.1 What personality traits do participants think of when dubbing their characters? (evil, mean, friendly, old/young, etc.)

4.2 What vocal properties do participants think they are using to portray the personality? (raspy, harsh, high/low pitch, creak, etc) note that terminology vary between researchers and performers.

4.3 What acoustic properties are we actually able to measure? (start with what the researcher hears, and then confirm with acoustic measurements)

4.4 How do the acoustic measurements relate to the properties described by the participants in 4.2?

4.5 To what extend do the personality traits and vocal properties described by the participants match with what the literature suggests? (laryngeal constriction for villains for example)

Answers to these questions will be sought out by analysing specific recorded speech samples produced by the participants who were prompted to read a Swedish sentence in their normal voice and then in two different character voices.

1.4 Approach

Four participants were recorded producing the same sentence, in their normal voice and in two different non-human/monster style character voices of their own choice. They were asked to imagine the character traits for these monsters themselves and give a vocal performance they associated with these traits.

The vocal quality we are trying to define is that of a monster, a non-human being that still communicates in human language. What specific traits this character has is up to the participants. They were asked to define these characteristics and explain their thought process behind it. Whether it would be a scary or mean monster, or perhaps an old and wise monster, one character could be calm while the other sounds aggressive. These characteristics result in vocal traits, like harsh voice, or breathy voice, which in turn were measured in an acoustic analysis.

Although this study is not necessarily linked to a specific language, the participants are all Swedish and so the prompt also included a Swedish sentence. A more detailed description will be given in chapter 3.

I am aware of the additional parameters that could be looked at within this research. But melody, pitch contour and accent, as well as temporal aspects will not be included. Duration for example can be measured relatively easily but does not really provide valuable information on how vocal characteristics can be altered, and therefore are left out.

Chapter 2 Theory & background

In chapter one the concepts of voice quality and phonation type have been briefly discussed. In this chapter we will go more in depth, binding together these concepts along with vowel formants using the Laryngeal Articulator Model or LAM, as well as discussing other relevant research.

Esling et al. (2019) provide the Laryngeal Articulator Model (LAM) as a framework to analyse voice quality. In this model the larynx is described not just as the source of vocal fold vibration, but as an articulator. (Esling et al. 2019, voice and voice quality, p4-6)

As mentioned in chapter one, voice quality can be described as the quasi-permanent characteristics that are present in one's speech more or less all the time that one is talking (Abercrombie, 1967). There are intrinsic and extrinsic features of voice quality. In this study we will be looking at the extrinsic features, but even if we just focus on the extrinsic features, voice quality is still a very broad concept.

2.1 Voice quality & pitch

Voice quality is a combination of different aspects. There are the laryngeal and supralaryngeal aspects, and phonation type. These different components will be discussed and defined further in this chapter.

Esling et al. (2019) describe voice quality as a property of accent or as a phonetic descriptor of accent. It is the long-term quality of one's voice, and as such, it is the longest-term phonetic strand of the aural medium of language. The two other strands are the prosodic and segmental strands, both of which are very short term. Segments last for tens or hundreds of milliseconds, while prosody, or voice dynamics occur over multiple syllables or phrases.

These strands are differentiated, but that does not mean there is no overlap. As mentioned before, voice quality pertains to the quasi-permanent characteristics in speech that are present while someone is speaking, this naturally includes phrases, syllables and segments as well.

At the phrase level, you are dealing with prosodic features, the melody of the sentence. The melody of a phrase can affect the meaning, but at the same time it also pertains to the voice quality. Measuring the pitch of the vocal fold vibration gives you information on both of these aspects.

In this study vowels are also measured, which could arguably be seen as an investigation of the segmental aspect. However, the placement of vowels can also provide valuable information on the setting of the larynx and tongue root. The more we change the setting of our articulators, the more those changes also influence the vowel quality.

There are different ways we can let our vocal folds vibrate. Sometimes phonation is created by only vibrating part of the vocal folds or it can incorporate the false vocal folds to create a very different sound. This can have a big impact on the vocal characteristic. This is why phonation type plays an important role within voice quality. In some literature the two notions are even used interchangeably. In this study phonation type will be referred to as one aspect of voice quality. In the next section the different phonation types will be discussed.

2.2 Phonation types

Phonation in a more narrow sense refers to the vibration produced in the glottis, but as mentioned above phonation can involve more than just the vocal folds. Phonation and phonation type in this study are used to refer to phonation in the broad sense. Laver (1980, 1994) defines several simple and compound phonation types. Simple types occur alone, compound types occur combined. Group 1 comprises modal voice & falsetto (chest & head voice), can occur alone or with other types, but not with each other. Whisper & creak, these can occur both as simple and compound types, and are able to combine with group 1 and with each other. Harshness & breathiness can only occur in compound types of phonation.

Laver also mentions another phonation type as a subgroup of harshness, the ventricular voice (Laver, 1994, pp. 114-118). For the ventricular voice phonation is caused, not just by the vibration of the true vocal folds, but by the combined vibration of the true vocal folds and the ventricular folds.

MODAL (also referred to as chest voice): The laryngeal characteristics of modal voice, or the neutral mode of phonation. Vibration of the larynx is regularly periodic, and an efficient way of producing vibration. There is no audible friction brought on by incomplete closure of the glottis.

FALSETTO (also referred to as head voice or thin register): Sub-glottal air pressure is lower than for modal voice, cricothyroid muscle puts tension on vocal folds making them thin, glottis often remains slightly open. But the opening is small so fricative component is whispery rather than breathy. The fundamental frequency of falsetto is typically considerably higher than the F0 of the modal voice (Laver, 1994). The vocal folds do not fully close, they come together just enough to make the edges vibrate (Esling et al., 2019).

The following three phonation types (whisper, creak & harsh) are the result of aryepiglottic structure rather than primarily glottal. They are the result of laryngeal constriction, the presence and extent of which changes the state of the larynx and generates specific sound qualities (Esling et al., 2019).

WHISPER: Widely agreed to be characterised by a triangular opening of the glottis. This opening lets the air flow through creating frication in the vocal tract and resulting in an inefficient production of speech. The frication causes a greater amount of inter-harmonic noise than the modal and falsetto phonation types (Laver, 1994).

CREAK (also referred to as vocal fry, glottal fry): Creaky voice is distinguished by an F0 below that of the modal voice. Harsh voice on the other hand consistently has a fundamental frequency above 100 Hz. Contrary to falsetto, creaky voice uses short thick vocal folds. They are generally loose, which is most probably caused by contracting the thyroid-arytenoid muscles (Esling et al., 2019). Creak is associated with aperiodic glottal pulses, resulting in a higher degree of jitter (Gordon & Ladefoged, 2001).

Keating et al. (2015) break down creaky voice into further categories. *Prototypical creaky voice* is characterised by low F0, irregular phonation, and a constricted glottis. Then there are five other types of creaky voice. Firstly, they distinguish *vocal fry* from prototypical creaky voice because although the glottis is constricted and the F0 is low, the phonation is not irregular, it is periodical. They also suggest the ventricular folds contribute to the low F0 in this type of creaky voice. *Multiply pulsed voice* involves two simultaneous periodicities, a low F0 and another that is roughly one octave higher. *Aperiodic voice* features another variant of F0 irregularity. This type of creak is so irregular that there is no discernible pitch, resulting in a very noisy signal. *Non-*

constricted creak is described as prototypical creak but instead of a constricted glottis, the glottis is spreading. This, combined with naturally low subglottal pressure is not an ideal condition for sustaining voicing, and is necessarily somewhat breathy. Finally there is *tense or pressed voice*. This type features a constricted glottis but not the prototypically low and irregular F0. It is used to make phonological distinctions in languages such as Mazatec, where creaky or laryngealised phonation can co-occur with a high tone (Keating et al., 2015). Esling et al. (2019) describe pressed voice as a form of harsh voice instead, but agrees that it features a higher F0 and further explains that it is produced with a forceful airflow through a tense glottis (more details in section 2.4).

HARSH: Harsh voice is distinguished by a rough, rasping sound, a raucous quality, and irregularity of the glottal wave-form. It is characterised by aperiodicity (jitter), varying amplitude (shimmer), and noise in the spectrum. It is caused by laryngeal tension and tightening of the aryepiglottic folds. The vocal and ventricular folds are compressed creating a relatively low pitch. It appears louder than modal voice, and when the pitch is low the aryepiglottic folds can also be induced to vibrate. This aryepiglottic trilling is also referred to as a growl (Esling et al., 2019). These effects are caused by the aryepiglottic constrictor mechanism, which will be further explained in 2.4.

BREATHY: Breathiness occurs when the vocal folds are vibrating without fully closing, making for inefficient vibration. Lessened glottal resistance leads to a higher rate of airflow than in modal voice. Whispery voice is distinguished from breathy voice by a narrowed epilaryngeal tube. On a glottal level whispery and breathy voice have the same configuration, it is laryngeal constriction that adds greater turbulent friction (Esling et al., 2019).

VENTRICULAR: Ventricular voice, also described as severely harsh voice, is caused by the vibration of the ventricular folds. It is distinguished by boosting the relative amplitude of the higher harmonics (Esling et al., 2019).

In addition to these phonation types, as described by Laver and Esling, there is a completely different way to create vibration in the vocal tract, using ingressive airflow rather than egressive. Egressive means the air moves from the lungs to the upper airway, whereas ingressive means the air moves toward the lungs. Egressive phonation (EP for short) is most commonly used since it is the more efficient form of voice production for humans (DeBoer, 2012). This is due to the configuration of the vocal folds. It is much easier to create and sustain phonation with EP, while ingressive phonation (IP for short) requires a much less economical use of air.

INGRESSIVE PHONATION: Creating phonation with an ingressive airflow instead of an egressive airflow. It is generally not used for speech production due to its inefficiency in generating vocal fold vibration. Compared to egressive phonation it is less sonorous and sounds harsher (Eklund, 2008). It has also been observed to produce a higher fundamental frequency. Orlikoff et al. (1997) found an average 5.1 semitone increase in F0 for IP compared to EP. This increase is caused by the lengthening of the vocal folds. The lengthening involves the cricothyroid muscles and makes the vocal folds thinner, which makes them vibrate faster. This is a natural process that happens during inspiratory voice production (DeBoer, 2012). Fornhammer et al. (2022) also mention the vocal fold vibration amplitude in IP is greater than with normal singing.

2.3 Vowels & tongue position

Vowel sounds are distinguished by their formants. Formants are amplified harmonics that are influenced primarily by tongue position. The first two formants are often used to define vowels within a vowel space (see figure 1 below).

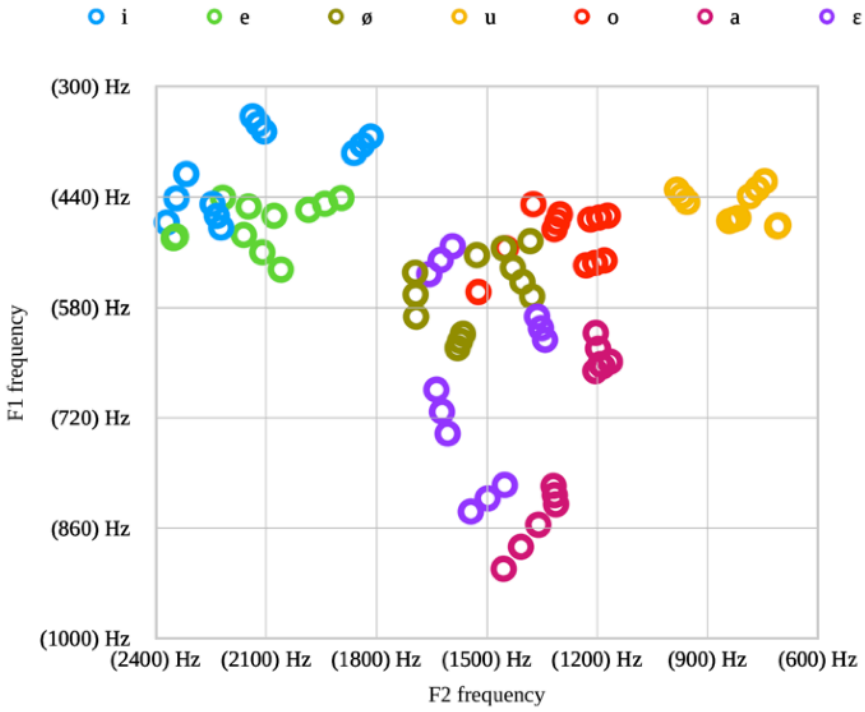

Figure 1 vowel space with normal voice of all participants

Swedish has 9 vowel phonemes that are often further differentiated into long and short vowels. McAllister (1998) sums them up as follows; /i/, /e/, /ɛ/, /a/, /y/, /ʉ/, /ø/, /o/, /u/. Note that each vowel phoneme has 2 to 4 allophones, this is where the differentiation between long and short vowels, as well as diphthongisation come into play (Engstrand, 2007; McAllister, 1998; Leinonen, 2008). There is also cross-dialectal variation in how vowels are realised (Ewald et al., 2019).

