

DETECTION OF DECEPTION

THROUGH VOICE ANALYSIS

R. D. Vooijs
Version 0.9
October 12 2004

Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Department of Mediamatics, Man - Machine Interaction

Graduation committee:
Prof. Dr. H. Koppelaar
Dr. Drs. L.J.M. Rothkrantz
Dr. K. van der Meer
Ir. P. Wiggers



Contents

- Contents2

- 1 Introduction.....4
 - 1.1 Problem definition.....4
 - 1.2 About this report4

- 2 Stress assessment through multimedia.....6
 - 2.1 Behaviour Feedback model6
 - 2.1.1 Behaviour6
 - 2.1.2 Media6
 - 2.1.3 Overall model7
 - 2.2 Sound model9

- 3 Stress.....11
 - 3.1 The human nervous system.....11
 - 3.1.1 Autonomic Nervous System.....11
 - 3.1.2 Sympathetic Nervous System12
 - 3.1.3 Parasympathetic Nervous System.....12
 - 3.2 Stress.....12
 - 3.2.1 What is stress12
 - 3.2.2 Stress response models.....13
 - 3.2.3 The General Adaption Syndrome13
 - 3.2.4 Cognitive neurochemical model.....14
 - 3.2.5 The effects of stress15

- 4 Speech signals and human speech production.....18
 - 4.1 Sound and speech signals18
 - 4.1.1 Sound18
 - 4.1.2 Fourier analysis and frequency power spectra.....18
 - 4.1.3 Speech.....20
 - 4.2 The human voice production system21
 - 4.2.1 Voiced sounds.....21
 - 4.2.2 Unvoiced sounds or fricatives22
 - 4.2.3 Plosives or stops23
 - 4.2.4 Conclusion.....23

- 5 The polygraph.....24
 - 5.1 The instrument.....24
 - 5.2 The polygraph session24
 - 5.2.1 Pre-test interview.....25
 - 5.2.2 The test questions26
 - 5.2.3 The first test.....27
 - 5.2.4 The card test.....27
 - 5.2.5 The third test.....28
 - 5.2.6 The mixed question test28
 - 5.2.7 The yes-test.....28
 - 5.2.8 The "Guilt complex" test.....28
 - 5.2.9 The peak of tension test.....28
 - 5.3 Evaluating a polygraph recording29
 - 5.3.1 Effects on respiration.....30
 - 5.3.2 Effects on blood pressure32
 - 5.4 Conclusions33

| | | |
|-------|--|----|
| 6 | Experimental design..... | 34 |
| 6.1 | Procedure..... | 34 |
| 6.2 | Calibration questions..... | 36 |
| 6.3 | Card guessing..... | 36 |
| 6.4 | Subjects and equipment..... | 37 |
| 7 | VoiceMaster..... | 39 |
| 7.1 | The original VoiceMaster..... | 39 |
| 7.1.1 | Analysis functions..... | 39 |
| 7.1.2 | Pitch determination using the Excursion Cycle method..... | 40 |
| 7.1.3 | Jitter determination using the Excursion Cycle method..... | 41 |
| 7.2 | Analysing voice samples with VoiceMaster..... | 41 |
| 7.2.1 | Clipping samples..... | 42 |
| 7.2.2 | Editing excursion cycles..... | 42 |
| 7.2.3 | Input/Output functions..... | 43 |
| 8 | Praat..... | 44 |
| 8.1 | Backgrounds..... | 44 |
| 8.2 | An introduction to Praat..... | 44 |
| 8.2.1 | Absolute Fundamentals: Opening and Closing Files..... | 44 |
| 8.2.2 | Basic Phonetics Functions: The Edit Window..... | 44 |
| 8.2.3 | Running and Implementing Scripts..... | 45 |
| 8.2.4 | Doing Resynthesis..... | 46 |
| 8.2.5 | Doing Other Types of Analyses..... | 47 |
| 9 | VoiceBase..... | 48 |
| 9.1 | Functions and features..... | 48 |
| 9.1.1 | Import samples..... | 48 |
| 9.1.2 | Analysing samples using Praat..... | 48 |
| 9.1.3 | Browse data..... | 49 |
| 9.1.4 | Generating graphs..... | 50 |
| 10 | Experimental results..... | 52 |
| 10.1 | Difference in analytical tools..... | 52 |
| 10.2 | General analysis..... | 53 |
| 10.3 | Voice characteristics for different persons..... | 54 |
| 10.4 | Stress detection per person..... | 55 |
| 11 | Conclusions and recommendations..... | 57 |
| 11.1 | Literature study on stress and the polygraph..... | 57 |
| 11.2 | The experimental setup..... | 57 |
| 11.3 | Stress detection through voice analysis..... | 58 |
| 11.4 | Analysis of test results..... | 58 |
| | Literature..... | 59 |
| | Appendix A: Praat scripts called from VoiceBase..... | 61 |
| | Appendix B: T-test for significant difference between sessions 1, 2 and 3..... | 63 |

1 Introduction

1.1 Problem definition

For some years now, a project called "Stress Assessment through Multimedia" has been going on. It is based on the "Behaviour Feedback Model", whose purpose it was to automatically recognize human moods and behaviour. Since a large part of this emotion recognition consisted of assessing the subject's stress level, the project shifted toward stress assessment.

Lie-detection consists mainly of stress detection. Existing lie-detection systems using physiological data, i.e. polygraphs, rely on detecting the stress that a person experiences when he is lying. The polygraph relies on data gathered from respiration, blood pressure and galvanic skin response.

The purpose of this project is to study the possibility of detection of deception as a supplement to the "Stress Assessment through Multimedia" project. Most psychologists nowadays agree that in the right circumstances stress can be detected in the voice. Voice-analyzing lie-detection equipment has been developed: the Psychological Stress Evaluator. There is however no agreement among psychologists about how valid its results are [Hollien '90].

Therefore a new research was done on 'Detection of deception through voice analysis'. In this experiment subjects were made to lie while their voice was recorded, so that research could be done about whether it is possible to detect stress in the voice caused by lying.

In final, the problem definition can be defined as follows:

- Perform a literature study on stress and the use of the polygraph
- Design an experimental design for stress assessment
- Develop a system for automatic stress assessment based on voice analysis
- Perform the experiment to gather data on deception behaviour
- Analyse the test results
- Draw conclusions based on these results

1.2 About this report

This report consists of three parts. Part one consists of a literature review on some existing projects and models on stress detection, detection of deception and voice analysis. Part two is about experimental setup and analytical tools. Part three contains the test results and conclusions and recommendations.

