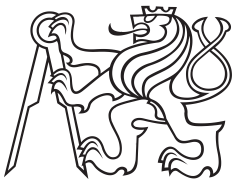
Master Thesis



Czech Technical University in Prague

F3

Faculty of Electrical Engineering
Department of Circuid Theory

Acoustic Analysis of Psychological Stress in a Speech Signal

Bc. Jana Zázvorková

Supervisor: prof. Ing. Roman Čmejla, CSc.
Field of study: Medical Electronics and Bioinformatics
Subfield: Signal Processing
May 2022

MASTER‘S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Zázvorková Jana** Personal ID number: **474400**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Circuit Theory**

Study program: **Medical Electronics and Bioinformatics**

Specialisation: **Signal processing**

II. Master’s thesis details

Master’s thesis title in English:

Acoustic Analysis of Psychological Stress in a Speech Signal

Master’s thesis title in Czech:

Akustická analýza psychického stresu v řečovém signálu

Guidelines:

• Conduct a research on the detection of psychological stress in the speech signal in order to select appropriate acoustic parameters.
• Check the selected acoustic parameters on the speech signals of phobic people taken at the Faculty of Science, Charles University.
• Perform the analysis in Praat and MATLAB environments. Also consider using the parameters obtained automatically [3] in the DYSAN toolbox.
• Evaluate the possibilities of psychological stress detecting from speech.

Bibliography / sources:

[1] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis and M. Tsiknakis, "Review on psychological stress detection using biosignals," in IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2019.2927337.
[2] H. Kurniawan, A. V. Maslov and M. Pechenizkiy, "Stress detection from speech and Galvanic Skin Response signals," Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, 2013, pp. 209-214, doi: 10.1109/CBMS.2013.6627790.
[3] J. Hlavnička, "Automated analysis of speech disorders in neurodegenerative diseases". Ph.D. thesis, CTU FEE, 2019, https://dspace.cvut.cz/handle/10467/79223?show=full

Name and workplace of master’s thesis supervisor:

prof. Ing. Roman Čmejla, CSc. Department of Circuit Theory FEE

Name and workplace of second master’s thesis supervisor or consultant:

Date of master’s thesis assignment: **26.01.2022** Deadline for master's thesis submission: _____

Assignment valid until: **30.09.2023**

prof. Ing. Roman Čmejla, CSc.	doc. Ing. Radoslav Bortel, Ph.D.	prof. Mgr. Petr Páta, Ph.D.
Supervisor’s signature	Head of department’s signature	Dean’s signature

III. Assignment receipt

The student acknowledges that the master’s thesis is an individual work. The student must produce her thesis without the assistance of others, with the exception of provided consultations. Within the master’s thesis, the author must state the names of consultants and include a list of references.

. _____ _____

Date of assignment receipt Student’s signature

Acknowledgements

I would like to thank my supervisor prof. Ing. Roman Čmejla, CSc. for guidance and RNDr. Eva Landová, PhD. for allowing me to participate in the experiment, making it possible to collect analysed recordings. I would also like to thank doc. Ing. Jan Rusz, PhD. for guidance in parameter selection.

Declaration

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

I further declare that all recordings have been made with the consent of the ethics committee and the consent of participants of the experiment.

Prague, 13. May 2022

Abstract

This master's thesis deals with the acoustic analysis of psychological stress in a speech signal. The effects of stress on human body, particularly on the respiratory system, serve as a foundation for research into the effect of stress on speech. Acoustic parameters, specifically the fundamental frequency f_0, are investigated in relation to stress exposure. The database was created from data collected as part of an experiment at the Faculty of Science at Charles University. The primary goals of this work are to create a stressed speech database, select appropriate acoustic speech features based on previous research, and analyse them. Despite the fact that this thesis is a pilot project, the results show a potential for detecting psychological stress in speech using acoustic parameters, especially f_0, and intensity. Preliminary results on vowel [a:] also show promising results for jitter, CPP and CPPs, and mean autocorrelation.

Keywords: speech, acoustic analysis, psychological stress, fundamental frequency, intensity, jitter, formants, CPP

Supervisor: prof. Ing. Roman Čmejla, CSc.

Abstrakt

Tato diplomová práce se zabývá akustickou analýzou psychického stresu v řečovém signálu. Projevy stresu v lidském těle, především v dýchací soustavě, slouží jako základ pro výzkum vlivu stresu na řeč. V této práci jsou zkoumány akustické parametry, převážně základní frekvence f_0, v závislosti na vystavení stresu. Použitá data pro analýzu se skládají z pavoučích a hadích fobiků pořízených v rámci experimentu na Přírodovědecké fakultě Univerzity Karlovy. Součástí práce je tvorba databáze nahrávek pořízených při psychickém stresu, vhodný výběr parametrů na základě rešerše předešlých prací a jejich akustická analýza. Přestože je tato práce formou pilotního projektu, výsledky ukazují potenciál pro detekci psychického stresu v řeči pomocí akustických parametrů, zejména f_0 a intenzity. Předběžné výsledky na samohlásce [a:] ukazují slibné výsledky pro jitter, CPP, CPPs a průměrnou autokorelaci.

Klíčová slova: řeč, akustická analýza, psychický stres, základní frekvence, intenzita, jitter, formanty, CPP

Překlad názvu: Akustická analýza psychického stresu v řečovém signálu

Contents

List of used abbreviations 1

1 Introduction 3
1.1 Motivation . . . 3

2 Theoretical Introduction 5
2.1 Speech production . . . 5
2.1.1 Source-filter Theory . . . 6
2.2 Acoustic speech characteristics . . 6
2.2.1 Fundamental frequency f_0 . . . 6
2.2.2 Formant frequencies . . . 7
2.2.3 Jitter and shimmer . . . 7
2.2.4 Intensity . . . 8
2.2.5 Speaking rate . . . 8
2.3 Psychological stress . . . 9
2.3.1 Definitional perspectives of psychological stress . . . 9
2.3.2 Stress and speech . . . 10

3 Related Works 11
3.1 Stress and its effects on human body . . . 11
3.2 Stress and emotions in speech . . 12
3.2.1 Stress-induced physiological changes affecting speech production . . . 12
3.2.2 Challenges of obtaining proper data . . . 13
3.2.3 Analysed acoustic characteristics in stressed speech . 13
3.2.4 Limitations of past research . 14
3.2.5 Overview or previous research 14

4 Methodology 17
4.1 Experiment . . . 17
4.1.1 Behavioural Approach Test . . 18
4.2 Data . . . 18
4.2.1 Recording environment and used recording devices . . . 18
4.2.2 Recordings . . . 19
4.2.3 Used protocol . . . 19
4.3 Praat . . . 20
4.4 Matlab . . . 20
4.5 DYSAN . . . 20
4.6 Data preprocessing and database creation . . . 20
4.7 Parameter analysis . . . 21
4.7.1 Mean fundamental frequency f_0 . . . 21
4.7.2 Standard deviation of f_0 . . . 22
4.7.3 Speaking rate . . . 22
4.7.4 Additional parameters' analysis using Praat . . . 23
4.7.5 Mean-energy intensity . . . 23
4.7.6 CPP and CPPS . . . 23

5 Results 25
5.1 Spontaneous speech . . . 25
5.1.1 Mean f_0 . . . 25
5.1.2 Standard deviation of f_0 . . . 26
5.2 Reading . . . 26
5.2.1 Speaking rate . . . 26
5.2.2 Mean f_0 . . . 27
5.2.3 Standard deviation of f_0 . . . 27
5.3 BAT by people . . . 28
5.3.1 Mean f_0 . . . 28
5.3.2 Mean f_0 by phobia category . 28
5.4 BAT by animals . . . 29
5.4.1 Mean f_0 . . . 29
5.4.2 Normalised mean f_0 . . . 30
5.5 Phobia and non-phobia stimuli results . . . 31
5.5.1 Mean f_0 . . . 31
5.6 BAT stages b/e . . . 31
5.6.1 Intensity for vowel [a:] . . . 32
5.6.2 Jitter for vowel [a:] . . . 32
5.6.3 Mean autocorrelation for vowel [a:] . . . 33
5.6.4 CPP and CPPS for vowel [a:] 33

6 Discussion 35
6.1 Spontaneous speech evaluation . 35
6.2 Reading task evaluation . . . 35
6.3 BAT evaluation . . . 36
6.3.1 Average phobia stimuli . . . 36
6.3.2 Additional parameters analysis for stages b/e . . . 37
6.4 Future research recommendation 38

7 Conclusion 39
Bibliography 41

A Detailed results 45

Figures

2.1 Scheme of human body systems used for speech production. 5
2.2 Source-filter theory scheme. 6
2.3 Formant frequencies example. . . . 7
2.4 Jitter and shimmer example 8

5.1 Mean f_0 results comparison of BAT snake phobia animal set between a spider and a snake phobic. 28
5.2 Spider BAT results of mean f_0 . 29
5.3 Normalised mean f_0 results for a spider. 30
5.4 Normalised mean f_0 results for a snake. 30
5.5 Boxplot comparison of mean f_0 for stimuli and control group. 31
5.6 Intensity results comparison for phobia and control group. 32
5.7 Jitter results comparison for phobia and control group. 32
5.8 Mean autocorrelation results comparison for phobia and control group. 33
5.9 CPPS results comparison for phobia and control group. 33

6.1 Comparison of average phobia and average control stimuli. 37

A.2 Overview of complete BAT results for PH3 46
A.1 Overview of complete BAT results for PH2 46
A.3 Mean f_0 results comparison of BAT spider phobia animal set between a spider and a snake phobic. 47
A.4 Snake BAT results of mean f_0. . 47
A.5 Shimmer results comparison for phobia and control group. 48
A.6 CPP results comparison for phobia and control group. 48

Tables

3.1 Related works and their results overview 15

4.1 Overview of animals used in BAT and their labels in the dataset. . . . 21
4.2 Overview of BAT stages and their labels in the dataset. 21

5.1 Results of mean f_0 for spontaneous speech 25
5.2 Results of standard deviation of f_0 spontaneous speech 26
5.3 Results of speech rate for the reading task 27
5.4 Results of mean f_0 for the reading task 27
5.5 Overview of mean f_0 per animal phobia group. 29

6.1 Normalised mean f_0 for average stimuli and control observations. . . 37

A.1 Standard deviation of f_0 for the reading task 45

List of used abbreviations

BAT - Behavioural Approach Test
SUSAS - Speech Under Simulated and Actual Stress
SR - Speaking rate
AR - Articulatory rate
f_0 - Fundamental frequency
F_1 - First formant
F_2 - Second formant
I - Sound Intensity
MFCCs - Mel-cepstral coefficients
CPP - Cepstral peak prominence
CPPS - Cepstral peak prominence smooth
XC - Control group identifier
PH - Phobia groupd identifier

Chapter 1

Introduction

This master's thesis focuses on the acoustic analysis of psychological stress in a speech signal. Since there aren't many available datasets, especially of Czech speakers, part of this work consists of collecting stressed speech recordings and creating a new database.
The first part of the thesis is dedicated to the theoretical introduction, such as the concept of psychological stress, speech production, and acoustic characteristics of speech. The research of related works and studies is presented to introduce appropriate acoustic characteristics for further analysis. Chapter 4 describes the chosen methodology, the experiment used for collecting data, data preprocessing, and programs used for data analysis. Thereafter, the methods and implementation of chosen acoustic parameters analyses' are briefly described. Results are presented in Chapter 5 and further discussed in Chapter 6.

1.1 Motivation

Various areas of expertise, such as engineering, psychiatry, or medicine, have taken an interest in speech analysis and the influence of stress on speech. In the past research, getting appropriate data has proven to be challenging. This was one of the reasons why studies used either actors for laboratory experiments or pilot recordings from the aviation industry to represent real-life situations.
Although it was not the main objective of this thesis, stress detection can contribute to all sorts of automatic emotion recognition systems used in spoken language interfaces in human-computer interactions.