This study is focussed on the interaction between vowel formants and altered laryngeal and epilaryngeal settings. As such it is unnecessary to create a full vowel system and delve into all the different allophones of each vowel phoneme. We just need a vowel space that will give us a general idea of the F1 and F2 ranges, in order to see how it shifts when a voice actor uses different vocal characteristics. For this, 7 vowels were measured with a focus on peripheral vowels. The /y/ and /ʉ/ were omitted due to their similarity to /i/ and /u/ respectively. Figure 1 below shows the vowel space with the combined measurements of each participant's normal voice. The relative relationship between the vowels for each separate participant is similar but comparison between participants is difficult due to each individual's physiology and possible dialectal differences.

Advanced tongue root (ATR) is a setting that can affect vowel quality. Advancing or retracting the tongue root are also connected to other articulatory settings, such as laryngeal constriction. Some language use ATR as a contrastive phoneme property. In those languages the presence or absence of ATR is used to distinguish minimal pairs (Garellek in Katz & Assmann, 2019).

Tongue retraction results in a lowered F2 and F3 because it enlarges the resonating spaces that amplify these formants. Tongue retraction can also be correlated with a constricted pharynx, something that Esling et al. classify as laryngeal constriction. The cavity volume reduction around and within the epilaryngeal tube yields a raise in spectral frequencies. Retracting the tongue root and raising the larynx typically results in a high F1 and a low F2 (Esling et al., 2019).

This is contrasted with lowering the larynx, expanding the pharynx and advancing the tongue root. A lowered larynx setting (the opposite of constriction) naturally moves the tongue forward, the airway opens and the volume of the pharyngeal cavity expands. This articulatory configuration generates different resonances and predisposes opposing phonatory effects, in other words a lower F1 and a higher F2 (Esling et al., 2019).

2.4 Laryngeal Articulator Model (LAM)

The laryngeal articulator model or LAM, as described by Esling et al. (2019), provides a structure that helps binding together pitch, vowel space and phonation type. They expand the notion of phonation to include not just the glottis but also other vibrating structure through the epilaryngeal tube (Esling et al., 2019).

Phonation is most commonly produced by the vocal folds at the glottis, but there are two other possible sources of periodic vibration within the laryngeal mechanism: the ventricular folds, and the aryepiglottic folds. As mentioned in 2.2 the ventricular folds are used during harsh types of phonation, throat-singing for example. The aryepiglottic folds can generate yet another distinct periodic signal. A particular configuration of the laryngeal constrictor mechanism can allow the aryepiglottic folds to trill against the epiglottis (Esling et al., 2019).

Harsh voice as introduced in 2.2 is the result of laryngeal constriction. Acoustically, it is identified by jitter (irregular or aperiodic voicing frequency) and shimmer (varying amplitude). Physically the vocal folds and ventricular folds get compressed because of the tightening of the aryepiglottic folds. With the right amount of tension the aryepiglottic folds can start vibrating too. This is more likely to happen when the pitch is low. The pitch is low when there is little tension to stretch the vocal folds. This aryepiglottic trilling is what occurs in a growl. The neutral state of harsh voice is at mid pitch, in this case the laryngeal structures are more compacted from bottom to top and the aryepiglottic folds are less likely to vibrate. Another type of harsh voice is often called pressed voice, and occurs at high pitch, when supraglottic laryngeal constriction is maintained but the glottal length is increased. These mechanisms create an isometric tension (tensing the muscle without contraction), keeping the glottis closed. Phonation is then created by a forceful airstream (Esling et al., 2019).

Esling et al. (2019, p17) explain that voice quality can affect formants. Specifically the increase of F1 and the decrease of F2. They argue that during raised-larynx voice, aryepiglottic constriction of the epilaryngeal tube occurs, as well as tongue retraction, resulting in the lowering of F2 and F3. This means vowel measurements can provide information about the state of the larynx and whether there is constriction of the epilaryngeal tube.

The aryepiglottic constrictor mechanism produces pharyngealisation effects, epiglottalisation effects, laryngealisation and glottalisation together with associated tongue retraction and larynx

raising. Laryngeal constriction normally involves a raised larynx, this compresses the epilaryngeal tube and shortens the vocal tract. Tongue retraction can then cause the pharynx to be compressed vertically as well, resulting in smaller pharyngeal volumes (Esling et al., 2019).

2.5 Vocal stereotypes (Teshigawara)

Teshigawara (2003) did a phonetic study on voices in Japanese animation, comparing vocal stereotypes of heroes and villains in Japanese culture. The study identified critical vocal components that differentiate good and bad characters. The majority of heroes' voices exhibited a presence of breathy voice and an absence of pharyngeal constriction. Breathy voice, along with other phonation types, have been discussed in 2.2.

The majority of villains' voices exhibited non-neutral epilaryngeal states, such as laryngeal sphincteringor pharyngeal expansion. A perception test (Teshigawara, 2003. *Voices in Japanese Animation: How People Perceive Voices of Good Guys and Bad Guys*) showed that these vocal components were indeed related to villain voices. For hero voices the results were not as clear. F0 range did not differ much between hero and villain voices, but vowel formant F2 proved consistently lower in villain voices, possibly due to pharyngeal expansion.

The current study is not focussed on the dichotomy between good and bad characters, but rather the distinction between normal human voices and voices of monsters and non-human characters. It is important to note that these monsters are not necessarily bad, but they are expected to exhibit non-neutral epilaryngeal states similar to the bad guys in Teshigawara's research, and quite possibly to a larger extent as well. For example, not much difference in mean F0 and F0 range was found, while the F0 in monster voices is expected to change quite drastically.

Teshigawara employed both an acoustic analysis and a perception test. This study will focus on the acoustic analysis of the voices. Rather than a perception test, the participants were asked to describe the vocal characteristics and personality traits of the characters they came up with. This way, the results of the acoustic analysis can be related to the characteristics described by the participants themselves.

Another important difference is that Teshigawara used voices from existing movies. Following digitisation of speech portions used for the analyses, Teshigawara performed an auditory analysis

using Laver's (1980, 1994, 2000) voice quality descriptive framework. In addition a spectrographic analysis was performed.

Common vocal characteristics for adult make heroes were modal phonation, a lax laryngeal tension setting (breathy voice), and no particular deviation from the neutral supralaryngeal settings. The mean F0 and F0 range did not differ significantly between heroes and villains, but F2 was consistently lower in villain voices. Teshigawara attributes this lower F2 to pharyngeal expansion and, for female voices, to pharyngeal constriction as well.

Teshigawara focussed specifically on good vs bad characters in anime, and ended up excluding the film *mononokehime* from the study because the film features no clear good or bad characters, for example the main female villain also has hero characteristics.

Instead of looking at good vs bad characters, the current study focusses on attributes that can make the character sound more non-human. An advantage of using recorded data samples rather than using the audio of existing movies is that a one to one comparison can be made between the normal voice and the character voice. One could possibly find vocal samples by the same voice actor in another movie, but having a full sentence said in different character voices allows for a more direct contrast analysis.

The current study will be utilising similar acoustic measurements as used in Teshigawara’s study to relate personality traits to vocal characteristics. It is important to distinguish between personality traits, the corresponding vocal characteristics, and the acoustic measurements. Especially because the study mainly revolves around the measurable acoustic features, which in turn will be related back to the personality traits and vocal characteristics as described by the participants themselves.

Chapter 3 Methodology

In this chapter the recordings, parameters, and measurements will be discussed.

3.1 Participants & recordings

Four Swedish native speakers with a background in singing, vocal performance or vocal coaching, were recorded for the study. The names have been anonymised. They were asked to come up with two different non-human monster voices. The properties of these monsters were left to the discretion of the participants themselves, so they were free to come up with their own monster as long as it would be as far away from their natural voice as they could come up with.

In order to be able to do a one on one comparison the participants were given a Swedish sentence to read out, in their normal voice and in the two non-human monster voices. The recordings were made as follows. The participant would read the sentence once in their normal voice (n1), then in a non-human monster voice of their own choice (k1), then once again in their normal voice (n2), and lastly in another character voice (k2).

The sentence is based on a sentence from a Swedish children's book, *Stora Emilboken* by Astrid Lindgren (1984), and contains a suitable variety of different vowels.

> *“Den lilla snälla bonden från Vena halade också fram en ettöring ur byxfickan, men han ångrade sig innan det var för sent, och stoppade ner den igen.”*

The recordings were done in a recording booth at Lund University’s Humanities Lab using a Tascam DR-40X. Each participant was also asked to provide a description of the properties and vocal characteristics of their different vocal performances.

The participants were also recorded performing a humming tone (in their normal voice and in the character voices), and a sweeping tone in order to acquire more data on their vocal capabilities. The details of this will be further described in the next section.

3.2 Parameters & measurements

In this section the parameters for this study will be presented. Voice quality cannot be analysed by measuring one specific component, so in order to analyse the acoustic signal, the following data will be measured.

pitch
- average pitch (median pitch & mean pitch in speech, pitch for humming tone)
- pitch variation (standard deviation, minimum & maximum frequency for speech)
- pitch range (semitone range between minimum & maximum frequency for sweep)

phonation
- harmonics-to-noise ratio (HNR)
- irregular phonation (jitter & shimmer)

vowel system
- formant frequencies (F1, F2 & F3)

3.2.1 Hertz & semitones

The pitch measurements were done in Hertz, which provides an absolute value. Frequencies increase on a logarithmic scale however, so in the result section semitone calculations are also used to properly compare pitch differences and pitch ranges. Below are the formulas that were used in order to calculate the pitch ranges in semitones or the semitones relative to 1 Hz.

pitch range = 12*log2(f0max/f0min)

semitones re 1 Hz = 12*log2(f0mean/1)

The standard deviation was calculated in Praat but these calculations are in Hertz and do not result in comparable values as the median pitch for each vocal performance is different. A 10 Hz deviation is relatively large with a deep voice, but a 10 Hz deviation at an F0 of 300 Hz is relatively small. This is why an additional calculation was made to get the relative pitch range in semitones.

Apart from the sentences, a few other recordings were made to collect additional voice data. The additional recordings include a sweeping tone, and a humming tone for the normal voice and the two character voices. The sweeping tone shows the range of the natural voice, going from the lowest to the highest frequency in one sweep.

The participants were also asked to produce a humming tone. The humming tone is closest to the natural F0 and reflects the physiology of the system, or the assumed physiology of the monster they are trying to portray.

The mean and median pitch for each different speech sample were also calculated, but did not always match up with the average pitch of the humming tone. This is most likely the result of irregular phonation and was seen mainly in very creaky or harsh voice samples.

3.2.2 Noise & aperiodicity

The voice signal contains harmonics, but also noise. The level of noise in the signal can be expressed by calculating the difference in amplitude between the harmonic and inharmonic components of the source spectrum (Garellek in Katz & Assmann, 2019). This is called the *harmonics-to-noise ratio* (HNR). The higher this ratio, the more clearly you can hear the harmonics. If there is more noise in the signal, perhaps due to turbulence in the vocal tract, the HNR value will be lower.