Part one starts with a description of the "Stress Assessment through Multimedia" project. The purpose of this study is to research a possible addition to this model. Therefore a description of this model and a more detailed description of the sound section of this model are given in chapter two.

Chapter three explains what stress is and how it is caused. It gives a description of the human nervous system, and how it reacts to stress.

An important part of this study consisted of searching for indications of stress in the voice. Chapter four first tells more about sound and about techniques for processing it. Then it tells something about the human voice and how it is produced.

The most widely used system in lie-detection is the polygraph. Chapter five explains how this polygraph works and how it is used to investigate whether subjects are lying.

The second part of this report starts with a description of the experimental setup in chapter six. An experiment was held in which subjects were requested to lie while their voice was recorded. Chapters seven, eight and nine describe the various tools that were used to analyze the collected data.

VoiceMaster, a program for the analysis of voice samples and the extraction of stress indicating features has been developed earlier. It was noticed however that the used algorithms were not very robust, and that this led to faulty stress levels. Therefore some modifications and additions were made to the existing program VoiceMaster and its additions are described in chapter seven.

Chapter eight describes the program Praat that was developed by researchers at the Institute of Phonetic Sciences in Amsterdam. Praat offers the possibility to extract all kinds of parameters from voice samples, but its real strength is the possibility to run scripts so that complete analyses of batches of samples can be done without manual intervention.

A database application called VoiceBase was developed to store the samples and the analyzed data. It has functionality to call Praat scripts in order to get parameters from samples. Furthermore it allows the user to select samples that meet certain criteria and generate plots and statistical analysis from these samples. VoiceBase is described in chapter nine and can be found on the CD-Rom included with this report.

Part three of this report contains the results of the voice analysis in chapter ten. Finally, chapter eleven tells what conclusions were drawn from all this research. Some recommendations are given about how future research into this subject could be approached.

2 Stress assessment through multimedia

The research described in this thesis is part of a larger project: "Stress Assessment through Multimedia". This again is a spin-off from the "Behaviour Feedback System" [Vark '93]. This project researched a system that is designed to observe human behaviour through several different media, and to deduce from this the subject's state of mind or emotions.

The media used are in correspondence with the way in which humans observe behaviour. Ears and eyes for instance, are implemented as the media sound and video. The purpose of the system is to produce feedback on observable behaviour in an automated way, in order to come to conclusions about the subject's state of mind or emotions.

2.1 Behaviour Feedback model

2.1.1 Behaviour

The model described in this chapter observes a person's behaviour, in order to come to conclusions about that person's state of mind. The definition of behaviour that is used here is:

"Those activities of an organism that can be observed by another organism or by an experimenter's instruments and which are reactions to social stimuli or to inner processes or a combination thereof."

Historically, there have been two schools of research in psychology: personal and environmental psychology. Personal psychologists strictly focused on a person's personality as the main determinant of his behaviour. Environmental psychologists state that a person's behaviour is mainly determined by his environment.

Nowadays it is widely accepted however that behaviour has an individual and an environmental component. This leads to the following justification of the behaviour feedback model:

Personal behaviour is influenced by the environment. In other words, when a person is performing different tasks, his behaviour will be influenced by the tasks being performed. It is therefore possible to develop a behaviour feedback system that interprets behaviour and produces feedback about the tasks being performed. On the other hand, behaviour is also influenced by personal differences. This means it is possible to train people in different situations, using the feedback from the behaviour feedback system.

2.1.2 Media

One of the requirements that had been defined before the behaviour feedback system was developed was the use of multimedia. This was induced by the problem to which the model will be applied. Since one medium is not enough for an unambiguous interpretation of behaviour, several different media have been chosen to examine different features of behaviour. For example, a smiling face is not enough to classify a subject as happy; he might as well be nervous. Therefore, when combining the information acquired by the different media,

hopefully, an unambiguous correspondence between the data and the observed behaviour can be found.

The media that have been chosen are sound, vision and physiology. From the physiological signals the nervous activity of the brain and the rest of the body can be derived. Every type of nervous activity accounts for one or more types of human behaviour. Moods can be deduced from voice recordings, and emotions from face analysis. This separation is of course not stringent; voice recordings for example can also help to classify emotions.

2.1.3 Overall model

The Behaviour Feedback System consists of four parallel subsystems and a central interpreting system. The subsystems are one each for sound and physiology and two for vision, one for facial expressions and one for motion.

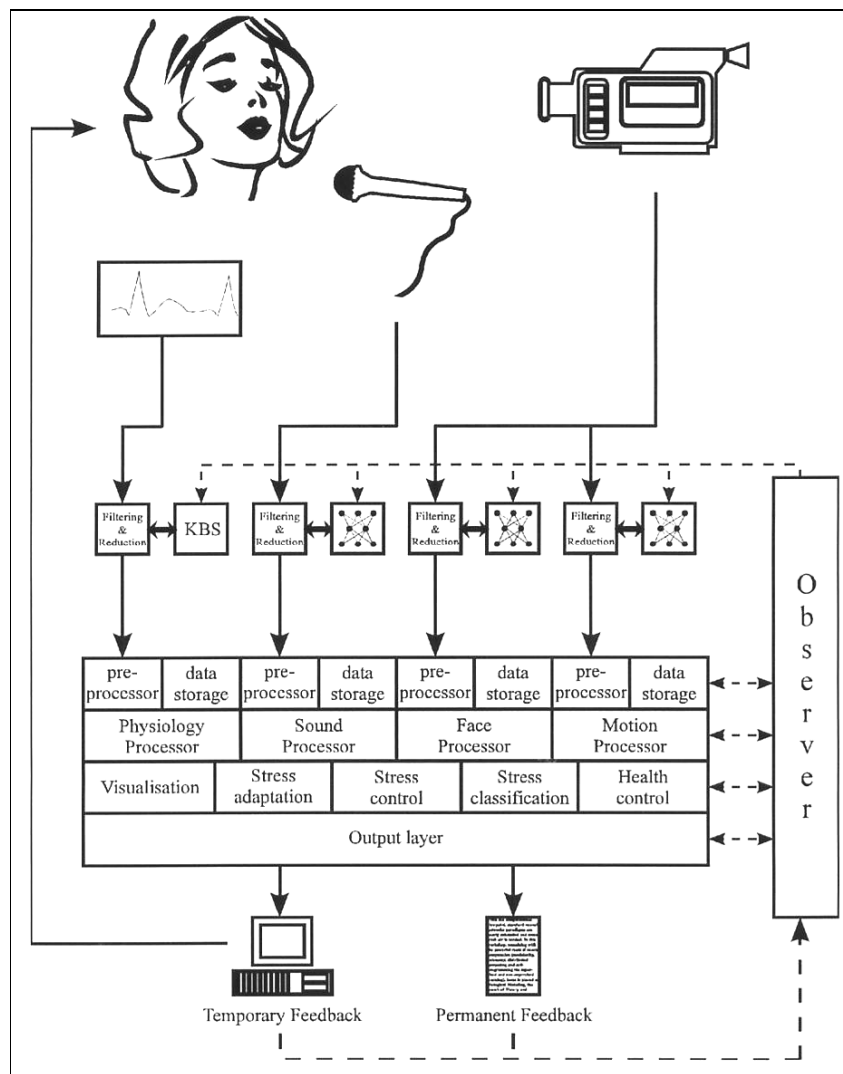


Figure 2.1 Behaviour Feedback System

The central system uses the hypotheses on the psychological state of the subject stated by the four subsystems, to generate the final interpretation and feedback.