Chapter 2

Theoretical Introduction

This chapter focuses on the theoretical knowledge needed to study the effects of psychological stress in a speech signal. Apart from a brief introduction to speech production and various acoustic speech characteristics, the concept of stress and emotions and their influence on speech will be introduced.

2.1 Speech production

Various areas of the human body are involved in speech production. Speech production, such as making articulated sounds or words, begins in the brain with the formation of the desired message, followed by conceptualisation and formulation. This mental stage is followed by a physical one, beginning in the lungs, which are part of the respiratory system. The respiratory system consists of organs that enable breathing. The exhaled air stream from the lungs serves as the foundation for sound creation in the phonatory system, causing oscillations in the vocal chords in the larynx.

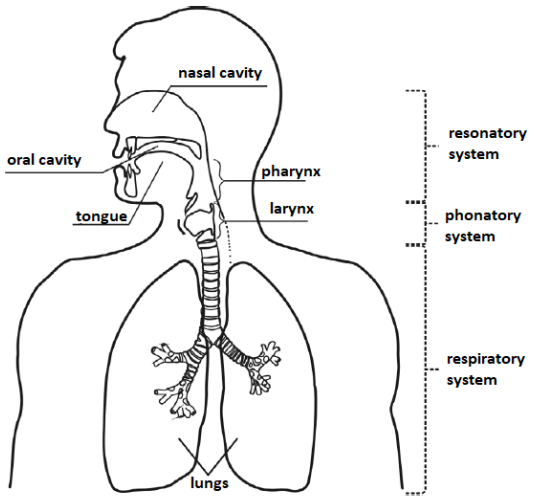


Figure 2.1: Scheme of human body systems used for speech production. Translated and taken from [31].

After passing through the larynx and vocal chords (also known as vocal folds), the air enters the nasal or oral cavity. There are two possible vocal fold positions: closed (no airflow) or open. The space between vocal chords is called the glottis. The final step in speech production is articulation, which occurs in the resonatory system, also known as the vocal tract. The oral cavity is responsible for the majority of the sounds that people can make. The positioning of the lips, teeth, and tongue shapes the sound. In real-life situations, articulatory movement deviations caused by emotional stress impact speakers' utterances. [31] [14]

2.1.1 Source-filter Theory

In the source-filter theory, introduced by Gunnar Fant, speech production is described as a two-stage process. As can be seen in Figure 2.2, first the sound is generated and then filtered or shaped by the vocal tract. [11]

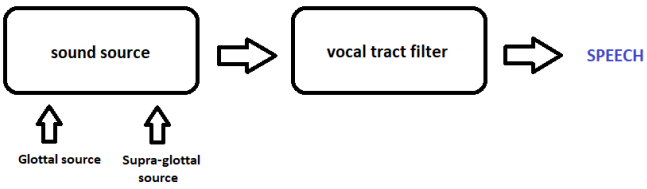


Figure 2.2: Source-filter theory scheme.

The filter in the vocal tract changes depending on whether the source is glottal or supra-glottal. For example, the fundamental frequency, also known as vocal pitch, is a characteristic of glottal source acoustics and vowel formants, both of which are the result of acoustic resonance of the vocal tract. [35]

2.2 Acoustic speech characteristics

The analysis of acoustic features of a speech signal is one of the main objectives of this thesis. The following sections describe acoustic characteristics used in stressed speech analysis and emotion recognition research, which can be divided into multiple categories - frequency parameters, amplitude parameters - such as intensity, spectral balance parameters, and temporal parameters. [24]
Based on research conducted in Chapter 3, the focus of this thesis is mostly on frequency parameters, in particular fundamental frequency, and temporal parameters, such as speaking rate.

2.2.1 Fundamental frequency f_0

Speech sounds are compound sounds that follow a (quasi-) periodic structure. Quasi-periodic behaviour means the signal shows irregular periodicity. In a speech signal, the approximate periodicity is expressed by the fundamental

frequency, usually denoted as f_0. It is determined by vocal folds oscillations during voiced speech and can be expressed in Hertz (Hz) or semitones (ST). The f_0 range for human speech is usually between 50 and 450 Hz, where males typically reach lower registers than females or children. [2]
When describing a voice, people use the term *pitch*, f_0 is the term for the physical phenomenon. During speech utterances, f_0 is not constant, so mean f_0 is used for acoustic analysis.

2.2.2 Formant frequencies

Closely related to fundamental frequency are formant frequencies. These parameters are observed in vowels and denoted by index numbers - mostly F_1 and F_2 are used in acoustic speech analysis of psychological stress. They are defined by local maxima in speech spectrograms, as can be seen in Figure 2.3. Formants characterise the resonance frequencies of vowels.

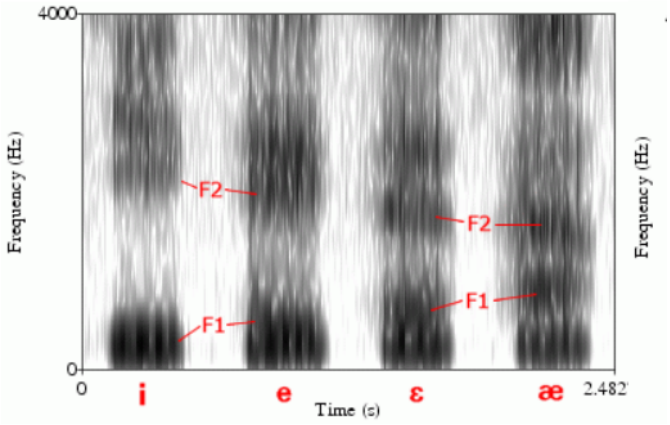


Figure 2.3: Formant frequencies example. Taken from [1].

2.2.3 Jitter and shimmer

Speech production is a process occurring in a biological system, the human body. Because such a system is made of soft tissue, its behaviour is not as precise as that of mechanical systems, resulting in the constant presence of perturbations in speech. Although all speech utterances contain some jitter and shimmer, they can be used to detect specific pathologies, with long-sustained vowels yielding the best results. [34] [2]
As shown in Figure 2.4, jitter represents frequency variation from cycle to cycle, whereas shimmer represents amplitude variation. Jitter can be measured in two ways: absolute, which examines consecutive periods, and relative, which examines the average period.

[1] `https://home.cc.umanitoba.ca/~krussll/phonetics/acoustic/spectrogram-sounds.html`

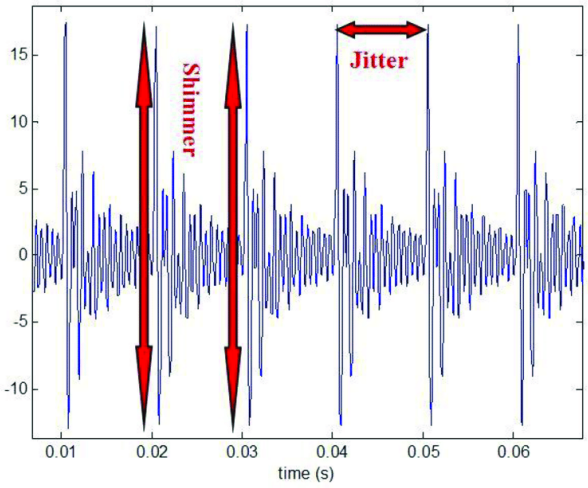


Figure 2.4: Jitter and shimmer example. Taken from [34].

2.2.4 Intensity

Sound intensity I represents the power and loudness of the wave defined as power per unit area, see 2.1.

$$I = \frac{P}{A} \, [\mathrm{W} \cdot \mathrm{m}^{-2}] \tag{2.1}$$

In speech analysis, amplitude is used as the measurement of loudness - intensity level L, in decibel scale defined as follows:

$$L = 10 \log_{10}(\frac{I}{I_0}) \, [\mathrm{dB}] \tag{2.2}$$

where I represents observed intensity and I_0 is a reference value using threshold of hearing defined as $I_0 = 1 \cdot 10^{-11}$ [W$\cdot$ m^{-2}].

2.2.5 Speaking rate

Speaking rate (SR), also known as speech rate, expresses the number of units in a given amount of time. Syllables per second is a common unit in speech analysis because words per minute is not suitable for varying word lengths. Pauses in speech also have a significant impact on SR. Articulatory rate (AR), which excludes such pauses, is defined as the ratio of the duration of articulation periods to the duration of voiced periods. SR, on the other hand, includes pauses during speech, making it a more accessible parameter for minimally labelled data.

2.3 Psychological stress

One of the goals of this thesis is to investigate the impact of psychological stress on speech. In general, psychological stress is thought to have negative effects on health and well-being. Despite being a widely used term, there are challenges and ongoing debates about how to define psychological stress appropriately. One of the reasons for definition inconsistency is the fact that the concept of stress is essential in many areas of expertise and present in a wide range of conditions. As a result, a single definition cannot cover all instances of stress.
The word stress itself has many meanings - below are some examples provided by Collins English Dictionary [20]:

- special emphasis or significance attached to something
- mental, emotional, or physical strain or tension
- emphasis placed upon a syllable by pronouncing it more loudly than those that surround it

2.3.1 Definitional perspectives of psychological stress

The basic concept proposes that psychological stress sets off distressing emotions related to deviations in the biological system of the human body. There are three definitional perspectives for defining psychological stress arising from research traditions. First, there are field studies and clinical observations, followed by animal laboratory studies, and finally, psychological phenomena and sequelae associated with stimuli and environments, which combine the previous two approaches. [23]

- **Environmental perspective**

  The main objective of the environmental perspective is the fact that psychological stress derives from the confronted environment. In such a case, these external circumstances can be labelled as psychologically demanding and threatening.

- **Response perspective**

  In response perspective, psychological stress is an organism's complex of reactions to environmental change, coming from natural selection and Darwinism.

- **Environment-Organism transaction perspective**

  Due to the limitations of the previous perspectives, another definition has been introduced: psychological stress is a relationship between an organism and its environment over time. This definition allows for the inclusion of both environmental conditions and individuals' actions. [19]

2.3.2 Stress and speech

As described in the Section 2.3.1, stress is a reaction to any psychologically demanding and threatening situation, in other words, an adaptation to a stressful situation. The human body can manifest physiological symptoms such as headaches, dizziness, chest pain, stomach problems, and difficulty sleeping. The variety of physiological symptoms leads to the question of whether stress accompanied by strong emotions like fear, disgust, or anger impacts speech as well. Some occupations rely on voice communication, such as air traffic controllers, pilots, military, police, or emergency responders, especially in time-sensitive and life-threatening situations. It is advantageous for them to be able to assess the speaker's mental state and determine whether or not they are under stress.

Chapter 3

Related Works

This chapter serves as the basis for conducted analyses in this thesis. It provides an overview of previously published related works on the subject, including emotions' analysis and detection in speech signals, acoustic analyses of stressed speech, and the overall effects of stress on the human body. Based on the research, some acoustic characteristics were chosen to be investigated further.
Due to the pilot project nature of this thesis, it can also be used as an inspiration for future work, with multiple ideas and possible parameters left to explore.

3.1 Stress and its effects on human body

Numerous studies have examined the effects of stress on the human body. It is now known that stress affects multiple bodily systems and can have different symptoms and courses of manifestation. There is a difference between acute and chronic stress; the latter, experienced over a long period of time, can result in long-term problems and manifestations of other seemingly unrelated conditions. This section will provide a brief overview of possible affected areas and the effects of stress on them. [1] [8] [37]

- **Respiratory system**

  One of the affected areas is the respiratory system. Experiencing stress can cause trouble breathing, hyperventilation, or even panic attacks. It can also worsen pre-existing conditions like asthma or chronic obstructive pulmonary disease.

- **Cardiovascular system**

  One of the stress manifestations in the cardiovascular system is an increase in heart rate and stronger heart muscle contractions. These changes are communicated within the body with adrenaline and noradrenaline. If such a situation develops into chronic stress, it can increase the risk of other conditions such as hypertension, a stroke, or a heart attack.