The signal can also be aperiodic. This is expressed in jitter, the higher the jitter value, the more aperiodicity in the pulses. Shimmer is the irregularity in the amplitude of the pulses, the higher the shimmer, the more irregularity in amplitude (Gordon & Ladefoged, 2001).

To measure the noise and aperiodicity, only one part of the recorded sentence was used. This part contains all voiced phonemes. The part was selected and then Praat’s ‘voice report’ function was used to get the necessary data.

3.2.3 Measuring vowels

The first, second and third formants of vowels were measured in Praat. The measurements were made by taking the average of the interval in the middle of the vowel, excluding approximately the first and last 14 milliseconds to avoid interference from the surrounding consonants. The first three formants (F1, F2, F3) were measured, this data can be found in the appendix.

Table 1 Vowels used in analysis

word	Swedish orthography	phonemes
lilla	i	i (ɪ in some sources)
snälla	ä	ɛ
bonden	o	u (ʊ in some sources)
vena	e	e
också	å	o
fram	a	a (ɑ in some sources)
öring	ö	ø (œ in some sources)

The table above contains the vowels used in this study. It shows both the symbols used in Swedish orthography, and the corresponding phonemes (Engstrand 2007, McAllister 1998).

Chapter 4 Results

4.1 Pitch & phonation type

The normal voice was recorded twice, both recordings were analysed and the average of the measurements is shown in the first column in the tables below. The three rows for humming pitch, and the minimum and maximum pitch in sweep, are not mean values since these were recorded only once. The same goes for the second and third columns, character voices k1 and k2 were recorded once each. The semitones re 1 Hz for the median pitch for each voice has also been calculated, as well as the semitone range from the minimum to the maximum pitch in speech.

The results specifically related to phonation type (jitter, shimmer, HNR) are shown in the last three rows in tables below. The normal voice column shows the average of the two recordings that were made. The k1 and k2 voice were recorded once each.

A humming tone was recorded for the normal voice and for the character voices. However the results were not always trustworthy, the humming tone for M-k1 did not even yield a pitch measurement, and the results for M-k2 came out at 871 Hz, which is way too high to be representative of the actual pitch. Instead the median pitch measurements were used, as they were deemed more accurate and reliable. (goes into the methodology chapter)

Table 2.1 Pitch measurements participant E ♀

E ♀	normal voice	voice k1	voice k2
median pitch in speech (in Hz)	207	159	305
mean pitch in speech (in Hz)	201	166	315
standard deviation (in Hz)	49	42	86
median pitch in semitones re 1Hz	92	88	99
min pitch in speech (in Hz)	152	102	70
max pitch in speech (in Hz)	283	291	559
min to max pitch range in semitones	11	18	36
humming pitch (in Hz)	198	180	543
minimum pitch in sweep (in Hz)	112	*	*
maximum pitch in sweep (in Hz)	1265	*	*
jitter (irregular frequency)	1.5 %	6.9 %	1.6 %
shimmer (irregular amplitude)	6.6 %	20.0 %	7.4 %
harmonics-to-noise ratio (in dB)	15.642	2.844	13.297

If we look at table 2.1 we can see that the median pitch in semitones shifts -4 and +7 semitones respectively for each character voice. The pitch range increases a bit for the k1 voice from 11 semitones to 18 semitones. For the k2 voice it increases to 36 semitones.

Character voice k1 has higher jitter and shimmer, and much lower HNR, along with a lower pitch, this indicates creaky or harsh voice. For character k2 she goes up with the pitch while there is only a slight increase in shimmer and slightly more noise in comparison to the normal voice.

Table 2.2 Pitch measurements participant M♂

M♂	normal voice	voice k1	voice k2
median pitch in speech (in Hz)	110	305	71
mean pitch in speech (in Hz)	110	305	74
standard deviation (in Hz)	16	164	15
median pitch in semitones re 1Hz	81	99	74
min pitch in speech (in Hz)	75	58	57
max pitch in speech (in Hz)	152	648	111
min to max pitch range in semitones	12	42	12
humming pitch (in Hz)	98	*	871
minimum pitch in sweep (in Hz)	72	*	*
maximum pitch in sweep (in Hz)	882	*	*
jitter (irregular frequency)	3.0 %	4.5 %	5.0 %
shimmer (irregular amplitude)	11.8 %	25.0 %	23.3 %
harmonics-to-noise ratio (in dB)	7.776	4.235	1.068

The humming pitch for the k1 voice (*) was too distorted to measure, but both the median and mean pitch in the speech sample were measured at 305 Hz. The median pitch shifted +18 semitones for k1 and -7 semitones for k2. The pitch ranges for the normal voice and the deeper k2 voice are roughly the same at 12 semitones, while the k1 voice (using inhaled speech) is much wider with a range of 42 semitones. The humming pitch for the very harsh k2 voice could only be measured at 871 Hz. The median and mean pitch measurements give a more realistic pitch measurement at roughly 72.5 Hz.

For character voice k1 ingressive phonation was used. A strong increase (+18 semitones) in pitch can be seen, as well as an increase in irregular phonation and noise. However, the k2 voice, which uses egressive phonation, shows even more noise, a strong indication for extremely harsh voice, along with the decrease in pitch (-7 semitones).

Table 2.3 Pitch measurements participant S♀

S♀	normal voice	voice k1	voice k2
median pitch in speech (in Hz)	190	96	499
mean pitch in speech (in Hz)	202	102	529
standard deviation (in Hz)	45	38	116
median pitch in semitones re 1Hz	91	79	108
min pitch in speech (in Hz)	131	33	339
max pitch in speech (in Hz)	324	220	873
min to max pitch range in semitones	16	33	16
humming pitch (in Hz)	220	64	894
minimum pitch in sweep (in Hz)	124	*	*
maximum pitch in sweep (in Hz)	2168	*	*
jitter (irregular frequency)	1.8 %	2.1 %	0.8 %
shimmer (irregular amplitude)	9.8 %	16.9 %	6.4 %
mean harmonics-to-noise ratio (in dB)	13.855	3.550	20.074

The median pitch for participant S♀ is generally higher again since it is a female participant. There is an impressive -12 semitone shift for her k1 voice, and a +17 semitone increase for her k2 voice. Her pitch range actually increases with the deeper k1 voice, a 33 semitone range compared to 16 semitones for both her normal voice and k2 character voice. However, the increased shimmer, jitter and noise make it very likely that the minimum and maximum pitch measurements for the k1 voice are not representative of the actual pitch range, or that the measurements themselves are skewed because of the irregularities in the speech signal.

Participant S♀ showed the most versatility in terms of pitch, with a -12 and +17 semitone pitch shift for her k1 and k2 character voices. The roughly one full octave lowered k1 voice comes with a lot more noise and increased jitter and shimmer. Her k2 voice, highest out of all the character voices in this study, actually shows cleaner phonation, with less jitter, shimmer and

noise than the normal voice. This is a strong indication for falsetto, with boosted harmonics possibly due to laryngealisation.

Table 2.4 Pitch measurements participant T♂

T♂	normal voice	voice k1	voice k2
median pitch in speech (in Hz)	100	62	155
mean pitch in speech (in Hz)	106	62	161
standard deviation (in Hz)	20	7	39
median pitch in semitones re 1Hz	80	72	87
min pitch in speech (in Hz)	61	53	89
max pitch in speech (in Hz)	164	89	268
min to max pitch range in semitones	17	9	19
humming pitch (in Hz)	90	49	185
minimum pitch in sweep (in Hz)	74	*	*
maximum pitch in sweep (in Hz)	558	*	*
jitter (irregular frequency)	3.0 %	5.7 %	1.8 %
shimmer (irregular amplitude)	11.6 %	14.4 %	8.6 %
harmonics-to-noise ratio (in dB)	5.624	0.476	10.956

Participant T shows a -8 semitone decrease for k1 and +7 semitones for k2. With a relatively big range in his normal voice (17 semitones), k1 is clearly more monotone with a 9 semitone pitch range. For k2 however, the range is slightly expanded to 19 semitones.

On average the participants lowered the median pitch of their voice -7.75 semitones and raised it by +12.25 semitones for the different character voices. The highest increase and decrease in semitones were +18 and -12. This shows it is generally easier to increase one's pitch but lowering one's pitch is not so easy. Important to note is that these numbers do not indicate the

maximum increase or decrease, but rather the amount deemed appropriate by the participants for their interpretation of a monster voice.

The minimum and maximum pitch for each participant were also measured. Interestingly, these values showed that some participants could have potentially increased their pitch over two times more than they did for their character voice. It also showed some participants went even lower for their character voice, lower than their minimum pitch indicated they could. This shows that using a different phonation type can extend the pitch range of the normal voice. For the minimum and maximum pitch measurements only modal voice and falsetto were used, while for the character voices different phonation types were employed.

4.2 Vowel formant measurements

The Swedish vowels used in the analysis are i, e, ö, o, å, a and ä (respectively i, e, ø, u, o, a, ɛ in IPA). An overview of the vowels and the corresponding phonemic symbols can be found in table 1. The /e/ (e in Vena) is realised as a diphthong, the end is lowered compared to the onset which is closer to /i/. The /ɛ/ (ä in snälla) was quite inconsistent, sometimes very central next to /ø/ (ö in öring), other times even lower than the open vowel a. Lastly, /o/ (å in också) seems to be rather centralised, especially in relation to /u/ (o in bonden).

There are three graphs for each of the four participants. The first graph shows the vowel formants of their normal voice, the second one shows their first character voice (k1), and the last graph shows their second character voice (k2). The graph for the normal voice shows three values for each vowel, these consist of two separate measurements and the average of the two. The normal voice was recorded and measured twice to make sure there were not any major inconsistencies. This is important as the normal voice is used as a baseline to compare the character voices to.

A complete table with all the formant frequency measurements (including F3, which is not shown in the graphs) can be found in the appendix.

4.2.1 Vowel shift participant E♀

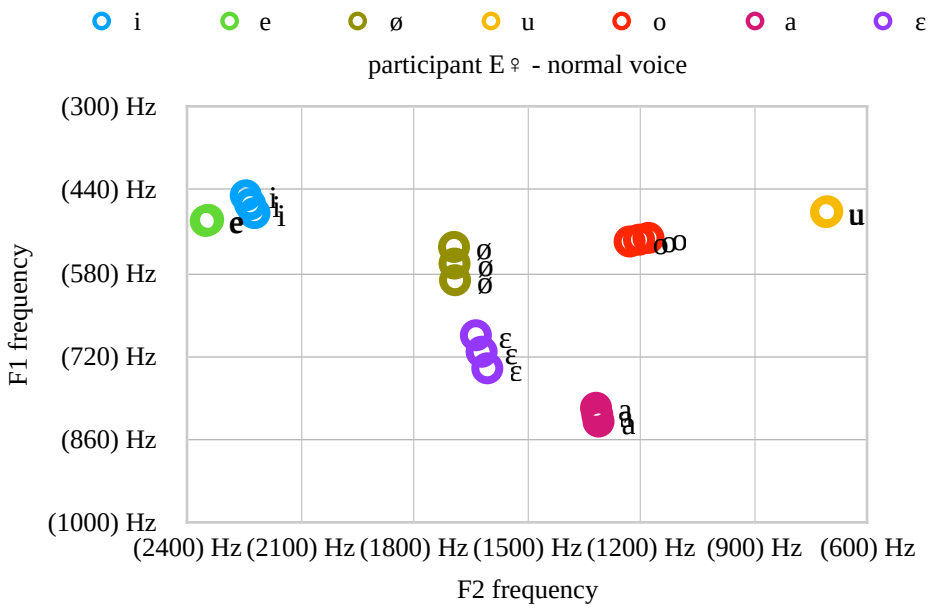

Figure 2.1 Vowel space of participant E♀ in normal voice

This is the vowel space of the participants normal voice. It shows the measurements of two separate recordings and the average value. This shows there is indeed some variation even if there is no difference in voice quality. The next two graphs show the measurements of two different character voices.