The functionality of the system is split into six layers, of which the first four are separately designed for each subsystem. These first four are:

1. Registration layer
In this layer the signals are registered and stored for later use. In the case of the sound subsystem this means recording samples through a microphone and storing them to disk.
2. Filtering and reduction layer
In this layer, several transformations are applied to the incoming signals. These consist of noise reduction or elimination, outlier correction and data reduction for faster processing and smaller storage capacity.
3. Pre-processor layer
This layer prepares the signals for the next medium specific layer. It also stores the data from the signals for later use, so no new measurements have to be made when new insights are obtained.
4. Medium specific layer
In this layer, each subsystem uses knowledge specific to their domain to interpret the pre-processed features and derive hypothesis concerning the possible type of behaviour.

The next two layers are the layers of the central interpreting system.

5. Application layer
This layer integrates information from the subsystems for use of several applications. At the moment, there are five applications for which the model seems suitable:
 - Visualization
This application visualizes information generated by the subsystems.
 - Behaviour (stress) adaptation
This application can help the subject to adjust his behaviour towards a type of optimal behaviour. This is especially helpful when the behaviour optimal for the task is known in advance.
 - Behaviour (stress) control
This application helps the subject to control and prevent undesirable behaviour (for example high stress level). It produces feedback about the subjects and the control behaviour.
 - Behaviour classification
This application combines the hypotheses from the subsystems, and determines the (most likely) type of behaviour the subject is showing. A knowledge based system is used in which all possible combinations of nervous activity, mood, emotion and behaviour are stored.
 - Health control
This application checks the appropriate information to ensure that the subject does not undergo unnecessary health risks.
6. Output layer
The output layer takes care of both temporary and permanent feedback. Temporary feedback is the instant feedback given during a session, and can be used by the trainee to adjust or monitor his behaviour. Permanent feedback produces a complete interpretation and can later be used for analysis.

2.2 Sound model

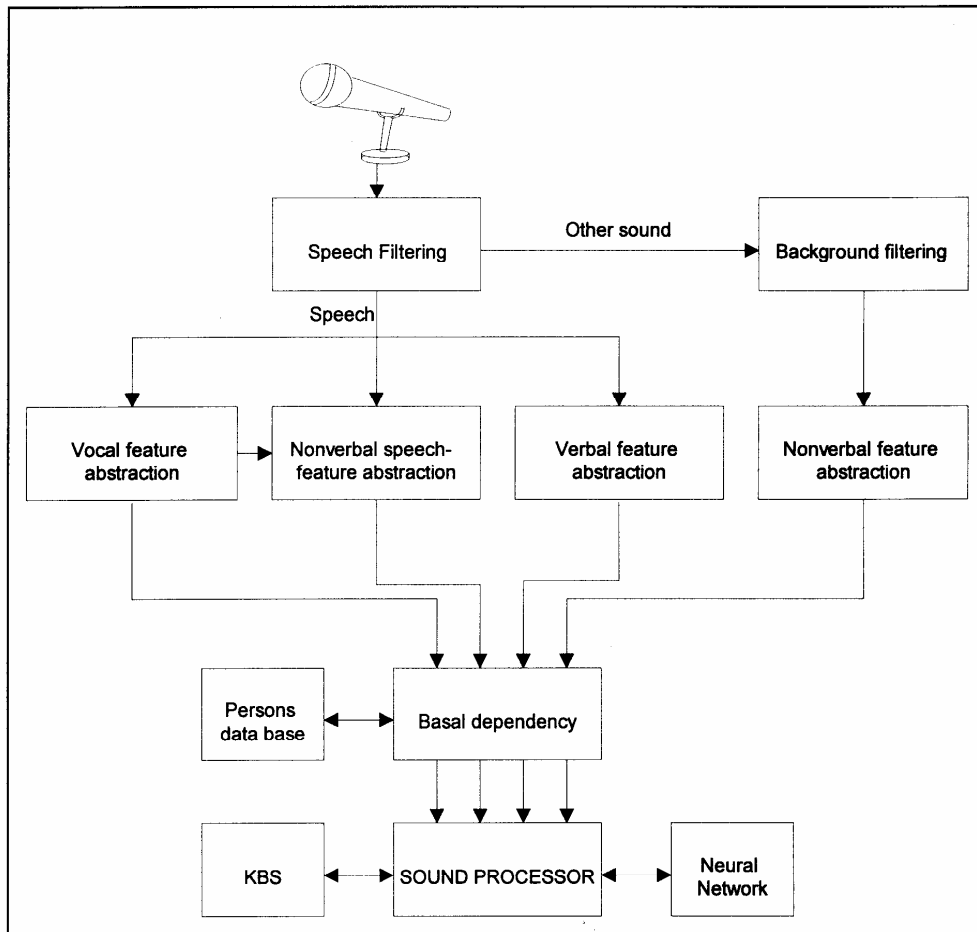


Figure 2.2 Sound model

A schematic model of the sound system is shown in Figure 2.2. The input for the sound system comes from the microphone. The main purpose is to analyze speech signals obtained from a person (the subject), but other sound components can also be taken into account. Therefore, the first part of the model is the separation of the actual speech from other sound components. These components consist of sounds other than speech, which are produced by the subject (like drumming with fingers, etc.) and background noises.

These background noises depend on the environment in which the system is used. For example, the background noise of a pilot in action will be quite different from that of a person who is teleconferencing. In this model the sound that is present in the background and is not produced by the subject will be regarded as background noise, and will not be used. The second part of the model involves the filtering of this background noise.