- **Gastrointensinal system**

  One way stress affects the gastrointestinal system is through gut bacteria. It can also cause bloating, discomfort, and nausea, all of which can lead to later gut dysfunction. Loss of appetite or in contrast stress eating are common side effects of stress.

- **Nervous system**

  The nervous system is responsible for triggering stress responses in all areas of the human body. As mentioned above, a stress response is conducted via the sympathetic and parasympathetic nervous systems that control hormone production responsible for other manifestations. What is concerning is that chronic stress causes continuous activation of the stress response with negative effects on other bodily systems.

3.2 Stress and emotions in speech

Many researchers have found it compelling to study stress-induced changes in speech and emotions for a variety of reasons. What can we learn from detecting stress or emotions in human speech? Where would such knowledge be applicable? In general, any situation in which the emotional and psychological state of a subject must be monitored. For example, automatic lie detection.[12] The polygraph, or physiological lie detector, is a well-established interrogation tool that works on the principle of measuring multiple physiological markers such as blood pressure, pulse, or skin conductivity. However, it is not foolproof. There are well-known workarounds. Including a speech-based lie detection in polygraphs could potentially improve their accuracy.
Another reason for studying the effects of psychological stress and emotions on speech is emotion detection. [4] [9] Automatic emotion detection would allow for further improvements of artificial intelligence, chatbots and all voice-enabled human-computer interactions.

3.2.1 Stress-induced physiological changes affecting speech production

As mentioned in the Section 3.1, the physiological effects of stress on the human body are well documented. It affects the respiratory system, which is closely linked to the process of producing speech, as described in Section 2.1. Since muscles tense up in response to stress, there is reason to believe that stress has an impact on speech production. The larynx muscle activity directly influences speech, in particular the fundamental frequency. Physiological stress manifestations, such as shortness of breath, excessive dryness of the mouth, and disrupted motor control, could all affect the output waveform passing through the vocal tract. [12] [36]

3.2.2 Challenges of obtaining proper data

As mentioned in Chapter 1, one of the biggest challenges of stressed speech analysis is gathering a proper dataset. Previous studies used either real recordings, such as pilots prio a crash or during emergencies ([26], [25]), emergency calls ([10]), or laboratory settings with professional actors ([36]), and subjects performing tasks under time pressure ([15]). Real-life data is limited, and artificially induced stress in a laboratory is an experiment in and of itself; it is difficult to determine the level of stress when using actors to act out emotions; the accuracy of their performance and resemblance to real emotions introduces more doubts about data quality and, thus, analysis results.
The results of such data-based acoustic analyses have been conflicting, and the following sections will go into further details of such discoveries.

3.2.3 Analysed acoustic characteristics in stressed speech

Due to the difficulty of precisely defining stress, most papers are dedicated to stress detection but do not classify the extent of psychological stress. Based on known physiological effects of stress on the human body, as described in the Section 3.1, researchers predicted possible acoustic speech parameters that could be affected by stress. One of the consequences of psychological stress, increased muscle tension, formed the hypothesis that a rise in f_0 should be observed in speech under stress. This assumption is based on the fact that increased muscle tension is likely to result in vocal folds tension, hence the suspected change of f_0. There may also be irregular respiration patterns, which can affect jitter. [17]
Historically speaking, one of the most researched acoustic speech characteristic is the above-mentioned f_0. Past research focused primarily on f_0 contour ([15]), mean ([36], [29] [3]), range ([10], [27]) and variability. Several studies have reported an increase of mean f_0 under stress in various scenarios. From real-life situations, mostly from pilots ([26], [36]), to cognitive tasks under time pressure ([26]), and oral examinations of university students ([29]). On the other hand there are papers contradicting such results ([3], [15]), or even presenting a conclusion that *"there are no valid acoustic indicators of psychological stress"* [32].
Apart from f_0, another explored acoustic parameter is the formant frequency. Studies showed an increase in both f_1 and f_2, as well as changes in formant bandwidth. [29] [14]
Pitch parameters are not the sole objective of stressed speech analysis. Overall voice quality has been observed, where stress in speech manifested in a breathy and strained voice. Both [15] and [36] reported occurrences of voicing irregularities when speaking under stress. Speech rate analysis has provided divergent results, where the general tendency is an increase in speed under stress, but not all studies have agreed with such findings. [17][36][22]
Cepstral features have also been explored, such as mel-cepstral coefficients (MFCCs), mostly for speech-based emotion recognition systems or stressed

speech recognition systems combining MFCCs and neural networks. [18]
One of the reasons for extensive cepstral features research is the fact that, compared to spectral features, cepstral features are less correlated with one another as well as being reasonably detailed descriptions of the short-term spectrum. [28] Past research reported a significant difference in MFCCs in the case of stressed students participating in an oral examination. [29]

3.2.4 Limitations of past research

Results of past research are difficult to compare due to inconsistencies in defining stress or controlling if all subjects are, in fact, experiencing stress. For example, the cognitive load does not have to necessarily result in stressed speech. With such deficiencies, the results are bound to be contradictory. Some of the limitations of the overall problem stem from the presence of individual variation, which makes developing a methodological and theoretical framework for the effects of stress on speech research more difficult. [15] [26] [17]
Apart from stress affecting speech, emotions also influence acoustic speech properties. More works ([24], [9], [4], [3]) shifted their focus on such phenomena in hopes to identify certain emotions, separate them to fully reveal the impacts on speech.
Part of the previous research is dedicated to the comparison of the results between laboratory and real-life situations in order to determine whether the stress of a pilot crashing an aeroplane can be compared to an actor simulating such emotions. The research suggests that artificially induced stress can be comparable in terms of acoustic analysis to natural stress. [26] It should be noted that results show a smaller increase in mean f_0 for laboratory-induced stress.
Unfortunately, most of the analysed acoustic speech parameters resulted only in tendencies. One of the most promising correlates of stress is fundamental frequency f_0 - an increase in mean f_0 under stress. Shimmer has been shown to be unaffected. Findings related to jitter are contradictory and mostly labelled as statistically insignificant.

3.2.5 Overview or previous research

For easier orientation, an overview of researched works, parameters they examined and their brief results will be provided in the Table 3.1.

paper	data	parameters	results
G. Demenko [10]	emergency calls	shift in f_0 range of f_0	↑ in stress related to fear ↑ in stress related to anger or irritation
R. Ruiz [27]	pilots radio announcer	range f_0 median f_0	↑ after or during the crash ↑ after or during the crash
C. Williams & Stevens [36]	actors	speech rate f_0	↓ for simulated fear ↓ for fear than for anger no clear correlate for fear
R. Banse [3]	actors	mean energy speech rate f_0	↑ for fear, panic, anger ↑ for fear, anger ↓ for sad no confirmation of previous findings of ↑ for fear
J. J. Congleton [7]	crew members	f_0 jitter shimmer	the most important indicator of psychological stress in speech less statistically significant vocal indicator of stress statistically insignificant
L. A. Streeter [32]	telephone recordings	f_0	no valid acoustic indicators of psychological stress
M. H. L. Hecker [15]	tasks under time pressure	f_0 spectograms	stress influences the contour of f_0 stress affects the shape of frequency spectrum
D. L. Mayer [22]	pilots	f_0 speaking rate	↑ during stress statistically insignificant
RothKranz L. [25]	Stroop test	f_0 jitter	↑ under stress ↓ under stress
R. Ruiz [26]	Stroop test pilot crash	f_0 f_0	gradual ↑ ↑ during emergency
M. Sigmund [29]	exam stress	f_1, f_2 f_0 MFCCs	↑ under stress ↑ under stress statistical difference
J. Hansen [14]	SUSAS corpus	f_0 f_1, f_2 vowel duration	↑ under stress ↑ under stress ↑ under stress

Table 3.1: Overview of related works and their results. The main objectives of the mentioned works were analysis of stressed speech, analysis of emotional impact on speech and emotion detection in speech. This table provides a brief insight into what data was used, which parameters were explored and simple results interpretation. ↑ = an increase of given phenomenom, ↓ = decrease.

Chapter 4
Methodology

One of the goals of this thesis was to create a dataset that could be used to examine the influence of stress on acoustic speech characteristics. There are certain limitations in collecting data for studying stressed speech; one cannot control the stress levels and must be confident that stress has occurred for all subjects. With this in mind, an experiment working with phobias, conducted by the Faculty of Science at Charles University, was chosen for recordings. Apart from creating the dataset, an analysis of selected parameters based on research described in Chapter 3 was done.

4.1 Experiment

The experiment, which served as the source of stress in the recordings, was part of a project called **Aversive response to spiders and the underlying emotions** (registered as 19-07164S at the Czech Science foundation - GAČR) led by RNDr. Eva Landová, PhD. and National Institute of Mental Health. The goals of this project are to determine whether the stimuli affect the emotions, focusing primarily on fear and disgust elicited by spiders and snakes.
The project consists of several experiments, including fMRI, but only the relevant part will be described in this section. First, a group of potential phobics had to be gathered. These individuals filled out specific questionnaires on spider and snake phobia, which were then followed by an in-person interview to obtain one's history, such as when they first became afraid of spiders or snakes. After signing the informed consent form, subjects were asked to participate in the experiment to measure physiological correlates of emotions such as heart electric activity (ECG), respiratory rate, and galvanic skin response (GSR) while watching various presentations of spider and snake stimuli. The Behavioural Approach Test (see 4.1.1) is also included, as is sorting and evaluating images of spider and snake stimuli. [13]
Typically, the experiment is divided into two sessions, with the first session consisting of a presentation, BAT, and image sorting (approximately 90 minute duration), and the second session, 2 weeks later, consisting only of presentations (approximately 60 minute duration). Due to commuting and other factors, some participants completed the entire experiment in one day.

All recordings were made when participants completed the entire experiment in one day or only the portion that included the BAT. No recordings took place only for the second session.

4.1.1 Behavioural Approach Test

Behavioural Approach Test (BAT) is used to assess the level of fear and avoidance for specific situations, which are commonly associated with phobias. During such a test, subjects are encouraged to approach the feared situation until they are unable to progress any further, while their physical symptoms, such as heart rate, are monitored.
The main idea is that the subject enters the room, which has a covered terrarium with an animal in the back. The identity of the hidden animal is known ahead of time, and the subject progresses through various stages and can leave whenever they see fit. For example, a person with arachnophobia (hereinafter referred to as *spider phobia*) does not have to finish all the stages for ophidiophobia (further referred to as *snake phobia*) despite the fact that they are not afraid of snakes.
There were 6 animals used for BAT, and their order depended on one's phobia. Each phobia was assigned a set of 3 animals. For arachnophobia it was a beetle *(Pachnoda iskuulka)*, a cockroach *(Eublaberus distanti)*, and a spider *(Tliltocatl vagans)*. For snake phobia the set consisted of leopard gecko *(Eublepharis macularius)*, a skink *(Tiliqua gigas)*, and a snake *(Epicrates (cenchria) maurus)*. If, for example, a subject had a snake phobia, they started with the set for spider phobia and advanced to the snake phobia set. It must be noted that the results of people with spider phobia served as a control group for snake phobia and vice versa.
There were 8 stages in total: 4 duct-taped lines approaching the terrarium, uncovering the terrarium, opening it, touching the animal with the blunt end of a pencil, and finally, touching the animal with one's finger.

4.2 Data

The recordings have been collected from 10 participants of the experiment, 3 of them having spider phobia, 2 control subjects not having any phobias, and the remaining 5 having snake phobia. All subjects were females, with ages ranging from 20 to 44 years old.

4.2.1 Recording environment and used recording devices

Recordings took place at the Faculty of Science, Charles University, in a designated lab for experiments using TASCAM DR-40X Linear PCM Recorder [33] with an external headset microphone. The external headset microphone was used to maintain a constant distance from the microphone, although this proved challenging during BAT because of subjects' movement.

After every animal, the microphone was manually adjusted, if needed.
The room was closed to limit outside noise for spontaneous speech and the reading task. In order to accommodate subjects during BAT, the doors were sometimes opened, especially for the final phobia-inducing animal. Colleagues supervising the experiment guided the participants throughout the test, contaminating some recordings, which had to be excluded from the dataset.