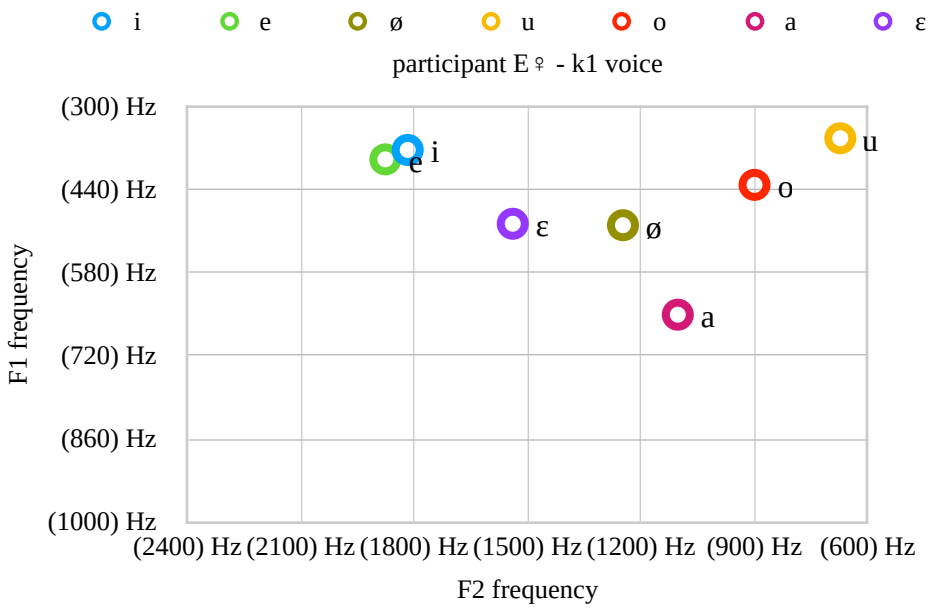

Figure 2.2 Vowel space of participant E♀ in character voice 1

In figure 2.2 all vowels are more backed, and generally higher compared to the normal voice. Both F1 and F2 are significantly lower compared to the normal voice.

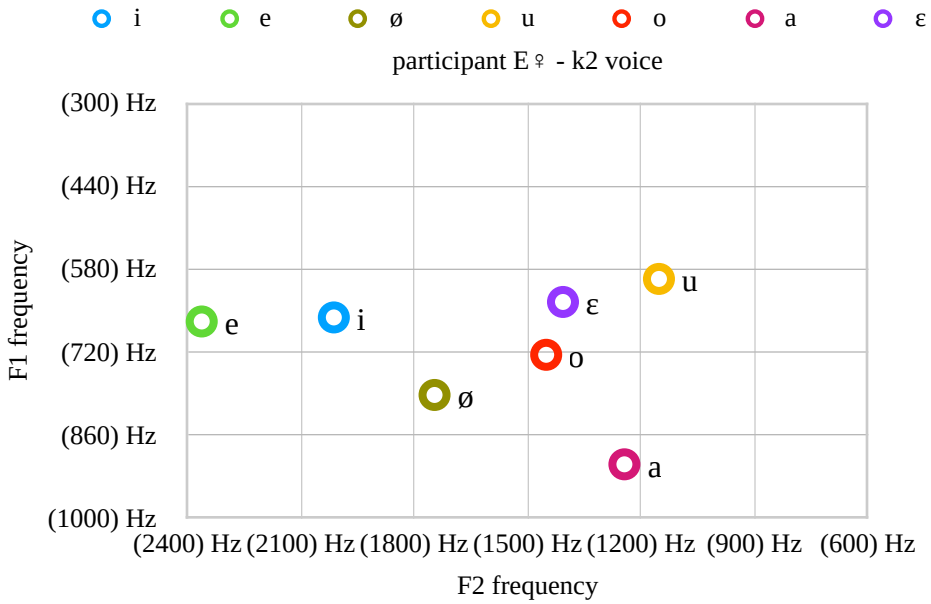

Figure 2.3 Vowel space of participant E♀ in character voice 2

In 2.3 we can see a general lowering of all the vowels. The F1 frequency is a lot higher in comparison to the normal voice. The system is also more fronted in its entirety with a surprisingly high F2 frequency for /u/. The relative positions of /ɛ/, /o/, and /ø/ seem to have

shifted as well which is strange, but even with these inconsistencies we can clearly see the frequencies of F1 and to a lesser extent, F2 have been raised.

4.2.2 Vowel shift participant M♂

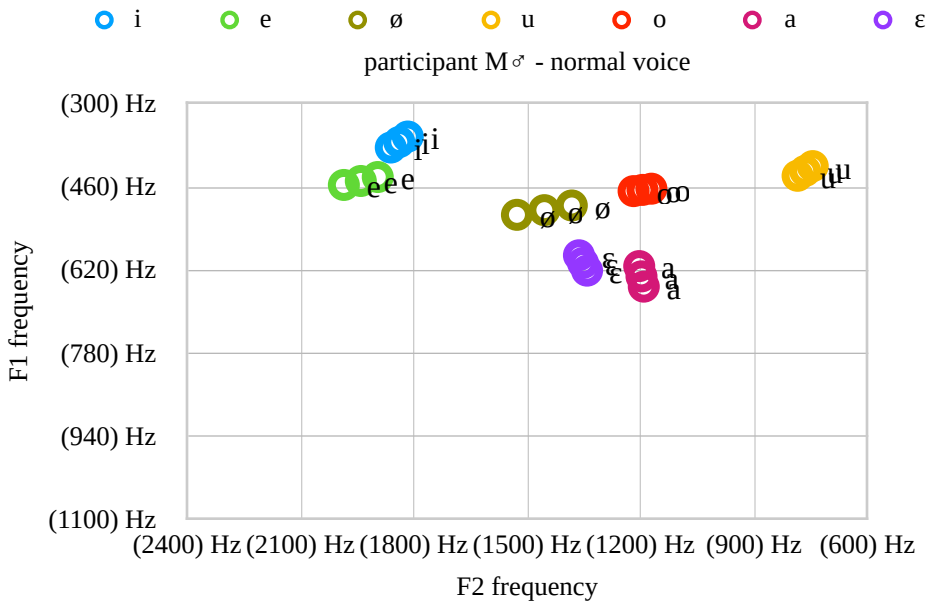

Figure 3.1 Vowel space of participant M♂ in normal voice

Vowels are where they are expected to be, generally low F1 and F2 frequencies compared to the character voices. The F1 frequency is also low compared to the normal voices of the female participants. This shows how the naturally larger laryngeal cavity in males affects the F1.

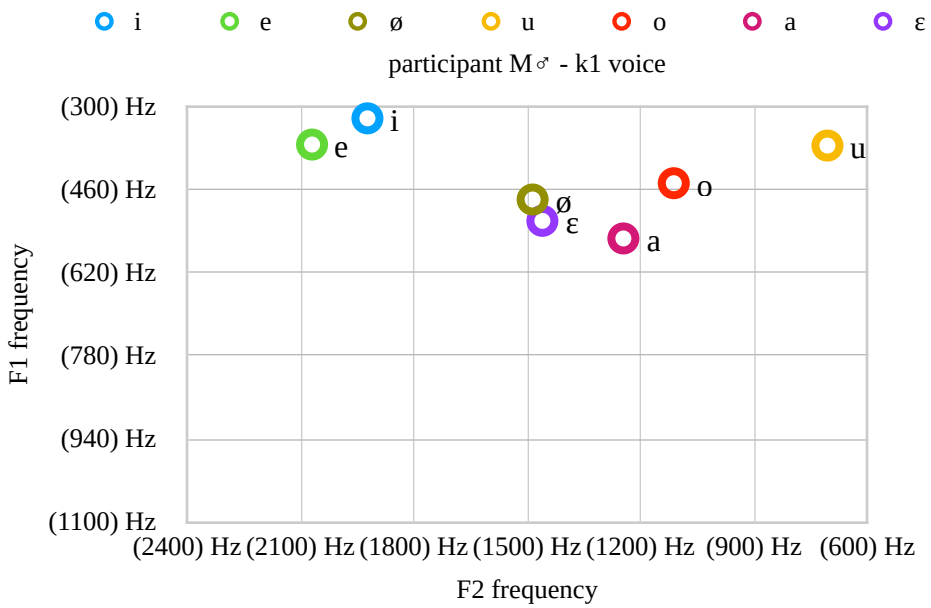

Figure 3.2 Vowel space of participant M♂ in character voice 1

Although quite similar to the normal voice all vowels are a bit higher. In other words, F1 is lower for all vowels, and the F1 range is also slightly shorter. Not a significant shift from the normal voice.

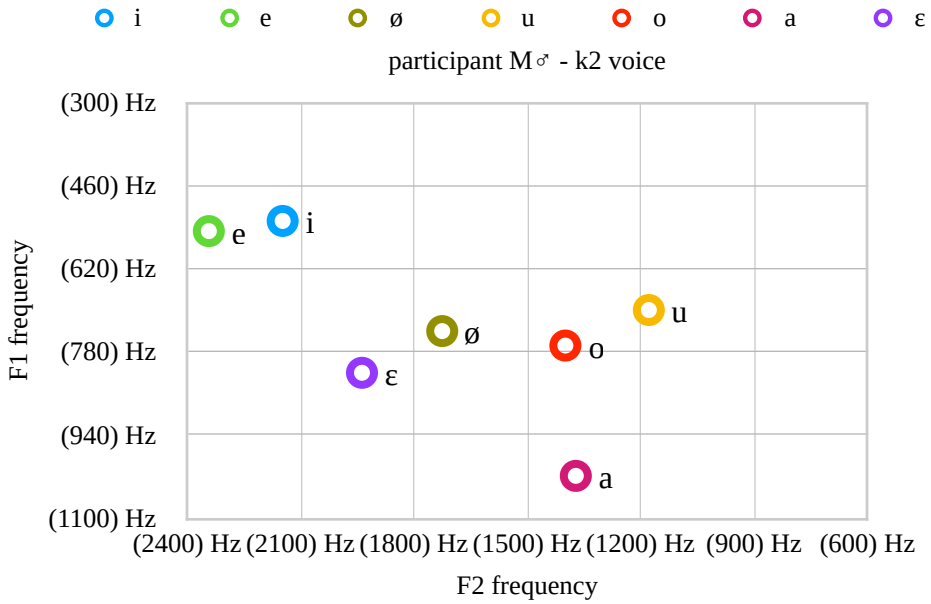

Figure 3.3 Vowel space of participant M♂ in character voice 2

Both F1 and F2 frequencies are much higher, which means fronted and lowered vowels in comparison to the normal voice. Especially /u/ is much more fronted than usual.

4.2.3 Vowel shift participant S♀

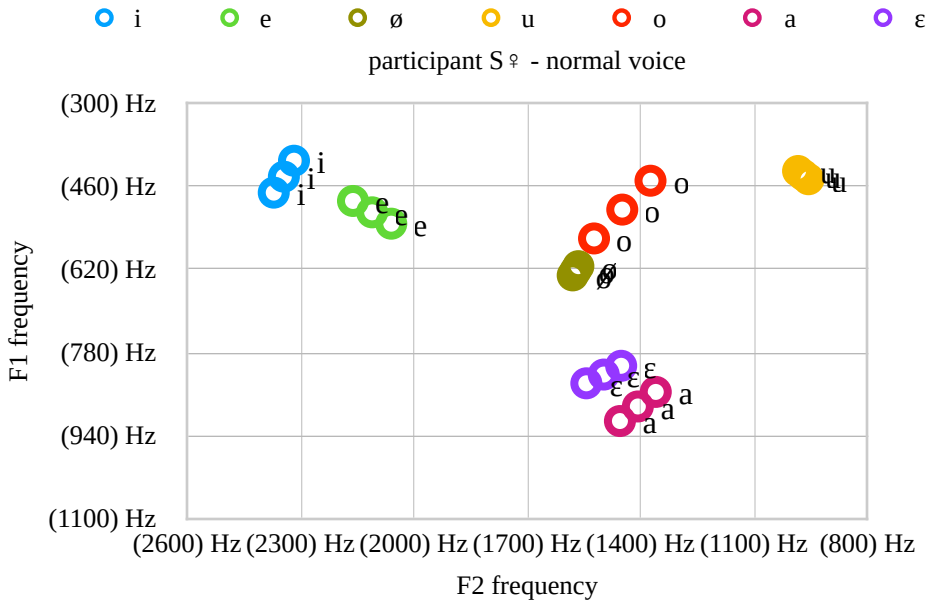

Figure 4.1 Vowel space of participant S♀ in normal voice

Quite some variation in the measurements for /o/, from almost centralised enough to assimilate into /ø/, to being at the same height as /u/. Additionally, /ɛ/ and /a/ are very close together. Generally /ɛ/ is expected to be more fronted with a higher F2.