After the described two filters, what remains as useful sound input is speech and sound which may be caused by the subject's non-verbal behaviour. This non-verbal behaviour could be a number of things, such as moving a chair, sighing, certain hand movements like scratching, rubbing, drumming or playing with objects such as keys or jewellery, or even lighting a cigarette. All these forms of non-verbal behaviour have distinguishing sounds and may be detected using

acoustic analysis of the sound signal, which will be the third part of the model. However, this will not be easy, because all forms of behaviour can occur at the same time.

The rest of the model deals with the extracted speech signal. It serves as input for two different speech analysis parts, namely:

1. Verbal feature abstraction.
This part deals with the contents of speech (language), and focuses mainly on choice of words.
2. Non-verbal feature abstraction.
Non-verbal feature abstraction deals not with what is said but with how it is said. It can be divided into two parts:
 - Vocal feature extraction.
This part deals with the physical speech signal. It examines the speech signal graph for certain characteristics such as timbre, voice quality and intonation. The output of this part consists of different acoustic parameter values.
 - Para-verbal feature extraction.
This part does not look at the words that are spoken, nor does it look at their meaning but it looks at the way they are spoken. Features are for example: repetitions, stuttering, speech rate, pauses, etc.

Since everybody has his own way of speaking, the features have to be compared to some set of basic values. These basic values are stored in a personal database. It is however possible that certain features can be used in a general set that is applicable to all subjects. The three sets of features (verbal, non-verbal and non-speech) now serve as input for the sound processor. The sound processor tries to combine the features in order to draw a conclusion on the behaviour the subject is showing. The sound processor contains the knowledge of relationship between the extracted feature values and different kinds of behaviour.

3 Stress

3.1 The human nervous system

The human nervous system is divided into two parts: the central nervous system and the peripheral nervous system. The central nervous system (CNS) resides in the brain and connected parts of the spinal column. It is responsible for cognitive interpretation and as such of little interest for this study and will therefore not be further discussed.

The peripheral nervous system is of more importance. It consists of the remaining neural tissues and can be further subdivided into a somatic and an autonomic part. The somatic nervous system is composed of the neurons that control the skeletal muscles. Our movements are controlled by these skeletal muscles. For example, lifting one's arm is controlled by the somatic nervous system. As can be expected, the somatic nervous system can be influenced consciously. The main interest of this study therefore lies with the autonomic nervous system.

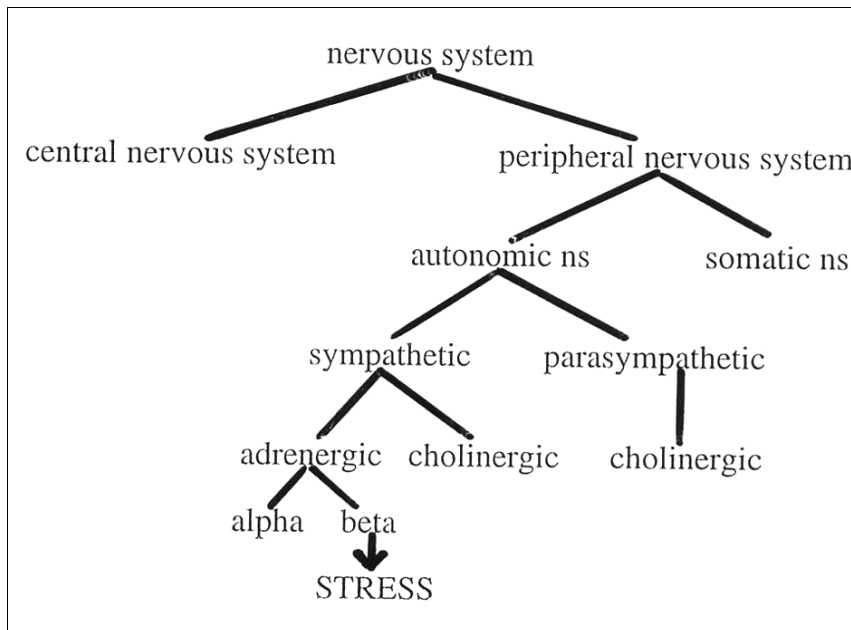


Figure 3.1 The human nervous system

3.1.1 Autonomic Nervous System

The autonomic nervous system (ANS) is composed of neurons regulating the 'vegetative' functions, in other words, functions that are usually not under direct voluntary control. An example of a reflex arc in the ANS is the baroreflex. This reflex is a regulatory mechanism that controls cardiac output, i.e. the blood volume passing through the heart per minute. Any change in blood pressure in the arteries is observed by baroreceptors. These receptors notify the ANS of this change, which in turn signals the heart, causing an increase in heart rate and thus levelling cardiac output.

The ANS can be further subdivided anatomically and functionally into two parts: the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). Both SNS and PNS are dominant under different circumstances

3.1.2 Sympathetic Nervous System

Neural stimulation in the SNS usually involves adrenergic neurotransmission, based on the chemical noradrenalin. Some targets however, like sweat glands, are based on cholinergic transmission and the chemical acetylcholine.

Adrenergic neurotransmission influences two kinds of receptors, alpha and beta. Both receptors are stimulated by the same chemical, but are differently stimulated or blocked by other hormones and neurotransmitters.

The SNS tends to be dominant when processes involved with expenditure of previously stored energy are required, so-called catabolic processes. For example, increased blood transport to skeletal muscles is an effect that originates from the SNS. Sympathetic activation is expected when input regulation as well as output regulation is required. Input regulation takes place when a person processes a change in input, i.e. the presentation of a new stimulus or a change in intensity or timing of a stimulus. Output regulation occurs when a person is spontaneously active and generates changes in the environment. Relative dominance of input or output regulation causes respectively adrenergic or 13-adrenergic enhancement

3.1.3 Parasympathetic Nervous System

Neural stimulation in the PNS is based on cholinergic neurotransmission. The PNS is dominant when processes concerned with preservation, accumulation and storage of energy are required, the so-called anabolic processes. The PNS plays an important role in routine functioning that is vital to maintaining life, like salivation.

Parasympathetic enhancement will prevail during input regulation at cost of output regulation (e.g. arousal, environmental intake and orientation). Parasympathetic inhibition will prevail during output regulation at cost of input regulation (e.g. activation, physical exercise and escape).

3.2 Stress

Nobody can live without a certain amount of stress. Crossing a road, trying to catch a train, or simply a feeling of joy is enough to 'trigger' the stress mechanism. Stress is not necessarily bad for us; stress is part of life as every emotion, every activity causes stress. Nevertheless, one must be able to cope with stress; the same stress that makes one person ill may be a strength giving experience for some other person. Although stress is also associated with positive emotions, for this report we will focus on negative stress such as stress caused by anger or fear.