4.2.2 Recordings

As mentioned above in 4.1, the experiment consists of different parts. The incorporation of recordings into the already established experiment was carefully considered in order not to disrupt the flow of the experiment by distracting participants with additional tasks. With this in mind, spontaneous speech and reading task recordings were made both at the beginning and at the end of the experiment. The starting point serves as the data representing the presence of stress in speech, whereas the final recording should show the normal state lacking psychological stress. During BAT, participants wore the recording equipment capturing the entire test.

4.2.3 Used protocol

For the spontaneous speech dataset, the participants were asked to paraphrase the Cinderella fairy tale for at least 90 seconds. In case they couldn't recall the story or wanted to remind themselves, a short overview, which they could read before the recording started, was prepared.
The reading task was based on a short text by Czech author Karel Čapek, which had previously been chosen for other speech-related experiments. You can see the full version below:

Když člověk poprvé vsadí do země sazeničku, chodí se na ni dívat třikrát denně: tak co, povyrostla už nebo ne? I tají dech, naklání se nad ní, přitlačí trochu půdu u jejich kořínků, načechrává jí lístky a vůbec ji obtěžuje různým konáním, které považuje za užitečnou péči. A když se sazenička přesto ujme a roste jako z vody, tu člověk žasne nad tímto divem přírody, má pocit čehosi jako zázraku a považuje to za jeden ze svých největších úspěchů. [6]

The subjects had to read out loud their heart rate shown on smart watch, which allowed us to incorporate speech recordings in BAT with a small modification without disrupting the experiment. The participants were asked to ideally say: "Moje aktuální tepová frekvence je XX úderů za minutu." *(translation "My current heart rate is XX beats per minute.")*. Due to apparent distress during this part of the experiment, subjects could change the sentence structure to their liking.

4.3 Praat

Praat [5] is a free software created by prof. dr. Paul Boersma and dr. David Weenik from the University of Amsterdam. Praat is used for speech analysis such as pitch, formant, or spectral analysis. In addition, it offers more advanced features, for example, principal component analysis (PCA) and feedforward neural networks.

4.4 Matlab

After initial data processing in Praat, most analyses were done in MATLAB. [21] MATLAB is a programming platform using matrix-based language designed by MathWorks specifically for engineers and scientists. Thanks to wide variety of toolboxes, MATLAB is a universal tool in many areas of expertise, including machine learning and neural networks.

4.5 DYSAN

DYSAN is a tool based on MATLAB used for automated analysis of speech disorders in neurodegenerative diseases. This tool allows for multiple acoustic analyses with different categories: rhythm test, connected speech, sustained vowels, and diadochokinetic test. [16]
When the main goals of this thesis were first formulated, there was an initial thought that DYSAN might be helpful, but it wasn't applicable in this case. The duration of BAT recordings was insufficient, and resulting analyses used in this work were implemented from scratch for easier modification.

4.6 Data preprocessing and database creation

As a part of the data preprocessing, the recordings have been divided into spontaneous speech, reading task, and BAT datasets. Using Praat, .wav files were created from the obtained raw recordings, while ensuring that no additional voices or avoidable noise were present. The resulting utterances were named in the following manner: *ID_+S+after/before+.wav*. IDs were either *PHxx* or *XCxx* depending whether one was a person with phobia or not. For example, *PH1_Sbefore.wav* is a spontaneous speech recording obtained at the start of the experiment for the first participant.
All mistakes, such as repeating words, were cut out from the reading recordings to maintain a set syllable count for further analysis. The utterances were named in the same manner as for the spontaneous speech dataset by replacing "S" to "R". Both versions are in the final dataset, cut out .wav files are labelled with "F" at the end. (For example *PH3_RbeforeF.wav*)
The BAT recordings were more time consuming to process and categorise. First, the animals were represented by numbers 1-6 (see Table 4.1) and stages with letters a-h (see Table 4.2). Depending on the subject's phobia, the order

of the animal is changed, which introduces a certain level of confusion when processing the full speech without any video recordings. The same goes for stages progression.

code number	animal
1	beetle
2	cockroach
3	spider
4	gecko
5	skink
6	snake

Table 4.1: Overview of animals used in BAT and their labels in the dataset.

stage name	stage description
a	starting line
b	advancing to the next line
c	advancing to the next line
d	advancing to the covered terrarium
e	uncovering the terrarium
f	opening the terrarium
g	touching the animal with a pencil
h	touching the animal with a finger

Table 4.2: Overview of BAT stages and their labels in the dataset.

4.7 Parameter analysis

Praat was used for preparing the recordings and generating .txt files with timestamps and fundamental frequency values in Hz for further analysis performed mostly in MATLAB. This process was fully done by pre-existing algorithms for obtaining pitch in Praat, in this case, autocorrelation has been used. Result interpretation, such as statistics and graphs were also implemented in MATLAB. During processing BAT recordings, a validation matrix A_{60x7} was created. The first column represents ID of recording, and the remaining columns represent stages a-h. (For example, PH5_B3d, stands for subject PH5, tarantula recording, stage d.)

4.7.1 Mean fundamental frequency f_0

Mean f_0 was calculated using the pre-generated .txt files in both Hz and ST. The mean value $\bar{f}_0$ is the average of given values per file and represents the central tendency of data. The following equation (4.1) defines mean as the sum of all observations f_{0i} (values of f_0 for i-th observation) divided by the total number of observations N. MATLAB's already existing function *mean()*

was used in the implementation.

$$\bar{f_0} = \frac{\sum_{i=1}^{N} f_{0i}}{N} \tag{4.1}$$

In order to calculate mean f_0 in ST, values in Hz had to be converted. For conversion following formula was used:

$$f_0 = 12 \cdot \log_2 \left(\frac{f_0^*}{f_{const}}\right) \text{[ST]} \tag{4.2}$$

where f_0^* is fundamental frequency in Hz, and f_{const} is a constant value, in this case 50 Hz was used. Mean f_0 [Hz] was primarily calculated to observe the effects described in previous studies as mentioned in Chapter 3. The conversion to semitones has been made in order to use the standard deviation of f_0 described in the section below (see 4.7.2) and to compare results of BAT for all subjects - each of them has unique values of f_0 and semitones were used to normalise the results for more straightforward graphical interpretation (see section 5.4.2).

4.7.2 Standard deviation of f_0

Standard deviation of data measures how a given set of values disperse in relation to their mean. After acquiring values of f_0 in section 4.7.1, standard deviation has been calculated accordingly:

$$std = \sqrt{\frac{\sum_{i=1}^{N} |f_{0i} - \bar{f_0}|^2}{N-1}} \tag{4.3}$$

where N is total number of observations, $\bar{f_0}$ is mean of given data, f_{0i} is the value of observation *i*. Normalisation by *N-1* is used in order to achieve a less biased estimate of standard deviation. In this case, MATLAB's already existing function *std()* has been used to get the resulting values.

4.7.3 Speaking rate

The reading task was the only part of the experiment that provided data appropriate for calculating the SR. Even with given text (for full text see section 4.2.3), subjects sometimes misspoke - these parts were cut out during the preprocessing described in 4.6. Due to limited time options, the recordings were not labelled, so the only way of calculating SR was to do it manually. First, the total number of syllables was counted. Then, each subject's individual starting and ending times were manually identified in Praat, and the total duration of reading was determined. Finally, SR was calculated using a simple formula as follows:

$$SR = \frac{total\_syll}{dur} \text{ [syll/s]} \tag{4.4}$$

where *dur* [s] is the time subjects took to read the text, *total_syll* is the number of total syllables in the text. Usually, *total_syll* equaled 151, but

there were a few occasions where the subject's mistakes could not be cut out of the speech.
Because each BAT and spontaneous speech recording is unique, labelling data or manually counting syllables would be time-consuming and out of scope of this thesis. Even if subjects were able to say the recommended sentence for BAT, the recordings would differ because they read their heart rate aloud.

4.7.4 Additional parameters' analysis using Praat

For vowel [a:] in BAT stages b and e, analyses of additional parameters were performed using Praat. Mean-energy intensity [dB], jitter [s], shimmer [dB], mean autocorrelation [-], F1 [Hz], and F2 [Hz] were analysed on manually selected duration of the vowel [a:]. Due to this thesis's pilot project nature, provided analysis relies on already implemented algorithms in Praat and serves as the first step for future research where more advanced approaches can be investigated further.

4.7.5 Mean-energy intensity

Already implemented algorithm in Praat for obtaining the mean intensity within specified time domain was used. The mean-energy intensity I_m [dB] is defined as follows:

$$I_m = 10 \log_{10} \left(\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} x(t) d(t) \right) \text{ [dB]} \tag{4.5}$$

where $x(t)$ is the intensity as function of time, $x(t) \in \langle t_1, t_2 \rangle$.

4.7.6 CPP and CPPS

Cepstral Peak Prominence (CPP) and Cepstral Peak Prominence Smooth (CPPS) are considered robust acoustic measures of voice quality. CPP has been studied in relation to specific voice disorders and dysphonia and was intended to measure breathiness and overall voice quality. MATLAB function for CPP and CPPS analysis was provided by the creators of the DYSAN toolbox. CPP and CPPS analysis should be performed only on voiced speech utterances, so a simple voiced speech detector was implemented in MATLAB. Based on the Praat generated .txt files containing f_0 values and timestamps, voiced speech intervals were identified and later used for the detector. In this thesis, CPP and CPPS are only used for preliminary analysis for future research.

Chapter 5

Results

The results of individual analyses are presented in this chapter. For an easier understanding, the chapter is divided into 6 parts. Results are presented separately for spontaneous speech, reading task, and BAT. BAT is further split into 4 sections. Firstly, the results of trends among individual participants are shown, then the focus shifts to phobia animals - a snake and a spider. In some cases, the snake control group consists of spider phobia and control subjects. The remaining two parts are based on Section 5.4. Section 5.5 shows results of phobia and control animal exposure. In Section 5.6, additional parameters are analysed for stages b and e. Interpretation of the following results are discussed in detail in Chapter 6.
Due to the large volume of graphs, some were moved to Appendix A. Absolute values of deciding thresholds for trend monotony were set to 2 Hz, 0.5 ST, and 0.1 syll/s, every occurrence under the given threshold was evaluated as insignificant.

5.1 Spontaneous speech

5.1.1 Mean f_0

ID	Mean f_0 [Hz]		
	before	after	trend
PH1	201.3	200.7	→
PH2	208.3	218.9	↘
PH3	176.7	186.8	↘
PH4	212.6	211.1	→
PH5	167.1	158.0	↗
PH6	202.2	208.7	↘
PH7	183.1	186.7	↘
PH8	180.6	175.7	↗
XC1	246.0	232.4	↗
XC2	216.2	221.7	↘

Table 5.1: Resulting values of mean f_0 [Hz] for spontaneous speech. The **before** column represents stressed speech recordings, the **after** column represents non-stressed speech. $\rightarrow$ = no change, $\searrow$ = decrease in comparison to non-stressed speech, $\nearrow$ = increase in comparison to non-stressed speech.

In the Table 5.1, the results overview of mean f_0 [Hz] analysis is shown.

5.1.2 Standard deviation of f_0

The Table 5.2 shows the results of f_0 standard deviation analysis.

ID	Standard deviation of f_0 [ST]		
	before	after	trend
PH1	3.47	3.48	$\rightarrow$
PH2	3.78	2.65	$\nearrow$
PH3	4.30	3.10	$\nearrow$
PH4	2.24	3.79	$\searrow$
PH5	2.85	5.22	$\searrow$
PH6	3.60	3.22	$\rightarrow$
PH7	4.04	4.09	$\rightarrow$
PH8	4.41	3.8	$\nearrow$
XC1	3.23	4.43	$\searrow$
XC2	2.94	2.41	$\nearrow$

Table 5.2: Resulting values of standard deviation of f_0 [ST] for spontaneous speech. The **before** column represents stressed speech recordings, the **after** column represents non-stressed speech. $\rightarrow$ = no change, $\searrow$ = decrease in comparison to non-stressed speech, $\nearrow$ = increase in comparison to non-stressed speech.