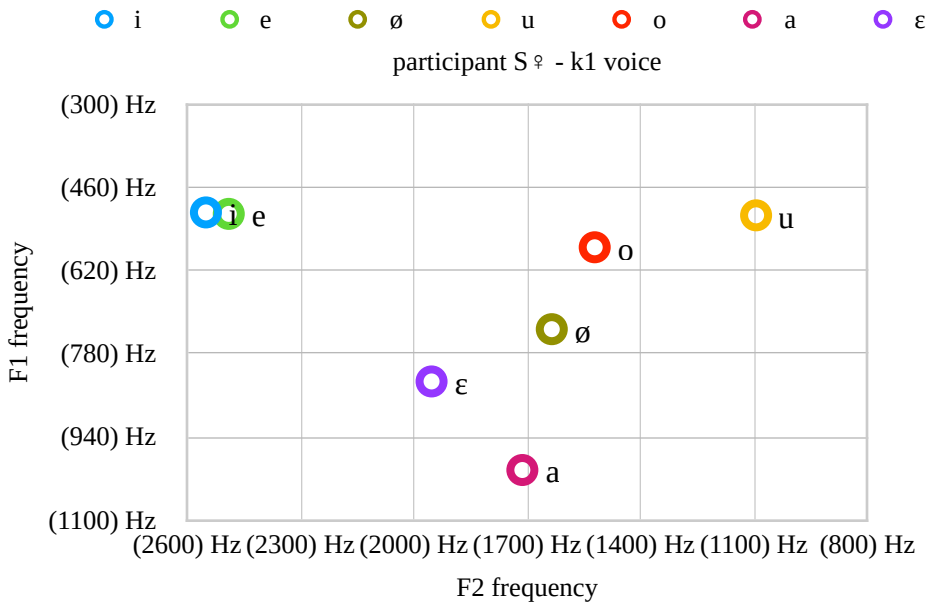

Figure 4.2 Vowel space of participant S♀ in character voice 1

The whole vowel system is lowered and more fronted. The relative position of /ɛ/ is shifted, it is more distinct from /a/ in the k1 voice than in the normal voice. The /i/ and /e/ contrast on the other hand, seems to be less distinct than in the normal voice.

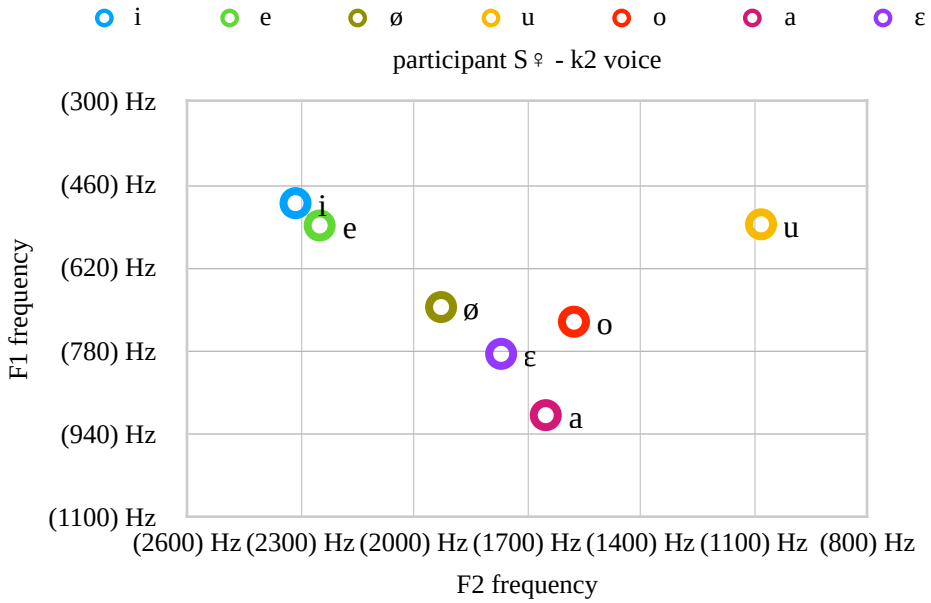

Figure 4.3 Vowel space of participant S♀ in character voice 2

A shortened F1 range compared to the normal voice, all the high vowels are lowered while low vowels like /a/ stay roughly the same. The /ø/ and /o/ vowels seem to have undergone a relative shift. Especially /o/ is more centralised with a bigger distance to /u/. Lastly, /u/ seems to be fronted a little bit, but not significantly.

4.2.4 Vowel shift participant T♂

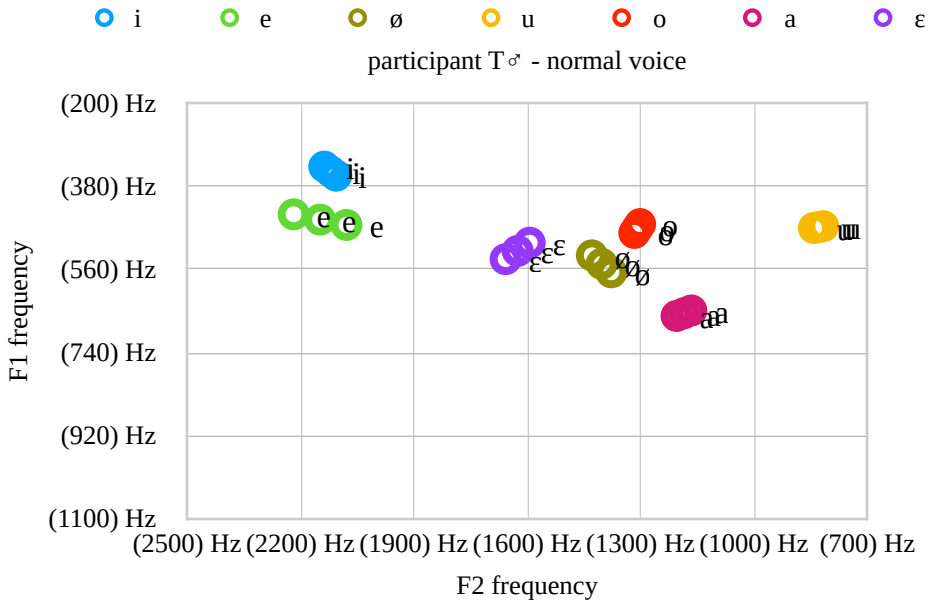

Figure 5.1 Vowel space of participant T♂ in normal voice

The F1 frequency is relatively low, even for low open vowels like /a/. Similar results are found for the normal voice of participant M♂, and it exemplifies that it is not just F0 pitch that differentiates male and female voices, but that in general the F1 frequency is also a lot lower for male speakers. The difference in F2 frequency is more subtle.

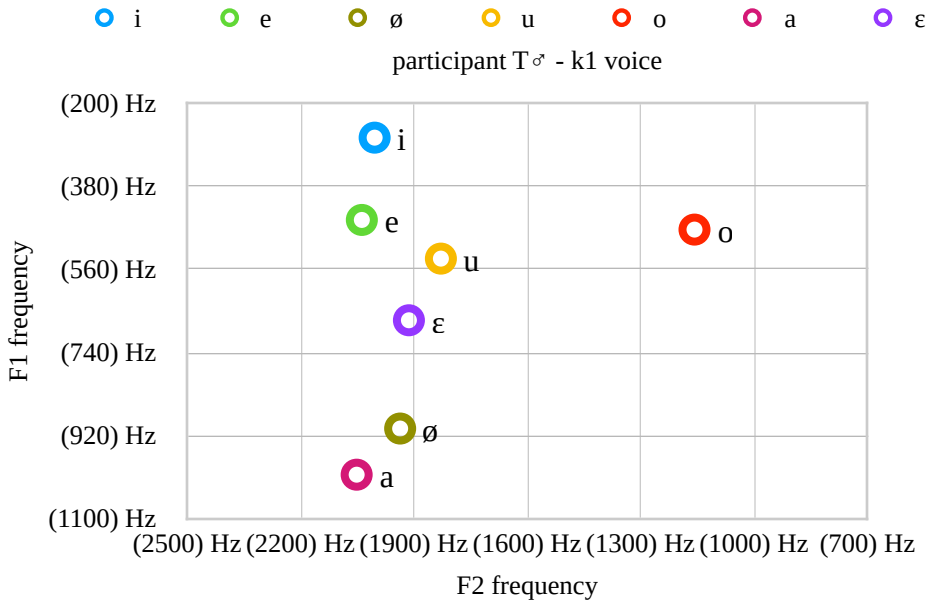

Figure 5.2 Vowel space of participant T♂ in character voice 1

With the exception of /o/ it seems the F2 range has been minimised so most vowels are at around 2000 Hz. The F1 range on the other hand has increased a lot, from /a/ at about 700 Hz in the normal voice to /a/ at about 1000 Hz in character voice k1, while /i/ goes from 350 Hz to 270 Hz. Meaning the F1 range from the highest to the lowest vowel for the normal voice is 350 Hz to 700 Hz while for k1 it is 260 Hz to 1000 Hz. The F3 frequency is higher for all but one of the vowels.

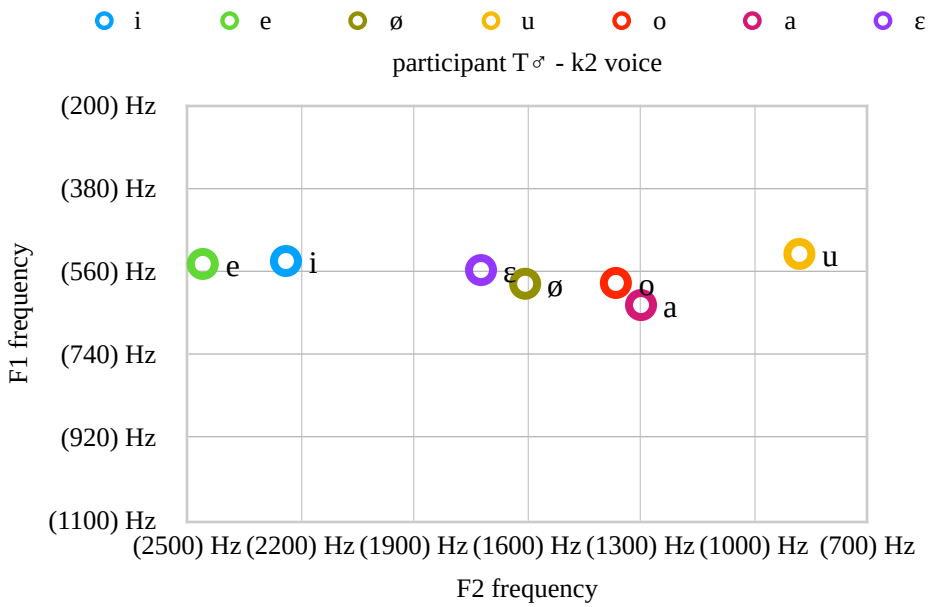

Figure 5.3 Vowel space of participant T♂ in character voice 2

High vowels are lowered while low vowels are higher than expected. The F1 range is very small. The F2 range on the other hand is slightly wider compared to the normal voice (figure 5.1). He is using an articulatory setting that evens out the F1 range, possibly a slightly raised larynx in combination with tongue root retraction to keep the laryngeal cavity relatively small and raising the F1 frequency. He does this while retaining the F2 range, meaning that tongue movement is maintaining a clear distinction between front and back vowels. The F3 frequency is generally lower for most vowels.