3.2.1 What is stress

Although everybody has a general notion of what is stress, the wide use of the term 'stress' has led to a multitude of different definitions as well as diverse theories on the underlying processes. While [Selye '56] defined stress as an autonomic physiological reaction pattern, others used the term stress to refer to the stimuli that provoke this pattern, or to describe the psychological evaluation and reaction to these stimuli. Furthermore the stimuli used to provoke a 'stress

reaction', the stressors, which are used in different studies, vary widely: exercising, arithmetic tasks, electric shocks, noise, gory pictures, conflicts, offences or 'life stresses' such as accidents or the death of relatives.

Although it is impossible to give a stringent definition of stress, there seems to be a general consensus that the stress concept involves external or internal stimulation of the organism, which upsets its internal balance. This requires an adaptive coping response in order to restore this balance. When such a coping response takes place, not only depends on the objective nature of the threatening stimulus or event, but also on psychological factors such as perceived significance, predictability, and controllability of the event ([Lazarus '79]).

Furthermore, there is a general agreement that stress involves a series of physiological and behavioural changes which prepare the organism for the appropriate coping response. These changes are characterized by some degree of activation (arousal or emotional tension in the organism) which insures that the body is ready for action ([Cannon '15], [Selye '56]). There is disagreement however on whether there is a specific or a non-specific physiological reaction to this activation.

3.2.2 Stress response models

After reading the previous paragraph, one may be left with the conclusion that stress is such an incomprehensible phenomenon, that it is impossible to accurately describe it. Nevertheless, in the next section two different attempts to describe the stress concept will be discussed. The first is the 'classical' model of the stress reaction, the General Adaption Syndrome ([Selye '56]). The second is a more modern approach that has its roots in neurobiology and can be regarded as a more detailed response model. It also takes into account some of the cognitive factors that determine the amount of stress perceived by a person.

3.2.3 The General Adaption Syndrome

The GAS is the observable manifestation of stress and contains all the non-specific changes in an organism, which occur when it is confronted with a stressor. By means of physiological processes, the organism tries to adapt to the changing environment. Selye defined the GAS as consisting of three phases: the alarm phase, the adaption phase, and the exhaustion phase.

In the alarm phase, a chain of events takes place (Figure 3.2) which affects numerous organs. Hormonal and chemical changes take place in this phase. The main function of the alarm phase is to mobilize its forces and prepare the body for action. The next phase, which is the adaption phase, can be described as the phase of resistance. The physiological processes are more general now, and pointed more towards the specific nature of the stressor. The resistance of the body should now be optimal. However, adaption, and therefore the resistance of the organism, has its limits. If the organism is confronted with a serious stressor during a long time, and it cannot return to a state of balance, it will finally go into the last phase of the GAS: the exhaustion phase. The organism is now running out of energy and the resistance is broken. The symptoms of the alarm phase will occur again, but they are now irreversible.

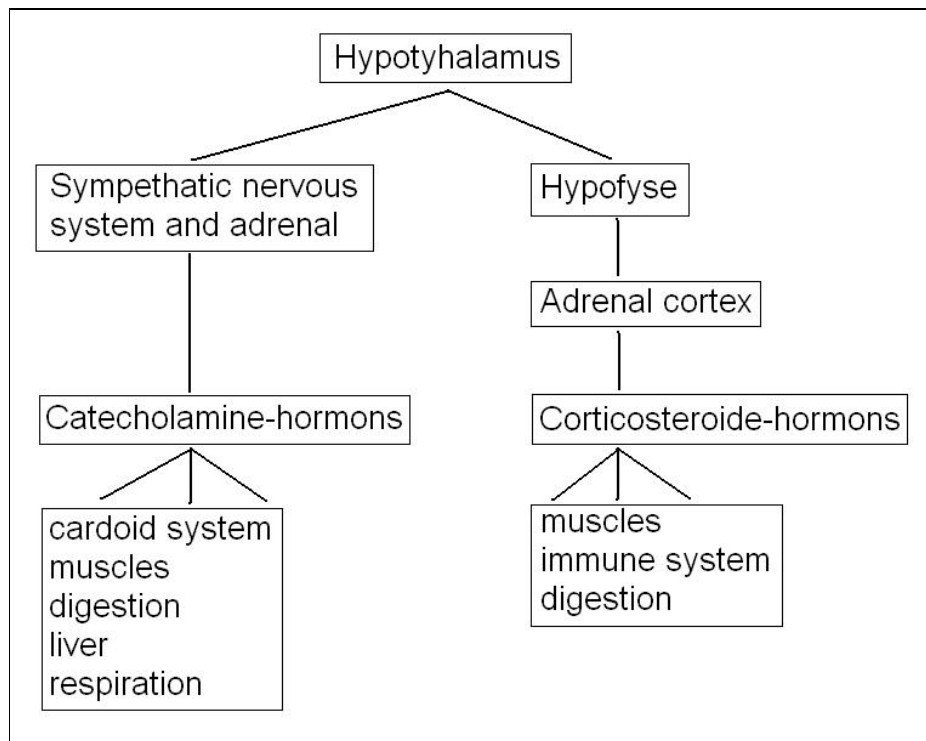


Figure 3.2 Organs involved in stress reaction

The main idea behind Selye's work is the fact that the GAS can result in 'diseases of adaption'. All of these diseases are caused by the physiological stress responses, and not by bacteria or viruses. For example allergies, in which the reaction to the stressor cannot compare with the relatively unarmfull nature of it.

Without trying to put down Selye's pioneering work on stress research, his theory can be criticized in some respects. For example, Mason found that the laboratory settings, in which Selye examined the physical stressors (cold, heat, underfeeding), could bring about many emotional disturbances and pain. He found that the psychological reactions to the stressors were the main factors in the relationship between stressor and stress. For example, when apes were given food without any nutritious value, rather than no food at all, no stress reaction took place. The fact that the apes could not be frustrated by the absence of their feeder or by the feeling of an empty stomach accounted for the absence of the stress reaction.

3.2.4 Cognitive neurochemical model

The model in Figure 3.2 assumes that the stressor directly works on the neurochemical system without cognitive mediation, and its impact is determined by the responsiveness of the relevant neurochemical systems, and the person's relevant learning experiences.

The response of a particular neurochemical system depends on its current state of arousal and also upon its arousability, or capacity to become aroused. 'Arousal' is the current state of activation; 'arousability' is a longer-term phenomenon and depends on a number of factors such as short-term influences like drugs and neurotransmitter levels, and long-term influences like heredity, disease and interactions with other neurochemical systems. Variations in the

arousal and arousability function increase or decrease the number of stimuli that will trigger emotional response. For example, when the neurochemical circuits associated with aggression have a high arousability, the person will react faster with feelings of anger and aggression, even to inappropriate stimuli.