5.2 Reading

In the case of the reading recordings SR, mean f_0, and std of f_0 were analyzed.

5.2.1 Speaking rate

The Table 5.3 shows the results of speaking rate analysis.

ID	Speaking rate [syll/s]		
	before	after	tendency
PH1	4.9	5.1	↘
PH2	4.3	4.9	↘
PH3	4.6	4.8	↘
PH4	4.6	4.5	→
PH5	5.1	5.0	→
PH6	4.9	4.7	↗
PH7	4.5	4.6	→
PH8	4.7	4.6	→
XC1	5.0	5.1	→
XC2	4.9	5.3	↘

Table 5.3: Resulting values of SR [syll/s] for the reading task. The **before** column represents stressed speech recordings, the **after** column represents non-stressed speech. → = no change, ↘ = decrease of speed in comparison to non-stressed speech (slower speaking rate, less syllables per second), ↗ = increase of speed in comparison to non-stressed speech.

5.2.2 Mean f_0

The results of f_0 standard deviation analysis are in the Table 5.4.

ID	Mean f_0 [Hz]		
	before	after	difference
PH1	203.5	211.4	↘
PH2	206.0	213.8	↘
PH3	203.2	197.1	↗
PH4	213.9	209.2	↗
PH5	175.4	172.5	↗
PH6	210.1	220.4	↘
PH7	193.1	200.2	↘
PH8	199.5	198.8	→
XC1	239.1	242.1	↘
XC2	219.2	226.4	↘

Table 5.4: Resulting values of mean f_0 [Hz] for the reading task. The **before** column represents stressed speech recordings, the **after** column represents non-stressed speech. → = no change, ↘ = decrease in comparison to non-stressed speech, ↗ = increase in comparison to non-stressed speech.

5.2.3 Standard deviation of f_0

The results of f_0 standard deviation [ST] analysis are shown in the Table A.1.

5.3 BAT by people

For easier interpretations of BAT analysis results, additional identifier was added to participants' ID labelling their phobia group: -H for snake phobia and -P for spider phobia.

5.3.1 Mean f_0

To illustrate what the complete BAT results for all animals look like, a representative for snake phobia and spider phobia were picked, as can be seen in Figures A.1 and A.2. Because such graphs with all 6 animals can become confusing, other detailed graphs are provided further.

5.3.2 Mean f_0 by phobia category

An example of how the results differ in snake phobia set and spider phobia set between a snake phobic and a spider phobic can be seen in Figures A.3 and 5.1

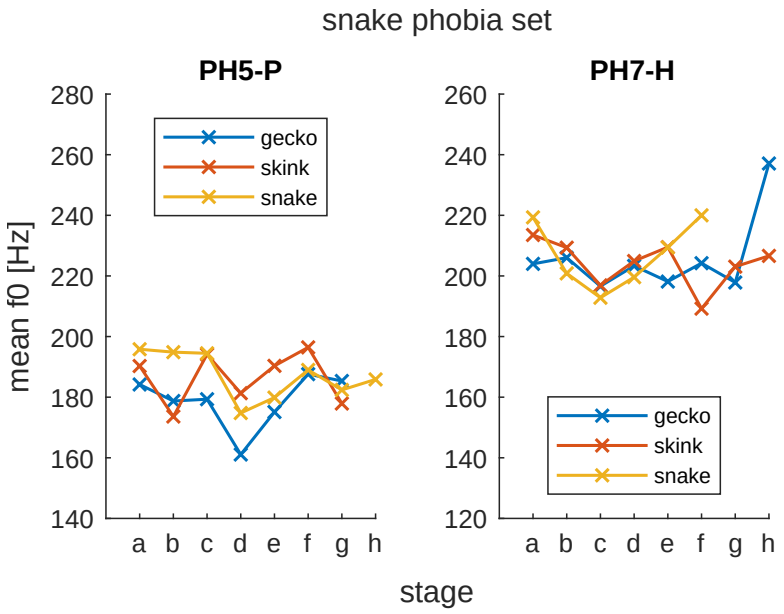


Figure 5.1: Mean f_0 [Hz] results comparison of BAT snake phobia animal set between a spider and a snake phobic. The left graph shows the results of a spider phobic and snake phobia animal set, in our case a control subject. The right graph shows the results of a snake phobic reacting to a spider phobia animal set, the stressed subject.

Table 5.5 shows an overview of mean f_0 per phobia set for all subjects and the difference between subject's phobia group and control group.

ID	Mean f_0 [ST] per phobia group		
	P_M	H_M	difference
PH1-H	25.49	25.78	0.29
PH2-P	26.94	26.76	0.18
PH3-H	23.71	24.11	0.4
PH4-H	26.44	26.47	0.03
PH5-P	23.17	23.09	0.08
PH6-H	27.00	26.92	-0.08
PH7-H	24.74	24.97	0.23
PH8-P	22.64	23.15	-0.51

Table 5.5: Overview of mean f_0 [ST] values for given animal phobia group and their difference. **P_M** = spider phobia animal group, **H_M** = snake phobia animal group. Resulting **difference** is calculated as individual's phobia group - control group.

5.4 BAT by animals

5.4.1 Mean f_0

The Figure 5.2 (Figure A.4) shows mean f_0 [Hz] of BAT results for a spider (snake) exposure.

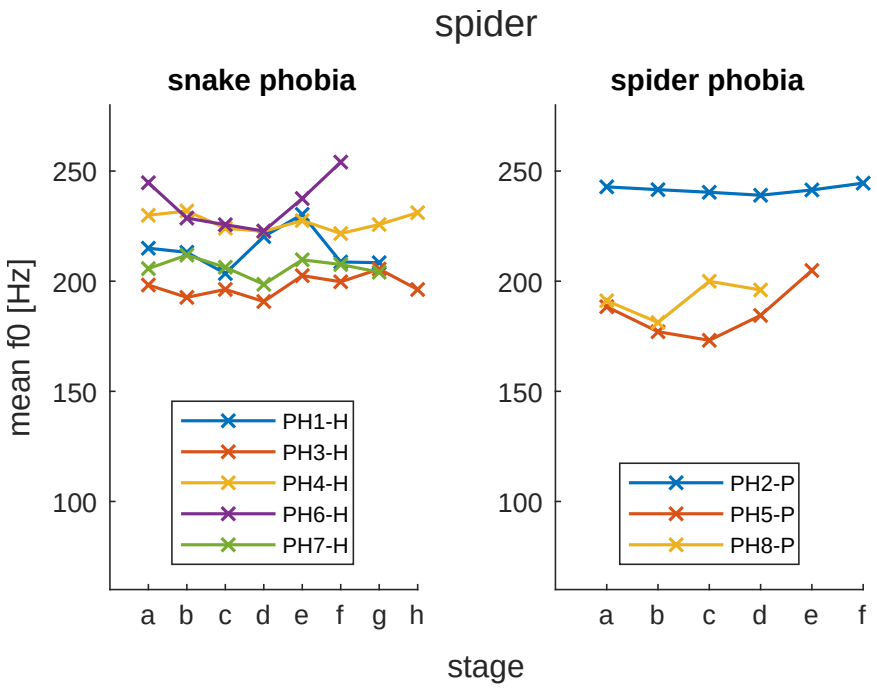


Figure 5.2: Spider BAT results of mean f_0 [Hz]. The left graph shows all snake phobia subjects' results of a spider exposure, in this case - a control group. The right graph shows all spider phobia subjects' results of a spider exposure - stressed speech.

5.4.2 Normalised mean f_0

Normalised mean f_0 [ST] results, where the values have been normalised by the stage b, are shown in the Figure 5.3 and Figure 5.4.

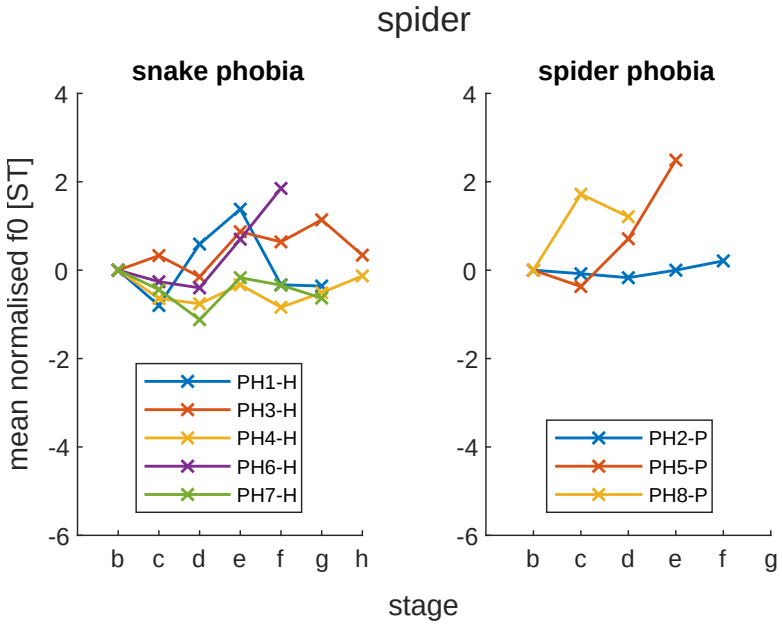


Figure 5.3: Normalised mean f_0 [ST] BAT results for a spider. The left graph shows all snake phobia subjects' results of a spider exposure, in this case - control group. The right graph shows all spider phobia subjects' results of a spider exposure - stressed speech.

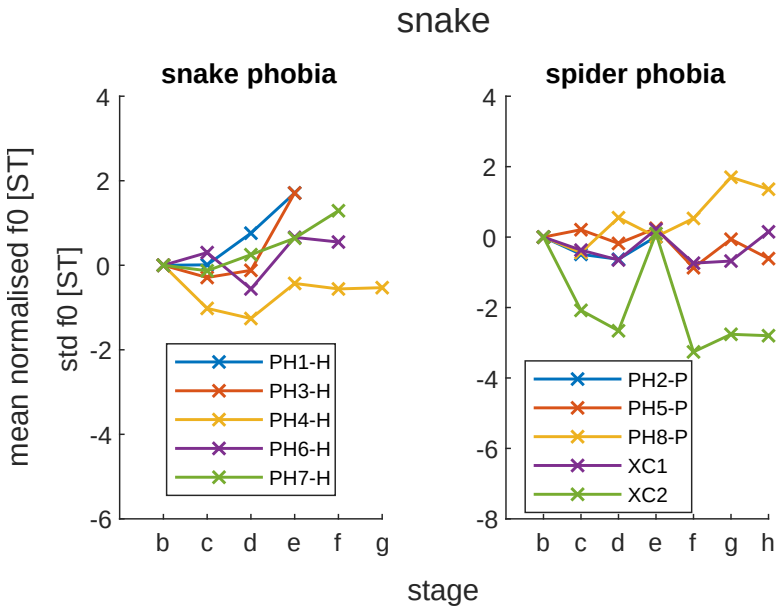


Figure 5.4: Normalised mean f_0 [ST] BAT results for a spider. The left graph shows all snake phobia subjects' results of a snake exposure - stressed speech. The right graph shows all spider phobia subjects' and control subjects' results of a snake exposure , in this case - control group.

5.5 Phobia and non-phobia stimuli results

5.5.1 Mean f_0

The resulting values of mean f_0 [ST] for phobia stimuli and non-phobia stimuli are shown in Figure 5.5 using boxplots. Each participant had their phobia-trigerring animal for stimuli results and the opposite for control (snake phobia subject = snake, control = spider). Phobia stimuli are denoted as S, represented by blue boxplots. Control stimuli are denoted as C, represented by red boxplots.