4.3 Character & vocal trait description

Each participant was asked to describe the character and the corresponding vocal traits they imagined for the non-human character voices (k1 & k2) they produced. This section includes a comprehensive list of their responses which will be further discussed in the discussion chapter.

participant E♀ - k1:

- character description: furry small alien, frightening, not that evil, soft voice, not sharp, loud, scary.
- vocal traits: overdrive, lot of resonance, very deep, dark, loud.

participant E♀ - k2:

- character description: animal crossing (pop-culture reference), non-human but speaking language, cute, happy, rascal, small, excited.
- vocal traits: high voice, singing intonations, tense larynx, not round, straight voice, more Scanian intonation. more stress on intonations, sharper.

participant M♂ - k1:

- character description: chaotic neutral, evil, damned soul, scales, purple, dark red details, cat-eyed.
- vocal traits: strained, lot of air, raspy, agonised, inhaled.

participant M♂ - k2:

- character description: selfish, mischievous personality, manic.
- vocal traits: loud, raspy, stressed (melody), strained, growling.

participant S♀ - k1:

- character traits: evil, cunning, animal, fearless, dominating, mean monster, male, abnormal, stranger things monster (pop-culture reference).
- vocal traits: lowered pitch, hoarse, distortion, deep/dark voice, no accent change just focussed on voice.

participant S♀ - k2:

- character description: puppy, boss baby (pop-culture reference), impatient, whiny, childlike, genderless or girl.
- vocal traits: bright, shrill, squeaky, more whiny Stockholmish.

participant T♂ - k1:

- character description: dragon, troll, orc.
- vocal traits: deep growly voice, breathy, false vocal cords instead of the regular.

participant T♂ - k2:

- character description: smaller, goblin, mosquito.
- vocal traits: buzzing, forward placement, higher pitch, nasal.

Chapter 5 Discussion

5.1 Pitch changes & vowel shifts

5.1.1 Pitch variation

The participants showed a tendency to lower their pitch for their first monster, while increasing it for the second one. Interestingly, the increase in their pitch was relatively higher than the decrease was, indicating physical limitations when lowering the F0. There was no clear indication that male and female speakers lowered their pitch in different ways. The male speakers went down 7 and 8 semitones, while the female speakers went down 4 and 12 semitones respectively. In terms of absolute values, the male speakers did have a lower frequency but relative to the normal voice, which was also naturally lower for the male speakers, the average decrease is the same as for the female speakers. In terms of relative pitch, it was in fact one of the female speakers who showed the biggest shift (-12 semitones for S-k1).

The minimum and maximum pitch measurements (sweeping tone using modal voice and falsetto) for each participant did not always correlate with the (spoken) pitch range of their character voices. For example, some participants had an even lower pitch for their character voice than the minimum pitch indicated was possible. This shows that pitch range can be extended beyond the minimum pitch of the modal voice. For the character voices with a low pitch creaky, harsh, and possibly even ventricular voice were employed. This is how each participant was consistently able to get to a lower pitch with their character voices than with their modal voice.

When the participants used a higher pitch for one of the character voices, the increase in pitch was between 7 to 18 semitones, on average a much bigger shift than for the low pitch voices. Interestingly, the measured vocal range could have allowed for an even bigger increase, but all participants stayed well below their maximum pitch.

5.1.2 Pitch range

On average the pitch range for the participants' normal voice was about 14 semitones, with 11 and 17 as highest and lowest values. There seems to be a trend to increase the pitch range for the character voices, as most of them, regardless of whether the average pitch went up or down, the pitch range increased. For one of the character voices (participant T♂'s k1) the range actually decreased, to a mere 9 semitones, while others went much wider, up to 36 semitones for egressive phonation and 42 semitones for ingressive phonation.

There were some difficulties analysing the recordings however, the semitone range for participant E's normal voice was first measured at 30 instead of 11, but upon further inspection the low pitch values turned out to be due to creaky voice and the higher readings were also caused by a sudden jump in pitch likely due to irregular phonation. This brings up the question of how to deal with these types of irregularities in the character voices. Perhaps the semitone range is not quite representative of the actual speech. But with the amount of creak and harshness in these recordings, correcting the measurements manually without skewing the results is a difficult task.

5.1.3 Vowel shifting

There seems to be a tendency to use laryngeal constriction and change tongue position, but only when the phonation type is still relatively normal. With ingressive phonation for example, the vowel system did not shift significantly compared to the normal voice (figure 3.2). This might be due to the already challenging task of producing ingressive phonation, or due to a more conscious decision to keep the vowel system the same to maintain comprehensibility.

Instead of shifting the entire vowel system some character voices, most noticeably in the k1 and k2 voices of participant T♂, almost completely neutralised either F1 or F2 distinctions, effectively compressing the vowel system horizontally or vertically. If the F2 range was compressed, as can be seen in figure 5.2, the F1 range was conversely expanded. If, on the contrary the F1 range was compressed, the F2 range was expanded (Figure 5.3). This could be an unconscious effort to retain vowel distinction when one of the formants is largely restricted to a very limited frequency range. It also tells us that F1 and F2 frequency ranges can be altered separately without necessarily affecting the other.

There were also individual shifts, especially the more central vowels /ɛ/ & /ø/ would be in a different position relative to the surrounding vowels. Perhaps there was a general centralisation which minimised this distinction in the first place, or perhaps it was due to anomalies in the measurements. Another anomaly is the /o/ in figure 5.2 (T-k1). All other vowels have lost most distinction in the F2 frequency, but /o/ stands out with a lower F2 frequency than even the normal voice has.

5.2 Phonation type & formant distribution

The literature provides descriptions of the phonatory effects of the various phonation types, so we can relate the acoustic measurements to these effects and infer from that what sort of phonation type was most likely used to produce the different character voices. It is however difficult to say with certainty which phonation type is used based purely on the results of the acoustic analysis.

5.2.1 Character voices with lower pitch

Starting with the male participants, whose natural pitch is already relatively low, we can see that when they lowered their pitch for one of the character voices, their mean pitch went down about 7 to 8 semitones. Both character voices are likely a kind of harsh voice which, according to Esling et al. (2019), is characterised by low F0, increased jitter, shimmer, and noise in the spectrum.

Participant M♂'s k2 exhibited more shimmer, where participant T♂'s k1 increased more in jitter. Participant T♂'s k1 was also much breathier which resulted in the lowest HNR value in any of the voice samples. This means it was extremely noisy and contains a lot of inharmonic components.

Participant S♀ produced a harsh voice for her k1 character as well and her relative pitch went down even more than the male speakers, with a decrease of roughly -12 semitones. Similar to participant M♂'s k2 there was much more shimmer, and compared to her normal voice the HNR was much lower, indicating breathiness (which increases the inharmonic frequencies in the signal).

Laryngeal constriction, as described by Esling et al. (2019), reduces the volume of the epilaryngeal tube, and together with the retracting of the tongue root it results in a higher F1 and a lower F2. The F1 and F2 frequencies both became higher for M-k2 and S-k1 indicating laryngeal constriction with a conversely fronted tongue root instead.

T-k1 also showed a relatively high F1, but the range was also drastically shortened, and when looking at F2 and F3 we can see some differences. The F2 range had conversely expanded, and the F3 frequency increased for nearly all the vowels as well. S-k1 also featured relatively high F3 frequencies, but for M-k2 there was no consistent F3 increase.

Participant E♀ also produced a lower pitch for her k1 voice, but rather than harsh voice, it seems she used creaky voice. The voice was 4 semitones lower than her regular voice, a smaller difference than we found for the harsh voices described above. In the vowel formants we can see a clear difference, instead of higher F1 and F2 frequencies, participant E♀'s k1 features both a lower F1 and a lower F2. Her F3 was also lower. The low pitch, along with increased jitter, shimmer and noise is probably caused by use of prototypical creaky voice, with lip-rounding for a low F2, constricted glottis, lowered larynx, and short thick vocal folds.

5.2.2 Character voices with higher pitch

Starting with the character with the higher pitch of all, at around 500 Hz, participant S♀'s k2 features less noise, less jitter, and less shimmer. The F1 range is shortened, with a slightly raised F1 frequency. This indicates there is some laryngeal constriction, while the high F0 is a strong indication of falsetto, but the lack of noise points to very low breathiness, something that often comes with falsetto as pulmonic airflow tends to escape through the incompletely closed parts of the glottis (Esling et al., 2019).

Participant E♀'s k2 similarly feature a high F0 and higher F1 frequencies as well. The phonation is less clean however, with a slight increase in jitter, shimmer, and noise compared to the normal voice. These results indicate a form of falsetto again, with higher degree of larynx raising. The F2 frequency is slightly higher, an indication there is only limited tongue retraction.

Participant T♂'s k2 also features a higher pitch, but not to the degree of falsetto as in the last two character voices. It has less jitter and shimmer, and a lot less noise. This rules out creaky, harsh, or breathy phonation. It sounds very nasal and tense. The F1 range is minimised and relatively high indicating a sort of constant larynx raising. This is done while retaining the F2 range,

meaning that tongue movement is maintaining a clear distinction between front and back vowels. All these results point toward tense or pressed voice, which features a constricted glottis but no irregular phonation or low F0 (Esling et al., 2019).

Ingressive phonation

Participant M♂ had an interesting approach for the first character voice (k1). Instead of using egressive airflow, he used ingressive airflow to create phonation. Interestingly enough, although the pitch and phonation parameters are very different or immeasurable in some cases, the vowel system is nearly identical to that of the normal voice. This indicates a lack of altered laryngeal states and absence of tongue retraction or advancement. Perhaps the ingressive airflow creates enough alteration or distinction from the normal voice and keeping the vowel system relatively unchanged is a way to maintain comprehensibility.

The results also correspond with Orlikoff et al. (1997) who found that there was an increase in F0 in IP compared to EP. However, they found an average increase of 5.1 semitones, while the results of participant M♂'s k1 voice showed an increase of 18 semitones. Perhaps the irregular phonation skewed the measurements, or perhaps the participant purposefully lengthened his vocal folds more than naturally happens during inspiratory voice production in order to sound even less 'human'. The character description provided by participant M contains adjectives like "strained, raspy, agonised" which indicates there was a conscious effort to add more tension to the vocal folds.

5.3 Voice types & character traits

The participants were asked to describe the vocal characteristics and personality traits of their two characters. Does what they describe match with their performance according to the measurements?

5.3.1 Character voices with lower pitch - harsh & creaky

Ventricular voice

Participant M♂ describes his k2 character as "selfish, mischievous personality, manic", with the following vocal traits: "loud, raspy, stressed (melody), strained, growling". The measurements

show a lower pitch, although the pitch measurements in the humming tone are extremely high. The median pitch is lowered -7 semitones, and the sample features increased jitter, even more so for shimmer, and a lot more noise. The vowels are fronted and lowered, with F1 and F2 frequencies both much higher. F3 higher or lower depending on the vowel.

M-k2's measurements point to a form of harsh voice, characterised by aperiodicity and noise. It has a rough rasping sound, low pitch and is slightly nasalised. Due to the extremely loud raspy sound, it is likely that the ventricular folds are vibrating as well. For the other two participants (S-k1, T-k1) it is more likely that the growling sound is due to vibration of the aryepiglottic folds. They also sound more whispery, which concurs with increased noise measurements.

As discussed in chapter 2.2, the ventricular voice is distinguished by a boost in the relative amplitude of the higher harmonics (Esling et al., 2019). The study does not go into an in-depth analysis of the spectral tilt, but it is an additional parameter that can contribute to an accurate analysis. A superficial look at the spectrum of M-k2 indicates the presence of strong higher harmonics.