The impact of a stressor is also a function of the learning experience. For example, a negative experience with high places might cause fear of heights. This subjective experience can be strengthened, when the relevant neurochemical systems are aroused or in a state of high arousability.

Stressors are 'filtered' by learning experiences and arousal/arousability, and the result determines their impact. This impact occurs both at cognitive and emotional level. At the emotional level adaptive and homeostatic processes are activated, which can be compared with the alarm phase of the GAS, and the physiological response takes place as a result of these processes. On the cognitive level the stressors are 'labelled' or 'appraised' on the ground of past experience and, the present situation ([Schachter '75], [Lazarus '79]). Once the situation has been labelled, the appropriate goal-directed 'coping' responses can be made. The labelled state can, however, become a stressor in itself and start a new cycle of response. At the same time the behaviour and the spontaneous expression are influenced by display rules, so the observed responses might not reveal the true feelings.

3.2.5 The effects of stress

In the following discussion, no justice can be done to the complex pattern of findings of the stress response of the human body. For now, the discussion will be restricted to the physiological responses, which have been repeatedly observed to follow stress events.

Physiological effects

It is well known that the human body will respond physiologically to a stressful situation. Responses of this type are induced by the ANS, which co-ordinates the activities of the endocrine system and smooth muscle tissue associated with the intestines, blood vessels and heart.

More importantly, the system essentially operates involuntary and can normally not be brought under conscious control. Of the two branches comprising the ANS, it is the sympathetic nervous system (SNS) that is of interest here. That is, while the PNS is dominant when an individual is at rest, the SNS takes over when energy mobilization is required. Specifically the SNS is activated by a perceived threat, and in response prepares the body to cope with the emergency, i.e. to flee or to fight.

When emotions such as fear, anger and anxiety (i.e. stress) occur, they stimulate physiological change. Among these are responses are increases in blood pressure and heart rate, decreases in skin resistance and changes in respiratory rates ([Vark '93], [Hollien '90]).

Effects on speech

As we have seen, the physiological stress response of an organism is dominated by sympathetic arousal, which is characterized by changes in heart rate, blood pressure, respiration patterns, and muscle tension. Of these, respiration and muscle tension are likely to have an effect on speech production.

Respiration will affect the sub glottal pressure in phonation, and muscle tension will affect the laryngeal mechanisms involved in phonation, as well as the characteristics of the vocal tract's resonance walls and the articulatory mechanism. Furthermore, disturbances in the co-ordination of neural impulses that control phonation and articulation, may affect the speech production.

Figure 3.3 shows a rather simplified diagram of the speech production system. For each of the three major speech production mechanisms (respiration, phonation and articulation), the most important functional variables which are likely to be affected by sympathetic arousal are shown.

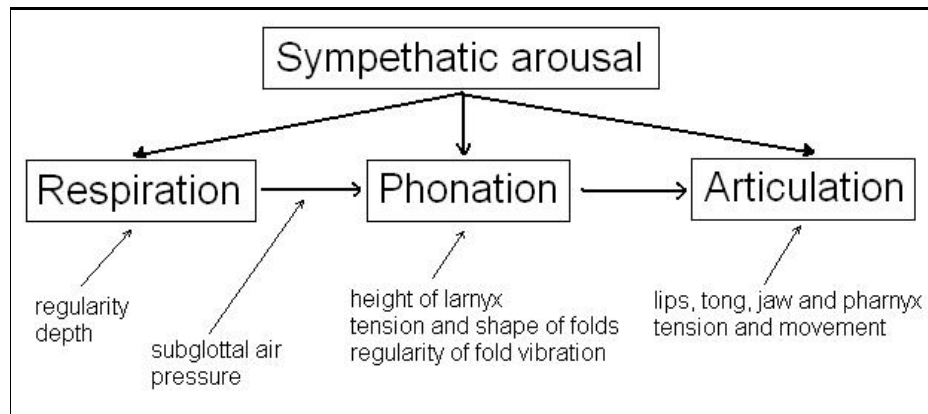


Figure 3.3 Speech production system

Unfortunately, little is known about the effect of emotional arousal on the physiology of the speech production mechanism. Nevertheless, in spite of the lack of evidence, one can cautiously state some hypotheses on the relationship between stress, physiological response and the speech signal. Given the deepening of respiration and dilation of the bronchi ([Gray '71]), one can expect higher intensity and higher fundamental frequency due to the increased sub glottal pressure and higher tension of the vocal fold, as well as a shift of the energy concentration in the spectrum towards higher frequencies.

Almost all of the parameters summed up in Table 3.1 have been shown, in recent research, to be affected by the vocal expression of emotion, emotional disturbance or stress.

| Parameter | Description |
|------------------------------|---|
| Fo mean | Fundamental frequency (vibration of the vocal folds as averaged over a speech utterance) |
| Pitch range | Difference between highest and lowest Pitch in an utterance |
| Pitch variability | Measure of dispersion of Pitch (e.g. standard deviation) |
| Pitch perturbation or jitter | Slight variations in the duration of glottal cycles |
| Pitch contour | Fundamental frequency values plotted over time (intonation) |
| F1 mean | Frequency of first (lowest) formant (significant energy concentration in the spectrum) averaged over an utterance |
| F2 mean | Mean frequency of the second formant |
| Intensity mean | Energy values for a speech sound wave averaged over an utterance |

| | |
|-----------------------|--|
| Intensity range | Difference between highest and lowest intensity value in an utterance |
| Intensity variability | Measure of dispersion of the intensity values (e.g. standard deviation) |
| High frequency energy | Relative proportion of energy in the upper region |
| Speech rate | Length of an utterance |
| Spectral noise | A-periodic energy components in the spectrum |
| Zero Crossings | Number of times a sound wave graph crosses the zero line (e.g. measured in number of times per second) |

Table 3.1 Overview of major acoustic parameters [Scherer '89]

Only few of these parameters are most likely to qualify as indicators of stress. These are the Pitch mean, Pitch perturbation, high frequency energy and speech rate. The intensity may also qualify, but this variable has not been sufficiently studied yet (due to the difficulty in getting absolute amplitude levels). The other variables might possibly serve as vocal stress indicators, but until there is not any systematic evidence from carefully designed experiments, they cannot be taken for granted as indicators of stress.