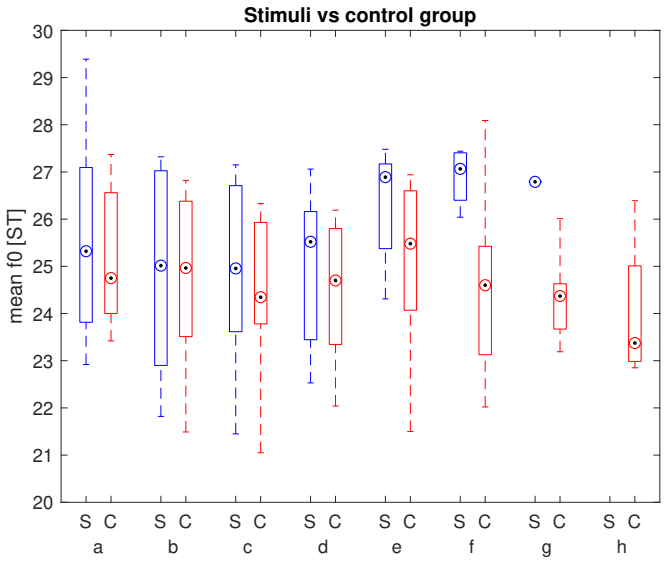


Figure 5.5: Boxplot comparison of phobia and control stimuli. Phobia stimuli are represented by blue boxplots. Control stimuli are represented by red boxplots. For phobia stimuli the results of phobia triggering animal exposure, snake (spider) for snake (spider) phobia, were used. For control stimuli the results of control animal exposure, spider for snake phobia subject and vice versa, were used. S = phobia group, C = control group.

5.6 BAT stages b/e

Following sections show results comparison between additional parameters' analyses between BAT stages b and e for vowel [a:] for the same stimuli and control group as in Section 5.5. Additional analysed parameters were intensity (Figure 5.6), jitter (Figure 5.7), shimmer (Figure A.5), mean autocorrelation (Figure 5.8), CPP (Figure A.6), and CPPS (Figure 5.9).

5.6.1 Intensity for vowel [a:]

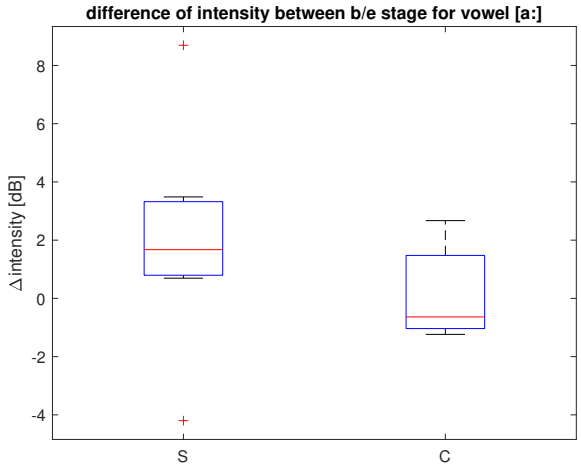


Figure 5.6: Boxplot comparison of the mean-energy intensity of vowel [a:] difference between stages b and e for phobia and control stimuli. For phobia stimuli the results of phobia triggering animal exposure, snake (spider) for snake (spider) phobia, were used. For control stimuli, the results of control animal exposure, spider for snake phobia subject and vice versa, were used. S = phobia group, C = control group.

5.6.2 Jitter for vowel [a:]

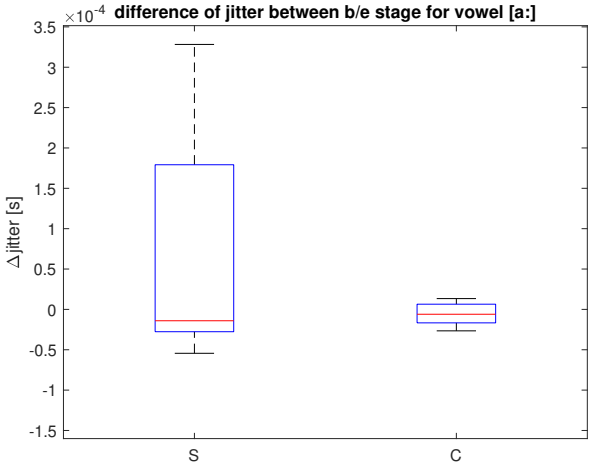


Figure 5.7: Boxplot comparison of the resulting jitter values of vowel [a:] difference between stages b and e for phobia and control stimuli. For phobia stimuli the results of phobia triggering animal exposure, snake (spider) for snake (spider) phobia, were used. For control stimuli, the results of control animal exposure, spider for snake phobia subject and vice versa, were used. S = phobia group, C = control group.

5.6.3 Mean autocorrelation for vowel [a:]

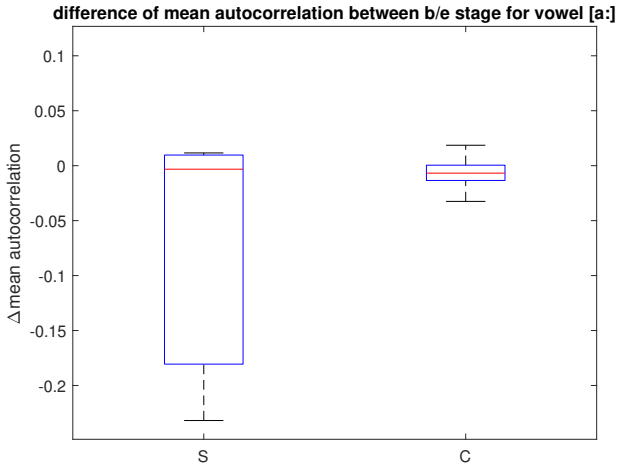


Figure 5.8: Boxplot comparison of the resulting shimmer values of vowel [a:] difference between stages b and e for phobia and control stimuli. For phobia stimuli the results of phobia triggering animal exposure, snake (spider) for snake (spider) phobia, were used. For control stimuli, the results of control animal exposure, spider for snake phobia subject and vice versa, were used. S = phobia group, C = control group.

5.6.4 CPP and CPPS for vowel [a:]

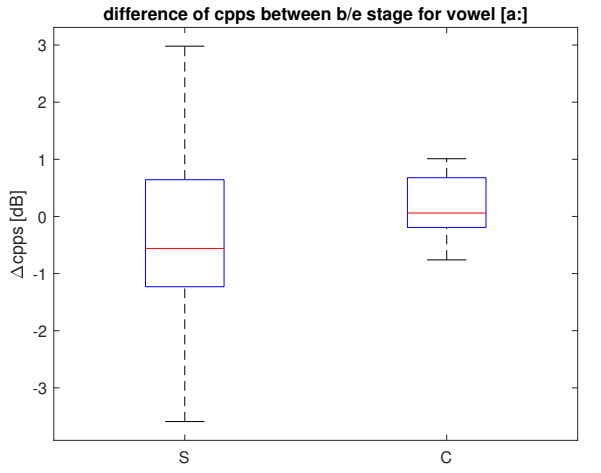


Figure 5.9: Boxplot comparison of the resulting CPPS values of vowel [a:] difference between stages b and e for phobia and control stimuli. For phobia stimuli the results of phobia triggering animal exposure, snake (spider) for snake (spider) phobia, were used. For control stimuli, the results of control animal exposure, spider for snake phobia subject and vice versa, were used. S = phobia group, C = control group.

Chapter 6

Discussion

This chapter focuses on the results interpretation reported in Chapter 5. The main goal of this master's thesis is to determine if there are any acoustic parameters indicating any influence of psychological stress on speech. Due to the nature of used experiment, as described in 4.1, a possible impact of emotions on speech must be taken into account. The database includes people with a severe form of either arachnophobia or ophidiophobia, so the source of psychological stress for all the analyses is the combination of extreme fear, disgust, and anxiety.
The main goal was not to differentiate between emotions, but to identify any promising correlates for stressed speech. The most promising results come from BAT recordings, in which the participant progresses through the experiment as far as possible before becoming overwhelmed. It must be noted that some subjects used the BAT portion of the experiment to overcome their fear and were able to suppress their reactions of how they would normally react.

6.1 Spontaneous speech evaluation

Spontaneous speech recordings were the least analysed in terms of used acoustic parameters. The results reported in Table 5.1 and Table 5.2 indicate that there is no trend in mean f_0 and std of f_0 values for stressed speech. In stressed speech analysis, it is important to control the stress present in participants if possible. These results indicate that spontaneous speech recordings taken before and after the experiment (1.5 - 3 hours time span) might not have the same value as direct BAT recordings.

6.2 Reading task evaluation

The reading task followed the spontaneous speech recordings. In addition to the already analysed acoustic parameters with the same conclusion of no observed trend (see Table 5.4, Table A.1), SR analysis has been performed, see Table 5.3. No clear trend has been observed, but it must be noted that speaking rate analysis has been done manually, which leaves more room for

human error, the same as the deciding threshold. The articulation rate and speaking rate could be calculated more precisely using labelled data.

6.3 BAT evaluation

The results of BAT recordings seem to be the most promising. More specifically, the results of mean f_0 [ST] and normalised mean f_0 [ST]. Barely any of the past studies used the semitone scale to illustrate results, making it harder to compare multiple participants.
Results reported in Section 5.3 are more of an illustrative nature to show the outcome of the complete experiment for a snake phobic and a spider phobic participant, wherein in both cases, we can see an increase of mean f_0 in the last stage for the phobic animal. Section 5.4 is more focused on the key animals - a snake and a spider. The graphs illustrate all of the phobic subjects' response and a new parameter normalised mean f_0 [ST] is used for easier results interpretation, see Figure 5.4 and Figure 5.3.
The tendency of an increase in mean normalised f_0 especially during the last stages in the graphs mentioned above, brought the idea of creating a phobic group, no matter which phobia, and a control group, as can be seen in Figure 5.5. Subsequently, additional parameter analyses for stages b and e were performed for vowel [a:] (Section 5.6).

6.3.1 Average phobia stimuli

As mentioned throughout the thesis, the main limitation of this database is the low number of samples. Normalised mean f_0 [ST] reported the most promising trend, so the participants were grouped by phobia (both snake and spider phobic subjects in one group) and their corresponding control stimuli. With this approach, we get 8 subjects for phobia and control group each. Unfortunately, individual observations are dependent on one another (each participant repeated the experiment for two animals), there are 8 categorical variables representing the stages, so these results are, again, not statistically evaluated.
An average course of mean f_0 [ST] for stimuli and control observations calculated from 8 observations can be seen in the Figure 6.1. For the average control animal (non-stressed speech), we can see an increase around stage e - revealing the animal, but overall the mean f_0 is decreasing. For average stimuli (stressed speech), we can see a much larger spike between the stages d-e but compared to non-stressed speech, the mean f_0 did not return to lower values. The values of average normalised mean f_0 [ST] of stimuli and control observations are in Table 6.1. It must be noted that for stimuli group, the more advanced stage, the less valid data there is as can be seen in Figure 5.5. The stage g would consist of only one sample for the stimuli group, and for this reason, it has been eliminated from the average stimuli calculation. Stage e consists of 7 samples and stage f of 4 samples. This does not change the significant jump between stage d-e, where stage f only supports the tendency.