Harsh voice with aryepiglottic trill

Participant T♂ describes his k1 character as a "dragon, troll, orc", with a "deep growly voice, breathy, false vocal cords instead of the regular". The results show that it does have a lower pitch, below even the minimum of the sweeping tone. It is very noisy, with an increase in jitter and to a lesser extend shimmer too. With the exception of /o/ it seems the F2 range has been minimised so the F2 of most vowels is centralised at around 2000 Hz. The F1 range on the other hand has increased a lot, from /a/ at about 700 Hz in the normal voice to /a/ at about 1000 Hz in character voice k1. The F3 frequency is higher for all but one of the vowels.

Participant S♀ describes her k1 character as "evil, cunning, animal, fearless, dominating, mean monster, male" with the following vocal traits: "lowered pitch, hoarse, distortion, deep/dark voice". The observations include low pitch, slightly more jitter, more shimmer, and a lot more noise. F1 and F2 frequencies are both raised, indicating laryngeal constriction, but possibly without tongue retraction (which would result in a lower F2 instead). The low growly quality indicates a sort of whispery harsh voice with a possible aryepiglottic trill. The character description features a lot of villainous qualities, which according to Teshigawara (2003) should correlate with not just a high F1, but also a low F2. These results do not quite match, but perhaps it is not completely fair to compare human villains to non-human monsters.

Creaky voice

Participant E♀ describes her k1 as a furry small alien that is scary but with a soft voice. The corresponding vocal traits she describes as “very deep, dark, with a lot of resonance and overdrive. This matches with the measurements that indicate creaky voice, featuring a lower F0, but not the tense laryngeal settings used to produce harsh voice. The overdrive she mentions also matches with the increased noise and aperiodicity in the signal.

5.3.2 Character voices with higher pitch - tense & falsetto

Participant S♀ describes her k2 character as: “puppy, bossbaby, impatient, whiny, childlike, genderless or girl”. The vocal traits are: “bright, shrill, squeaky, more whiny Stockholmish”. The description implies that the F0 frequency would increase, which it does with no less than 17 semitones. “Whiny, childlike” would suggest an increased pitch range, but it actually stays the same if we look at the relative pitch range in semitones.

Participant E♀ describes her k2 as a “cute, happy, small, excited animal crossing character”. The vocal traits she describes as: “high voice, singing intonations, tense larynx, not round, straight voice, more stress on intonations, sharper”. The F0 is clearly higher, which matches her description. “Happy, excited, singing intonations” imply an increased pitch range, which corresponds with the measurements (36 semitone range).

Participant T♂’s k2 character description is a “smaller, goblin, mosquito”. The vocal traits include: “buzzing, forward placement, higher pitch, nasal”. Although there is a 7 semitone increase in pitch (similar to E-k2), the median pitch is still well below 200 Hz and not likely the result of falsetto. There is also less jitter and shimmer, and a lot less noise, which could mean his normal voice is more creaky and breathy, whereas the k2 voice is cleaner in terms of periodic vibration, possibly due to higher subglottal pressure, elongated vocal folds, and a tenser laryngeal setting. The F2 range is slightly wider compared to the normal voice. This is unexpected because “forward placement” would suggest the F2 frequency would be higher instead. It seems he is using an articulatory setting that evens out the F1 range, probably a raised larynx to raise the F1 frequency, combined with tongue retraction to keep the cavity relatively small. He does this while retaining the F2 range, meaning that tongue movement is maintaining a clear distinction between front and back vowels.

5.3.3 Villainous characteristics

More stereotypically villainous qualities, such as “manic, evil, cunning, dominating, mean” were used to describe the characters with some form of harsh voice. The formant frequency measurements for these voices were also relatively high, which is indicative of laryngeal constriction. This matches with Teshigawara’s findings of villain characters displaying non-neutral states of the articulatory system (Teshigawara, 2003).

The more relaxed creaky voice (E-k1) on the other hand, featured a lower F1 and F2. The description for this character included “frightening, not that evil”, which concurs again with what Teshigawara’s findings suggest.

In terms of formants, the laryngealisation effects in Teshigawara’s study are described as having a high F1 and a low F2, while the results or the current study show raising of both F1 and F2 frequencies. This inconsistency might be due to the additional articulatory processes that raise F2 frequencies and make the voices sound less ‘human’, whereas the voices studied by Teshigawara, villainous as they may be, are not attempting to sound like anything more than human.

5.4 Limitations & future studies

Measurements

Pitch measurements were done for the humming tone, and also included the mean and median measurements of the speech samples. There were often inconsistencies between the humming tone and the median pitch, and even cases in which the humming pitch could not be properly measured due to irregular phonation. This issue was resolved by consistently using the median pitch measurements instead.

The F1 and F2 measurements for the vowels showed some anomalies, for example the /o/ in T-k1’s voice, which had an unexpectedly low F2 frequency in a system with an otherwise high and centralised F2 range. Voices that featured a high level of inharmonic frequencies and irregular phonation proved difficult to measure, which might have caused anomalous results.

The vowel system featured some inconsistencies, namely the /o/ and /ɛ/ vowels that featured varying degrees of centralisation. The vowels were taken from an ecologically valid sentence so

the environment of the vowels is not controlled. The /o/ vowel (å in också) was often centralised. It occurs again in *ångrade* but the following nasal interferes with the vowel quality there as well. F3 measures were also done but in many cases inconsistencies made it difficult to draw significant conclusions from them.

Additional parameters

As briefly mentioned in the previous section (5.3), there is no in-depth analysis of the spectral tilt. This is an additional parameter that could be explored further, something Teshigawara (2003) also demonstrates.

Temporal aspects, melody, and pitch contour were not included in the study, they can however contribute to a more detailed analysis. Melody and intonation for example, these were actually mentioned by some participants as aspects they altered for a specific character (E-k2, S-k2).

Relevance & future studies

The study revolves around voice acting and so the results are also most relevant for voice actors and voice coaches. It helps to understand what happens acoustically when different phonation types are employed. This can also contribute to speech synthesis and voice editing. For example plugin designers that work with voice effects.

The concept of monster is very culturally dependent. In this study was consciously restricted to Swedish participants, but in a future study it would be interesting to see if voice actors with different cultural and linguistic backgrounds conceptualise what constitutes a monster voice in a different way.

Chapter 6 Summary & conclusion

6.1 Back to the research questions

Answering the operationalised research questions

1. Pitch & phonation type

> 1.1 Compared to the average pitch in one's normal voice (in speech), how much higher or lower is the average pitch for the non-human characters? (change in semitones)

For the character voices that were lower in pitch, the pitch was lowered about -8 semitones on average. For the character voices that were higher in pitch, the pitch was raised about +12 semitones. The biggest increase in pitch was for one of the male participants (participant M with +18 semitones) who used ingressive phonation for his k1 voice. If we regard only egressive phonation the biggest increase was participant S♀ with a 17 semitone increase in a high-pitched tense falsetto voice for her k2. Her other character had the biggest decrease in pitch with -12 semitones.

> 1.2 Compared to the average pitch variation in one's normal voice, is there smaller or larger pitch variation in the non-human characters?

Generally the participants had an increased pitch range for the character voices, but there was one exception where the participant spoke more monotone for one of the characters (T-k1).

> 1.3 In the non-human characters, is the pitch closer to the normal pitch (humming tone) or to the minimum and maximum pitch of their voice? (compared to min & max F0)

Pitch ranges varied, for the characters with deeper voices the participants all went below the minimum pitch of their modal voice, showing that creaky and harsh voice can extend the vocal range below the range of normal phonation. For the higher pitch characters the speakers raised their pitch to a relatively higher degree. However, none of them came close to the max F0 they demonstrated in the sweeping tone recording. In short, they stayed closer to the natural pitch

when raising their F0, but were closer to their minimum pitch when they lowered their F0 for one of their character voices.

1.4 Do participants use different phonation types, and can these be measured in terms of jitter, shimmer and HNR?

Without looking at the formant frequencies it is difficult to judge based solely on irregular phonation and noise whether the speaker uses creaky voice or harsh voice as both of them can come with a lower pitch, and increased jitter, shimmer and HNR. The measurements can give a good indication as to which type of phonation is being used, but it is difficult drawing doubtless conclusions from them. As the laryngeal articulator model shows, the articulatory system is intricate with many different components that can all affect phonation in different ways.

2. What happens to vowels when dubbing a non-human/monster character that still speaks human language?

2.1 Are vowels centralised? (do they become less distinct, which could affect comprehensibility)

For the character voices that feature some form of laryngeal constriction and a more fronted tongue position we can see that the F1 and F2 ranges become smaller. There is also instances of extreme centralisation in one of the first two formants, but in those cases the other formant featured an expanded range instead, as if to make up for the lack of distinction.

2.2 Is the whole vowel system shifted? (is it stretched or are vowels individually dislocated)

Firstly, it is important to note that there does not necessarily have to be a shift. The formant frequencies can remain relatively stable even if the phonation type and F0 pitch are completely changed. The more central vowels like /ɛ/ and /ø/ did seem to shift individually in some cases but peripheral vowels like /i/, /u/ and /a/ generally remained on the relative edge of the vowel system. The vowel system was often shifted as a whole, with both F1 and F2 being raised (M-k2, S-k1, T-k1, E-k2), or lowered (E-k1). Most interestingly, the vowel system could be squished and stretched along either the F1 or the F2 axis (T-k1, T-k2).

2.3 Do vowel formants shift along with raised-larynx voice, or other forms of laryngeal constriction?

According to the literature the raised-larynx comes with tongue retraction, resulting in a raised F1 but a lowered F2 (Esling et al., 2019; Teshigawara, 2003). This effect was largely absent in the results, when the F1 was raised the F2 was raised as well, or slightly centralised in one case. An explanation could be that there is laryngeal constriction, but a lack of tongue retraction.

Perhaps one of the most important results we found regarding vowel shifting is that different formants can be centralised without necessarily affect the others. Especially if we look at participant T's k2 voice (figure 5.3) we can see that even if the F1 values feature nearly no distinction between individual vowels, the F2 range can conversely be expanded, retaining the vowel distinctions.

> 3. Do the different speakers use a similar approach, or is there for example a systematic difference between male and female voice actors when dubbing monsters/aliens/non-human characters?

The male and female participants show obvious differences in F0 pitch and vowel formant frequencies. However, in relative terms it was one of the female participants who was able to lower her pitch a relatively greater amount than any of the male participants. The female participants were the only ones to employ falsetto. Whether this is a systematic difference is difficult to say with such a small sample size, nor is it the case that the male participants would not be able to produce a falsetto, as they demonstrated it during the recording of the sweeping tone.

4. What do these vocal techniques mean for the character?

> 4.1 What personality traits do participants think of when dubbing their characters? (evil, mean, friendly, old/young, etc.)

There were only four participants so I will not attempt to make any generalisations. Something to keep in mind is that all of them have a Swedish cultural background, and people with other cultural backgrounds might interpret a non-human monster voice very differently. The participants were each given two options for different monsters and although they had the freedom to choose, they would usually start with a deeper, lower voice and do a higher one for the other monster. The monsters that featured a deep harsh voice (M-k2, S-k1, T-k1) were described as "selfish, mischievous, manic, evil, cunning, fearless, dominating, mean, male, abnormal, animal, dragon, troll, orc". The monsters that featured tensed up high pitch voices (E-k2, S-k2, T-k2) were described as "cute, happy, rascal, small, excited, puppy, impatient, whiny,

childlike, genderless or girl, goblin, mosquito". The two outliers featured creaky voice (E-k1), described as "furry small alien, frightening, not that evil, soft voice, scary" and ingressive phonation (M-k1), described as "chaotic neutral, evil, damned soul, scales, purple, dark red details, cat-eyed".

4.2 What vocal properties do participants think they are using to portray the personality? (raspy, harsh, high/low pitch, creak, etc) note that terminology vary between researchers and performers.