Because virtually all studies report strong individual differences in terms of the number and kind of vocal indicators that seem to accompany stress, this is a major concern for the research. One has either to use stressors that have the appropriate effect on virtually the entire subject, or measures have to be developed independent of vocal response to allow appropriate co-variation. In this case, psycho-physiological measurement seems the most promising, but when individual differences can not be eliminated, attempts can be made to get more grip on the sources of the individual differences. Knowing more about individual reactivity may enable us to predict differential vocal responses to stress.

Thus, the future of this research is divided over different disciplines of research. Lately we have been experiencing a major technological breakthrough, in the sense that rather sophisticated digital voice analysis techniques are available on standard personal computers. Researchers in these disciplines, for whom the use of tape recorders and video tape recorders are already counted as a major technological advance, can make use of this improved technology. Consequently, one could hope for an increased use of objective methods of acoustic measurement in the field of vocal expression research.

4 Speech signals and human speech production

Voice analysis requires a basic understanding of sound signals in general, and speech signals in particular. It also requires knowledge of how human speech signals are produced. This chapter gives a short introduction in these topics, explaining the basic theories and mechanisms and introducing terms used throughout the rest of this report.

The first section is about sound and speech signals and the second section gives a description of the human speech production system.

4.1 Sound and speech signals

This section will give an introduction into the physical aspects of sound and speech. The best way to gain insight into the structure of a sound signal is Fourier analysis, and therefore this will also be introduced.

4.1.1 Sound

Before talking about speech, one should take a look at sound in general. Sound is nothing more than waves of small pressure changes in the air. The loudness or intensity of a sound corresponds with the magnitude or amplitude of these pressure variations. The higher the amplitude, the louder the sound. The pitch of the sound corresponds with the number of pressure variations per unit of time. The higher this number of variations, the higher the pitch.

A sound signal is the variation of the amplitude of the air pressure over time. The simplest sound signal is a sinusoidal wave, or sine wave. A sine wave is a periodic signal as it repeats itself endlessly. The length of one repetition is called the period of the wave, and the number of repetitions per second is called its frequency. The length of the period is given in seconds (or milliseconds) and the frequency is expressed in hertz (Hz) or kilohertz (kHz).

Most sound signals, including speech signals, are far more complex. Nevertheless, complex signals can still be periodic. An example of a complex signal is given in Figure 4.1.

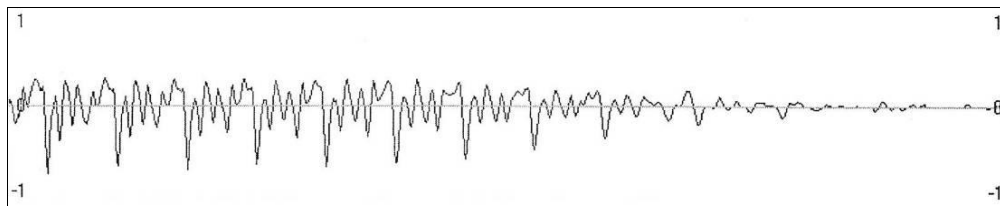


Figure 4.1 Waveform of a sound signal

It can easily be seen that this vowel sound is periodic. The length of one period of a complex periodic sound is called the fundamental period and the number of repetitions per second is called the fundamental frequency (or Pitch). The fundamental frequency is a very important characteristic feature of the human voice.

4.1.2 Fourier analysis and frequency power spectra

The French mathematician Fourier showed that every continuous periodic signal, no matter how complex, can be decomposed into a number of purely sinusoidal components with different frequencies. This decomposition is called Fourier transformation. If a periodic signal is decomposed into sinusoidal components,

the frequencies of these components are all multiples (or harmonics) of the fundamental frequency of the original signal.

Figure 4.1 is an examples of a signal displayed in the time domain. The time is on the horizontal axis, and the amplitude is on the vertical axis. Fourier analysis makes it possible to view signals in the frequency domain. In frequency domain displays, the amplitude of the frequency components is plotted against the frequency, so the frequency is on the horizontal axis. The frequency domain plot is also called the frequency power spectrum, or simply the spectrum of the signal. Figure 4.2 shows a complex periodic wave composed of two sinusoid waves, displayed both in the time and the frequency domains.

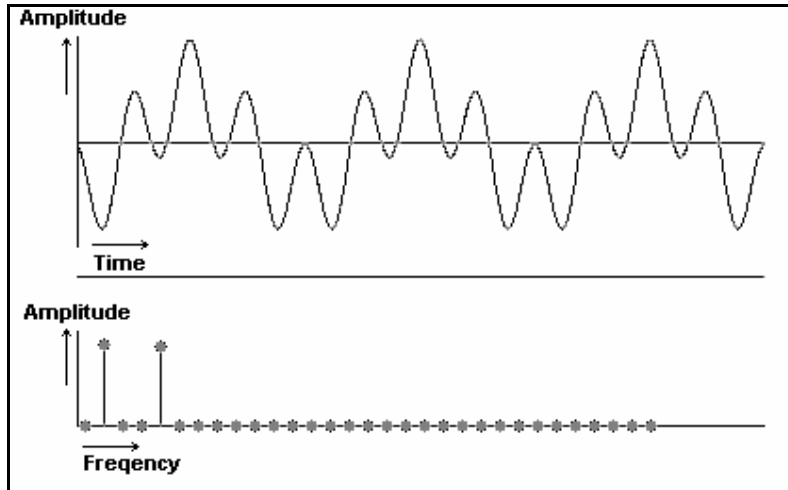


Figure 4.2 Fourier analysis of a periodic wave

Frequency spectra are a good means to get insight in the frequency distribution of a sound signal. From Figure 4.2 it is immediately clear that the signal consists of two sine waves, of which one has a frequency three times the frequency of the other.

The drawback with these frequency power spectrum displays is that they display the frequency distribution of the entire sound signal. In reality though, the frequency distribution of a signal can differ in time. If this time-dependency is relevant for the analysis, another method of displaying the power spectrum is used: the spectrogram. In spectrograms the frequency is displayed on the vertical, and time is displayed on the horizontal axis. The frequency distribution is displayed by using different shades of grey. An example of a spectrogram is given in Figure 4.3.