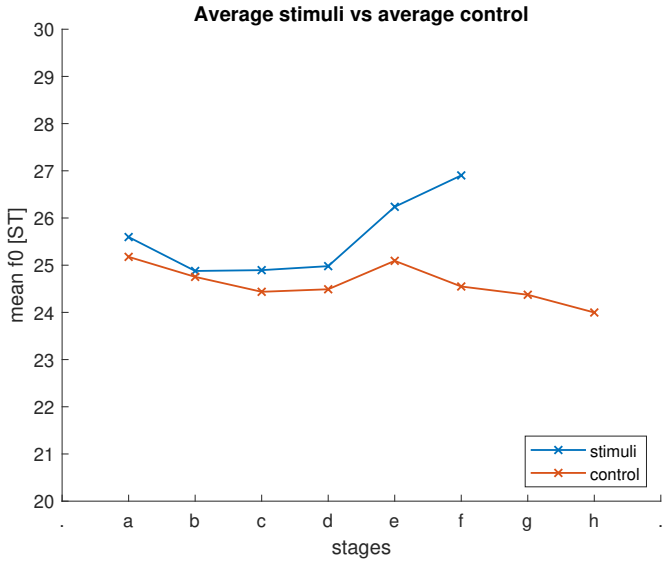


Figure 6.1: Comparison of average phobia and average control stimuli. Average phobia stimuli was calculated from the results of the phobia triggering animal exposure (snake phobia - snake). Average control stimuli was calculated from the results of the control animal exposure (spider for snake phobia subject and vice versa.)

subject	**Normalised mean f_0 [ST] at stage**						
	b	**c**	**d**	**e**	**f**	**g**	**h**
stimuli	0	0.02	0.10	1.36	2.03	-	-
control	0	-0.32	-0.26	0.34	-0.20	-0.38	-0.75

Table 6.1: Overview of normalised mean f_0 [ST] values for average stimuli and average control observations.

From the results of normalised mean f_0 in Table 6.1 we can see that in the case of average phobic stimuli, the difference between stage b and f is 2.03 ST, and the min-max range is also 2.03 ST. For average control observation, the difference between stages b and h is -0.75 ST, and the min-max range is also 0.75 ST. Overall, the trend of an increasing mean f_0 for stressed speech can be observed, especially from stage d for average stimuli observation.

6.3.2 Additional parameters analysis for stages b/e

Additional parameter analysis was performed on data from 7 phobic subjects on a vowel [a:]. Because the utterances of the last participant do not contain needed vowel [a:], the data was omitted. The results in Figure 5.8 and Figure 5.7 indicate that in the case of the control group (non-stressed speech) the changes in mean autocorrelation and jitter are smaller than for stressed speech. Mean-energy intensity tends to increase for stressed speech, see Figure 5.6. For this parameter, it is important to consider the possible effect of present emotions in speech.

CPP and CPPS show higher variability for stressed speech, as can be seen in Figure A.6 and Figure 5.9. The results of all additional analyses are statistically insignificant as expected due to a low number of data samples. Results of F_1 and F_2 analysis are inconclusive.

6.4 Future research recommendation

Research shows that past works reported divergent results but demonstrate potential for future works. In accordance with Chapter 3, most of the outcomes discussed in this chapter are only trends and not statistically significant results with the sample size being the main limitation. As a result, further statistical testing was not performed in order to avoid p-hacking and distorted claims. Additional recordings are still being collected but, due to time constraints, cannot be included in this work.
For future research on the acoustic analysis of psychological stress in speech, expanding the database introduced in this thesis is recommended. Further data processing, such as labelling mentioned in relation to SR analysis, could provide the foundation for a more in-depth analysis of, for example, articulation rate, formant frequencies, or spectral slope. Labelling would allow more vowel-centred analysis, not just [a:] as in the case of the above-mentioned additional parameters (Section 5.6).
Another promising acoustic parameter, not studied in the past research in relation to stress but Parkinson's disease, is CPP and CPPS [30]. CPP and CPPS could be suitable parameters for analysing psychological stress in speech. For proper CPP and CPPS analysis, it is needed to implement a voiced speech detector and, therefore, a robust f_0 detector - the simple approach used in this thesis using values of f_0 from Praat was selected due to the pilot project nature.
Aside from examining additional parameters, it would be interesting to see whether there are any differences in psychological stress in male and female speech. Because this database only includes female subjects, expanding it to include male phobic subjects could provide an interesting insight. Not to mention that the majority of real-life situation recordings come from the aviation industry, specifically pilots, which is a male-dominated field.

Chapter 7

Conclusion

The purpose of this master's thesis was to collect stressed speech recordings, create a database, preprocess the data, and perform acoustic analysis of promising speech parameters based on research. One of the objectives was to assess the possibility of detecting psychological stress through speech.
Because the experiment that allowed recording stressed speech was already in progress, only a subset of the total participants was included in the created dataset. The experiment focuses on aversive responses to spiders and snakes, with participants suffering from agoraphobia and ophidiophobia. Prior to any data preprocessing or analysis, research on stressed speech analysis and emotion detection was carried out. There are many contradictory reports due to the uncertainty in the definition of psychological stress and other complications such as not controlling whether subjects were, in fact, under stress.
As described in Chapter 3, previous research either used recordings of real-life situations, such as emergency calls or pilots prior to and during emergencies, or laboratory-induced stress with performing tasks under time pressure, or actors simulating certain emotions. Throughout all the past research, the most promising parameter was the fundamental frequency f_0, and that is why the main focus of performed analyses in this thesis lies in f_0 as the primary point of interest.
The recordings were manually preprocessed using Praat and divided into spontaneous speech, reading task, and BAT recordings. A .txt file containing pitch values was generated for each recording when creating individual utterances. These files were further processed in MATLAB, analysing the selected parameters - mean f_0, the standard deviation of f_0, and normalised mean f_0. In the case of the reading task, the speaking rate was calculated by extracting the starting and ending time manually from Praat.
Based on the results of mean f_0 for BAT recordings, additional parameters were analysed for stages b and e for vowel [a:]. Analysed parameters were mean-energy intensity, jitter, shimmer, mean autocorrelation, CPP, and CPPS. It must be noted that vowel [a:] was selected manually in Praat, and all the analyses were performed by either Praat or an already-existing function provided by the creators of the DYSAN toolbox.
Due to the small dataset size, the results of all analyses performed for

spontaneous speech and reading task recordings are inconclusive. The most promising recordings were taken during BAT, especially of the main phobic animals - a snake and a spider. In order to compare multiple participants, mean f_0 and normalised mean f_0 are both in semitones. A trend of an increasing mean f_0 has been observed after grouping the subjects into a phobic and a control group - the phobic group consisting of the results of the phobia-triggering animal exposure (snake for a snake phobic, spider for a spider phobic), and the control group consisting of the results of a control animal exposure (spider for a snake phobic, snake for a spider phobic). An average phobic and an average control observation were calculated from the phobic and control group confirming the tendency of an increased mean f_0 in stressed speech. One of the limitations of this approach is that participants may be scared of the control animal as well. Because the database is insufficient for proper statistical evaluation and for avoidance of p-hacking, all of the results of this work are simply observed tendencies that could be confirmed or rejected by expanding the created database.

In terms of future research on psychological stress in speech, additional data preprocessing, such as labelling, could be promising. This database can be used to study not only psychological stress in speech, but also emotion recognition and classification, specifically fear. It would be worthwhile to expand the database itself; part of the experiment is still ongoing, and additional recordings are being collected, but they are not included in this thesis due to time constraints. Further analysis of cepstral peak prominence, intensity, mean autocorrelation and jitter in other vowels and additional acoustic analysis of other parameters, such as formant frequencies, articulatory rate, vowel duration, spectral slope and speech rate from labelled data, could also provide insight into psychological stress in speech.

Bibliography

[1] Stress effects on the body, 2018. https://www.apa.org/topics/stress/body.

[2] Bäckström, T., Räsänen, O., Zewoudie, A., Pérez Zarazaga, P., and Koivusalo, L. Introduction to speech processing. https://wiki.aalto.fi/display/ITSP/Introduction+to+Speech+Processing.

[3] Banse, R., and Scherer, K. R. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology 70*, 3 (1996), 614–636.

[4] Bhimavarapu, J. P., Sarvana, K., Achanta, V. K. S., Kadiyala, C., and Yadhavkareti, C. Modelling of emotion recognition system from speech using MFCC features. In *ESSENCE OF MATHEMATICS IN ENGINEERING APPLICATIONS: EMEA-2020* (2021), AIP Publishing.

[5] Boersma, P., and Weenink, D. Praat: doing phonetics by computer [Computer program]. Version 6.1.10, 2022. http://www.praat.org/.

[6] Čapek, K. *Spisy. Od člověka k člověku 3.* Českosloven. Spisovatel, Praha, 1991.

[7] Congleton, J. J., Jones, W. A., Shiflett, S. G., McSweeney, K. P., and Huchingson, R. D. An evaluation of voice stress analysis techniques in a simulated AWACS environment. *International Journal of Speech Technology 2*, 1 (May 1997), 61–69.

[8] Cooper, C. *Handbook of stress medicine and health.* CRC Press, Boca Raton, 2005.

[9] Davletcharova, A., Sugathan, S., Abraham, B., and James, A. P. Detection and analysis of emotion from speech signals. *Procedia Computer Science 58* (2015), 91–96.

[10] Demenko, G., and Jastrzębska, M. Analysis of natural speech under stress. *Acta Physica Polonica A 121*, 1A (Jan. 2012), A–92–A–95.

[11] Fant, G. *Acoustic theory of speech production with calculations based on X-ray studies of Russian articulations*. Mouton, s'Gravenhage, 1960.

[12] Gadallah, M., Matar, M., and Algezawi, A. Speech based automatic lie detection. In *Proceedings of the Sixteenth National Radio Science Conference. NRSC'99 (IEEE Cat. No.99EX249)*, Ain Shams Univ.

[13] GAČR. *Návrh standardního projektu 19-07164S*.

[14] Hansen, J. H. L., and Patil, S. Speech under stress: Analysis, modeling and recognition. In *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 108–137.

[15] Hecker, M. H. L., Stevens, K. N., von Bismarck, G., and Williams, C. E. Manifestations of task-induced stress in the acoustic speech signal. *The Journal of the Acoustical Society of America 44*, 4 (Oct. 1968), 993–1001.

[16] Hlavnička, J. *Automated analysis of speech disorders in neurodegenerative diseases*. PhD thesis, Czech Technical University in Prague, 2018.

[17] Kirchhübel, C., Howard, D. M., and Stedmon, A. W. Acoustic correlates of speech when under stress: Research, methods and future directions. *International Journal of Speech Language and the Law 18*, 1 (Sept. 2011).

[18] Kumar, N., Thangjam, S., and Kaur, H. Robust stress speech recognition using mfcc and neural network.

[19] Lazarus, R. *Stress, appraisal, and coping*. Springer Publishing Company, New York, 1984.

[20] LLC, H. P. Collins english dictionary, 2018. `https://www.collinsdictionary.com/dictionary/english`.

[21] MATLAB. *version 9.5.0 (R2018b)*. The MathWorks Inc., Natick, Massachusetts, 2018.

[22] Mayer, D. L., Brenner, M., and Cash, J. R. Development of a speech analysis protocol for accident investigation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 38*, 1 (Oct. 1994), 124–127.

[23] Monroe, S. M., and Cummins, L. F. Stress: Psychological perspectives. In *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier, 2015, pp. 583–587.

[24] Nordström, H. *Emotional Communication in the Human Voice*. Department of Psychology, Stockholm University, Stockholm, 2019.

[25] Rothkrantz, L. J. M., Wiggers, P., van Wees, J.-W. A., and van Vark, R. J. Voice stress analysis. In *Text, Speech and Dialogue*. Springer Berlin Heidelberg, 2004, pp. 449–456.

[26] Ruiz, R., Absil, E., Harmegnies, B., Legros, C., and Poch, D. Time- and spectrum-related variabilities in stressed speech under laboratory and real conditions. *Speech Communication 20*, 1-2 (Nov. 1996), 111–129.

[27] Ruiz, R., Legros, C., and Guell, A. Voice analysis to predict the psychological or physical state of a speaker. *Aviation, space, and environmental medicine 61*, 3 (1990), 266–271.

[28] Sethu, V., Epps, J., and Ambikairajah, E. Speech based emotion recognition. In *Speech and Audio Processing for Coding, Enhancement and Recognition*. Springer New York, Sept. 2014, pp. 197–228.

[29] Sigmund, M. Introducing the database ExamStress for speech under stress. In *Proceedings of the 7th Nordic Signal Processing Symposium - NORSIG 2006* (June 2006), IEEE.