There were varied results, as some participants were more knowledgeable than other and used more scientific terminology when describing the vocal characteristics of their character voices. Participant T described his k1 voice as a deep growly voice for which he uses his false vocal cords instead of the regular vocal folds. Technically speaking the deep growl he describes is more probably a result of the combined vibration of the vocal folds and the aryepiglottic folds, but terminology tends to differ even in academia and both the ventricular folds (or false vocal cords) and the aryepiglottic folds can be used to create a lower rougher pitch.

4.3 What acoustic properties are we actually able to measure?

For the different types of harsh voice (M-k2, S-k1, T-k1) we could generally see an increase in F1 and F2 frequency and a decrease in F0 pitch of anywhere between -7 and -12 semitones. Especially shimmer and a noise measurements were much higher compared to the normal voice. For E-k1's creaky voice a strong increase in jitter was found, along with lower F1 and F2 frequencies and a lower F0 pitch. These are the some of the acoustic properties that are indicative of phonation type and tell us to what degree the voice quality is altered.

4.4 How do the acoustic measurements relate to the properties described by the participants in 4.2?

The participants used a variety of different adjectives and references to describe the properties of the characters they came up with. When they mentioned "deep, dark, lowered pitch" the pitch measurements reflected a decrease in F0. The same goes for the character with a higher pitch, these were generally described as "bright, small, childlike, cute, high voice".

There were other correlations as well, descriptions like "happy, excited" showed an increased pitch range (E-k2), while "deep, growly" correlated with measurements indicating harsh voice with aryepiglottic fold vibration (T-k1).

There were also some attributes that did not seem to match the measurements. A character described with “buzzing” did not show an increase in noise for example (T-k2), nor did “whiny” come with an increased pitch range (S-k2) as one might expect. Lastly there were descriptive words that were generally difficult to connect to any sort measurement in specific.

> 4.5 To what extend do the personality traits and vocal properties described by the participants match with what the literature suggests? (laryngeal constriction for villains for example)

With harsh voice, at a low pitch, the aryepiglottic folds can start vibrating, resulting in a growl (Esling et al., 2019). The measurements indicate the use of this technique for three of the character voices (M-k2 S-k1 T-k1) two of which are aptly described as “growling, or deep growly voice” in the vocal trait description. The other participant used “lowered pitch, hoarse, distortion, deep/dark voice ” in their vocal trait description, which is pretty accurate and matches with our findings based on the measurements and literature.

More stereotypically villainous qualities, such as “manic, evil, cunning, dominating, mean,” were given to describe the characters with some form of low pitched harsh voice. The high formant frequency measurements for these voices are indicative of laryngeal constriction. This matches with Teshigawara’s findings of villain characters displaying non-neutral states of the articulatory system. The more relaxed creaky voice (E-k1), which features a lower F1 and F2, was described as “frightening, not that evil” which again concurs with what Teshigawara’s findings suggest.

6.2 Conclusion

Back to the main research question

> *In what way do voice actors change their voice when dubbing a non-human character?*

In terms of voice quality, changes in pitch, phonation type, and vowel quality were found. However, there were instances where not necessarily all of these parameters were altered. In M-k1 for example, a big shift in pitch and phonation type was measured, but the vowel system remained relatively unchanged. A popular phonation type for monster voices seems to be a form

of harsh voice, as there were three voices that exhibited some version of harsh voice, with ventricular or aryepiglottic vibration to create a low pitch. There was another instance where the pitch was much lower compared to the normal voice but this was more likely the result of creaky voice.

Interestingly, the F0 for these lower voices was actually lower than the pitch range measurements of their modal and falsetto voice indicated they could produce. This shows that certain types of creaky and harsh voice can extend the pitch range below the limits of the modal voice.

In case of a higher pitch, participants generally had a larger pitch shift compared to their normal voice. However, the upper limit of their pitch range was much higher still.

The vowel system can shift, sometimes as a whole, sometimes compressed in either the F1 or the F2 frequencies. There were also individual vowel shifts but they occurred mostly within the already central vowels and could be interpreted as anomalies.

The results of this study demonstrate the capabilities of voice actors to alter their natural voice to create vocal performances that portray non-human characters with a multitude of different personality traits. It demonstrates in which way and to which degree the acoustically measurable parameters of voice quality are shifted.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburg U.P.

Akita, K. (2021). Phonation Types Matter in Sound Symbolism. *COGNITIVE SCIENCE, 45*(5). https://doi.org/10.1111/cogs.12982

DeBoer, A. R. (2012). Ingressive Phonation in Contemporary Vocal Music. *Doctor of Musical Arts Dissertations.* 16. Bowling Green State University.

Eklund, R. (2008). Pulmonic ingressive phonation: Diachronic and synchronic characteristics, distribution and function in animal and human sound production and in human speech. *Journal of the International Phonetic Association*, *38*(3), 235–324.

Engstrand, O. (2007). *Fonetik light : [lajt] : kortfattad ljudlära för språkstudier och uttalsundervisning* (första upplagan). Studentlitteratur AB, Lund.

Esling, J.H., Moisik, S.R., Benner, A. and Crevier-Buchman, L. (2019). *Voice Quality. The Laryngeal Articulator Model*. Cambridge: University Press.

Ewald, O., Asu, E. L., & Schötz, S. (2017). The formant dynamics of long close vowels in three varieties of Swedish. In F. Lacerda (chair) (Ed.), *Interspeech 2017* (pp. 1412–1416). (Interspeech). https://doi.org/10.21437/Interspeech.2017-1134

Fornhammar, L., Sundberg, J., Fuchs, M., & Pieper, L. (2022). Measuring Voice Effects of Vibrato-Free and Ingressive Singing: A Study of Phonation Threshold Pressures. *Journal of Voice*, *36*(4), 479–486. https://doi.org/10.1016/j.jvoice.2020.07.023

Garellek, M. (2019). The phonetics of voice. In Katz, W. F. & Assmann, P. F. (Eds.), *The Routledge handbook of phonetics* (pp. 75-106). London and New York: Routledge Taylor and Francis.

Gordon, M., & Ladefoged, P. (2001). Phonation Types: A Cross-Linguistic Overview. *Journal of Phonetics, 29*(4), 383–406. https://doi.org/10.1006/jpho.2001.0147

Katz, W. F., & Assmann, P. F. (2019). *The Routledge handbook of phonetics*. Routledge, Taylor & Fancis Group.

Keating, P. & Garellek, M. & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. *18th International Congress of Phonetic Sciences, ICPhS 2015, Glasgow, UK, August 10-14, 2015*. University of Glasgow.

Kreiman, J & Sidtis, D. (2011). *Foundations of Voice Studies. An Interdisciplinary Approach to Voice Production and Perception.* Wiley-Blackwell.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge University Press.

Laver, J. (1994). *Principles of phonetics*. Cambridge University Press.

McAllister, R. (1998). *Talkommunikation* (2. uppl.). Studentlitteratur, Lund.

Orlikoff, R. F., Baken, R. J., & Kraus, D. H. (1997). Acoustic and physiologic characteristics of inspiratory phonation. *Journal of the Acoustical Society of America*, *102*(3), 1838–1845. https://doi.org/10.1121/1.420090

Teshigawara, M. (2003). *Voices in Japanese animation: a phonetic study of vocal stereotypes of heroes and villains in Japanese culture.* University of Victoria, British Columbia.

Appendix

This section includes additional tables and data for reference. Tables 3.1 to 3.4 comprise formant value measurements for F1, F2, and F3.

Table 3.1 Formant values participant E ♀

E - normal	F1	F2	F3
i	464	2235	2958
ä	712	1622	3124
o	477	711	2640
e	492	2352	2860
å	523	1206	2859
a	818	1316	2637
ö	564	1694	2668
E - k1	**F1**	**F2**	**F3**
i	373	1818	2493
ä	498	1540	2592
o	354	673	2885
e	389	1877	2656
å	432	900	2588
a	651	1103	2587
ö	500	1248	2436
E - k2	**F1**	**F2**	**F3**
i	661	2013	3216
ä	635	1407	2114
o	597	1153	1684
e	668	2363	3101
å	724	1451	3519
a	909	1244	1832
ö	792	1747	1993

Table 3.2 Formant values participant M♂

M - normal	F1	F2	F3
i	374	1839	2411
ä	606	1353	2837
o	429	765	2903
e	449	1941	2593
å	466	1195	2796
a	632	1198	2688
ö	505	1455	2246
M - k1	**F1**	**F2**	**F3**
i	323	1923	2745
ä	520	1460	2621
o	375	706	3121
e	373	2069	2650
å	448	1112	2777
a	554	1246	2659
ö	479	1486	2680
M - k2	**F1**	**F2**	**F3**
i	528	2147	2749
ä	820	1937	2591
o	699	1178	2829
e	547	2343	2551
å	767	1399	2869
a	1018	1372	2596
ö	740	1725	2313

Table 3.3 Formant values participant S ♀

S - normal	F1	F2	F3
i	441	2344	2566
ä	822	1498	2542
o	438	970	3170
e	510	2111	2672
å	505	1449	2928
a	794	1408	2686
ö	623	1572	2781
S - k1	**F1**	**F2**	**F3**
i	508	2550	3164
ä	833	1953	3217
o	513	1095	2941
e	511	2490	3104
å	575	1522	2919
a	1003	1713	2874
ö	732	1635	2905
S - k2	**F1**	**F2**	**F3**
i	495	2313	2931
ä	785	1769	2345
o	536	1082	2762
e	538	2249	2666
å	723	1577	2719
a	903	1651	2317
ö	695	1928	2322

Table 3.4 Formant values participant T♂

T - normal	F1	F2	F3
i	347	2121	2619
ä	520	1625	2776
o	468	829	2776
e	452	2148	2559
å	471	1309	2432
a	655	1186	2738
ö	548	1403	2728
T - k1	**F1**	**F2**	**F3**
i	275	2004	3092
ä	670	1913	2890
o	536	1829	3231
e	453	2038	2661
å	474	1158	2363
a	1004	2052	3188
ö	904	1936	3046
T - k2	**F1**	**F2**	**F3**
i	537	2239	2627
ä	557	1723	2445
o	522	880	2013
e	543	2459	2809
å	584	1366	2284
a	632	1299	2512
ö	586	1606	2430

The list below contains the character descriptions and corresponding vocal traits as provided by the participants. The participants answered using descriptive words in either English or Swedish:

participant E♀ - k1:
- character description: furry small alien, frightening, not that evil, mjukröst, inte vassen, horigt, scary.
- vocal traits: overdrive, lot of resonance, very deep, mörk, högljud.

participant E♀ - k2:
- character description: animal crossing, non-human but speaking language, gulligt, glad, filur, small, excited, happy.
- vocal traits: högröst, sjungande intonationer, spänd struphuvud, inte rund, rak röst, more skånsk intonation/göteborg change dialect. more stress on intonations, sharper.

participant M♂ - k1:
- character description: chaotic neutral, evil, damned soul, scales, purple, dark red details, cat-eyed.
- vocal traits: strained, lot of air, raspy, agonised (plågad), inhaled.

participant M♂ - k2:
- character description: selfish, mischievous personality, manic.
- vocal traits: loud, raspy, stressed (melody), strained. In Swedish: raspig, ansträngd, morrande.

participant S♀ - k1:
- character traits: evil, cunning, djur, fearless, dominating, mean monster, male, abnormal, stranger things monster.
- vocal traits: lowered pitch, hoarse, distortion, mörk röst, no accent change just focussed on voice.

participant S♀ - k2:
- character description: hundvälp, bossbaby, otåliga, gnällig, barnslig, könslös eller tjej.
- vocal traits: ljust, pippigt, squeaky, gnälligt. more winey stockholmska.

participant T♂ - k1:
- character description: dragon, troll, orc.
- vocal traits: deep growly voice, breathy false vocal cords instead of the regular.

participant T♂ - k2:
- character description: smaller, goblin, mosquito.
- vocal traits: gnisslande (buzzing), forward placement, higher pitch, nasal.