[30] Šimek, M., and Rusz, J. Validation of cepstral peak prominence in assessing early voice changes of parkinson's disease: Effect of speaking task and ambient noise. *The Journal of the Acoustical Society of America 150*, 6 (Dec. 2021), 4522–4533.

[31] Skarnitzl, R., Šturm, P., and Volín, J. *Zvuková báze řečové komunikace*, první ed. nakladatelství Karolinum, Praha, 2016.

[32] Streeter, L. A., Macdonald, N. H., Apple, W., Krauss, R. M., and Galotti, K. M. Acoustic and perceptual indicators of emotional stress. *The Journal of the Acoustical Society of America 73*, 4 (Apr. 1983), 1354–1360.

[33] TASCAM. *Linear PCM Recorder*. `https://tascam.com/downloads/products/tascam/dr-40x/e_dr-40x_rm_vd.pdf`.

[34] Teixeira, J. P., Oliveira, C., and Lopes, C. Vocal acoustic analysis – jitter, shimmer and HNR parameters. *Procedia Technology 9* (2013), 1112–1122.

[35] University, M. Acoustics. `https://www.mq.edu.au/about/about-the-university/our-faculties/medicine-and-health-sciences/departments-and-centres/department-of-linguistics/our-research/phonetics-and-phonology/speech/acoustics`.

[36] Williams, C. E., and Stevens, K. N. Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America 52*, 4B (Oct. 1972), 1238–1250.

[37] Yaribeygi, H., Panahi, Y., Sahraei, H., Johnston, T. P., and Sahebkar, A. The impact of stress on body function: a review. *EXCLI Journal; 16:Doc1057; ISSN 1611-2156* (2017).

Appendix A

Detailed results

ID	Standard deviation of f_0 [ST]		
	before	**after**	**trend**
PH1	3.22	3.32	$\rightarrow$
PH2	3.91	4.12	$\rightarrow$
PH3	4.12	4.34	$\rightarrow$
PH4	2.58	3.47	$\searrow$
PH5	2.14	3.76	$\searrow$
PH6	3.33	2.45	$\nearrow$
PH7	3.21	2.00	$\nearrow$
PH8	4.18	3.97	$\rightarrow$
XC1	4.34	4.37	$\rightarrow$
XC2	3.12	2.59	$\nearrow$

Table A.1: Resulting values of standard deviation of f_0 [ST] for the reading task. The **before** column represents stressed speech recordings, the **after** column represents non-stressed speech. $\rightarrow$ = no change, $\searrow$ = decrease in comparison to non-stressed speech, $\nearrow$ = increase in comparison to non-stressed speech.

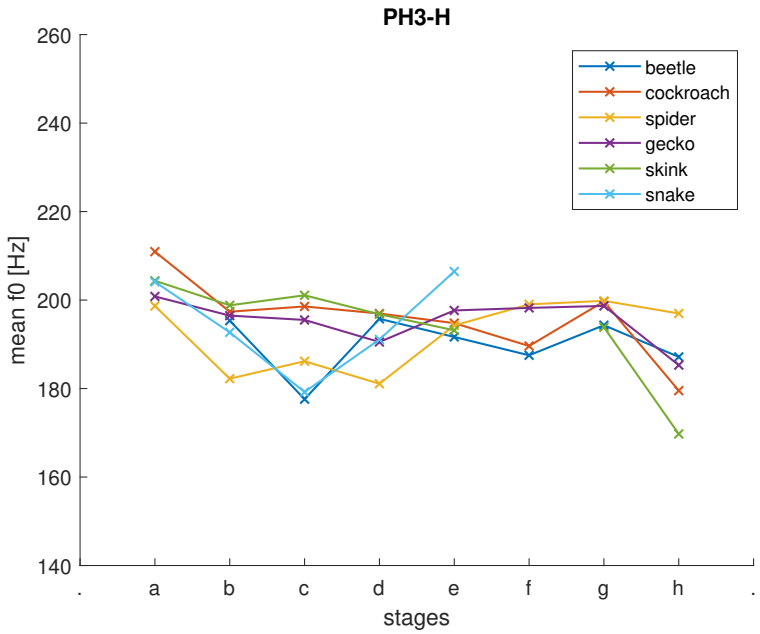


Figure A.2: Overview of complete BAT results for a specific snake phobic - PH3.

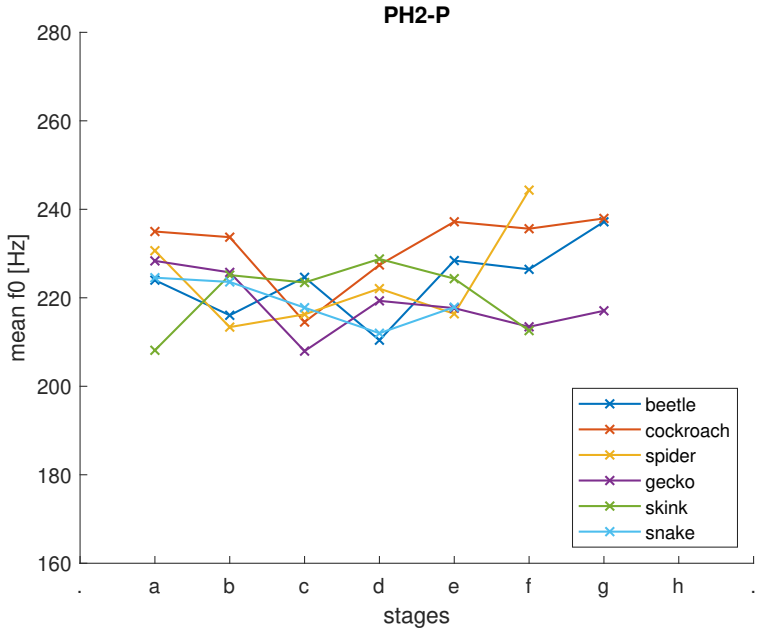


Figure A.1: Overview of complete BAT results for a specific spider phobic - PH2.

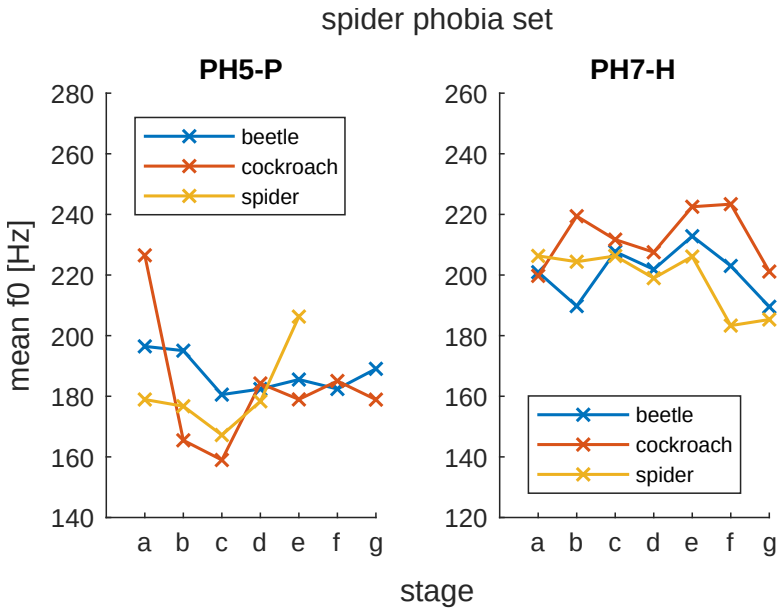


Figure A.3: Mean f_0 [Hz] results comparison of BAT spider phobia animal set between a spider and a snake phobic. The left graph shows the results of a spider phobic with spider phobia animal set, the stressed subject. The right graph shows the results of a snake phobic reacting to a spider phobia animal set, in our case a control subject.

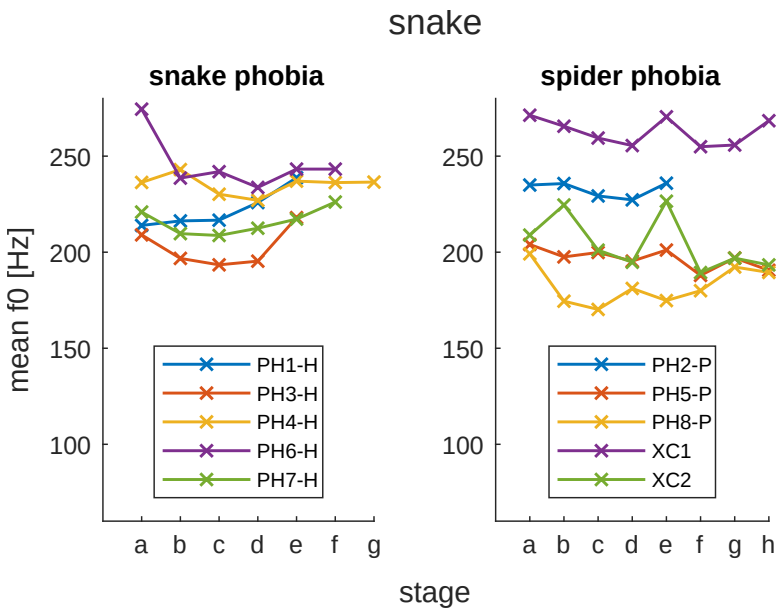


Figure A.4: Snake BAT results of mean f_0 [Hz]. The left graph shows all snake phobia subjects' results of a snake exposure - stressed speech. The right graph shows all spider phobia subjects' and control subjects' results of a snake exposure, in this case - a control group.

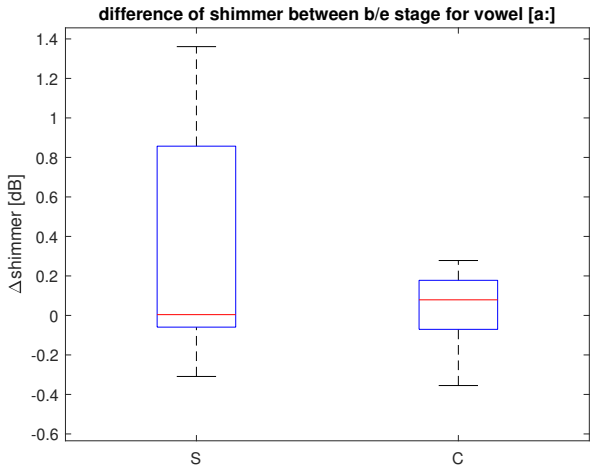


Figure A.5: Boxplot comparison of the resulting shimmer values of vowel [a:] difference between stages b and e for phobia and control stimuli. For phobia stimuli the results of phobia triggering animal exposure, snake (spider) for snake (spider) phobia, were used. For control stimuli, the results of control animal exposure, spider for snake phobia subject and vice versa, were used. S = phobia group, C = control group.

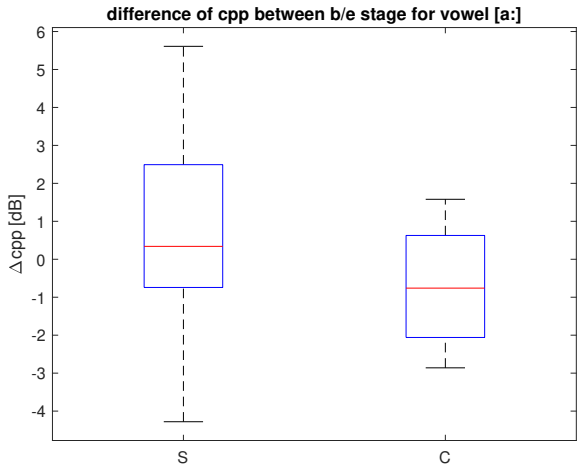


Figure A.6: Boxplot comparison of the resulting CPP values of vowel [a:] difference between stages b and e for phobia and control stimuli. For phobia stimuli the results of phobia triggering animal exposure, snake (spider) for snake (spider) phobia, were used. For control stimuli, the results of control animal exposure, spider for snake phobia subject and vice versa, were used. S = phobia group, C = control group.