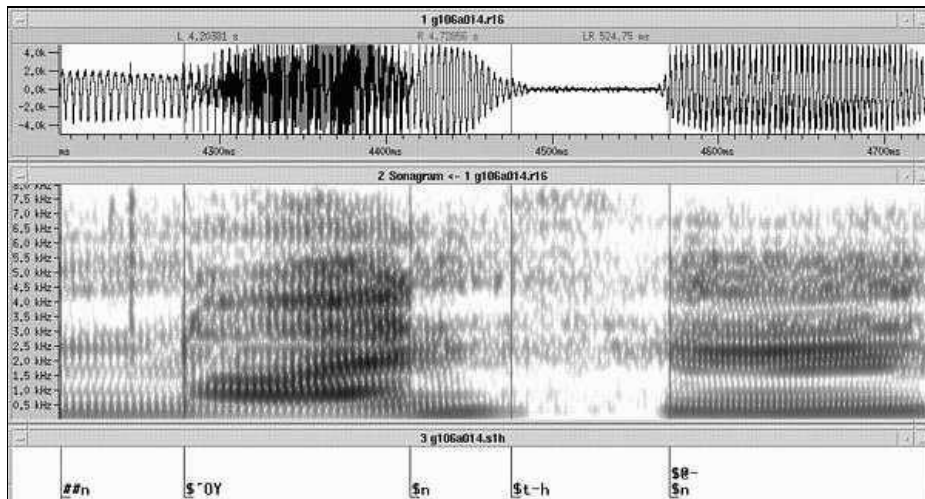


Figure 4.3 Example of a spectrogram

4.1.3 Speech

Speech signals are complex sound signals that contain frequencies ranging from about 50 Hz to about 10,000 Hz. Speech signals cannot be classified as either periodic or non-periodic signals. They consist of both periodic and non-periodic parts. All vowels and some of the consonants are periodic. These are called voiced sounds. Section 4.2 will give a more precise description of voice and unvoiced sounds and how they are produced. The rest of this report will show that the voiced segments of speech are more important to non-verbal voice analysis. Especially the fundamental frequency of these voiced periodic segments is an important feature.

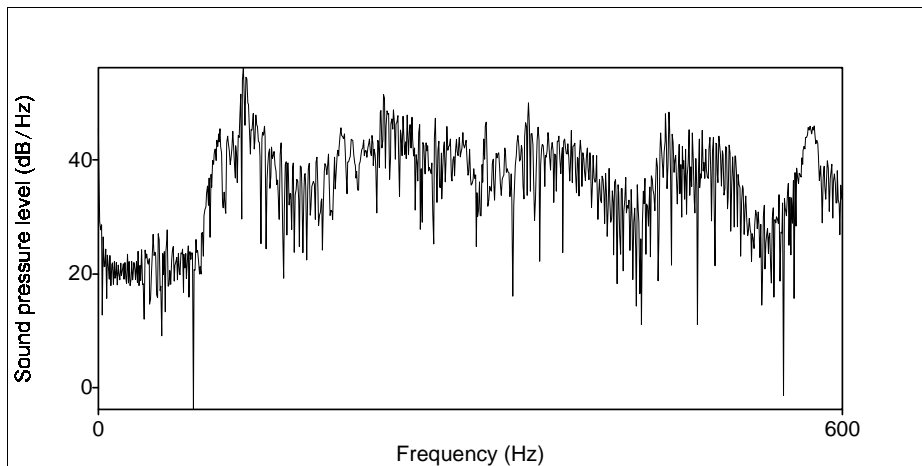


Figure 4.4 Power spectrum of a sentence

Figure 4.4 is an example of a frequency power spectrum of a complete sentence of speech. The graph shows several peaks. These peaks are harmonics of the fundamental frequency and are not like the narrow peaks in Figure 4.2. The main reason for this is that speech signals have both periodic and non-periodic parts.

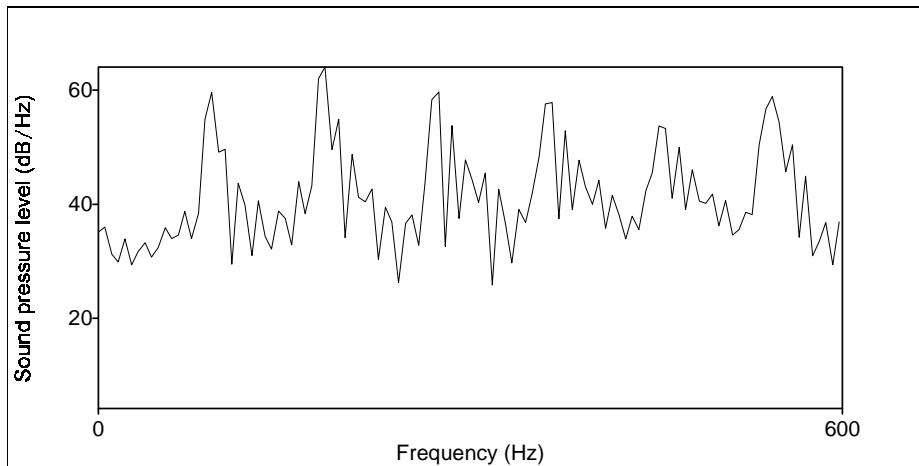


Figure 4.5 Power spectrum of a short voiced sound

Figure 4.5 shows the spectrum of only a short voiced segment of speech. The peaks should be at distinct, individual frequencies, but examining the graph learns that this is not the case, although the peaks are a lot sharper than in the full sentence. The reason for these still relatively broad peaks is that periodic parts of speech are never exactly periodic. The individual periods of the signal are never exactly the same.

For many applications it is important to know how the frequency distribution of a speech signal changes over time. Speech spectrograms are very useful for displaying this information. A spectrogram shows a large number of short-time frequency spectra in one graph. These spectra are drawn vertically, the intensity of the colour reflecting the frequency distribution. Consequently, peaks in the spectra are now visible as dark horizontal stripes in the spectrogram.

4.2 The human voice production system

This section will describe the human voice production system and its components, and will explain how this system is used to produce the different classes of speech sounds ([Poulton '83], [Scherer '82]).

The general process of voice production can be described as follows: the lungs, the power source of speaking, produce a flow of air through the glottis, which is the opening between the vocal cords. These are situated in the larynx, or Adam's apple. This flow of air results in a sound waveform, which is periodic or non-periodic depending on the tension of the vocal cords. This sound signal enters the vocal tract, which consists of the throat, mouth and nose. The vocal tract then acts like an acoustical tube, and filters this sound signal, amplifying some frequencies and attenuating others. The resulting sound signal comes out of the mouth and/or nose as speech.

In short this is the way in which speech production works. The rest of this section explains in detail how the different classes of speech sounds are produced. These classes are: voiced sounds, the most important for our research; unvoiced sounds and plosive sounds.

4.2.1 Voiced sounds

The group of voiced sounds consists of all of the vowels and some of the consonants. The consonants that are voiced include the semi-vowels /w/, /l/, /r/

and /v/, the voiced fricatives /v/, /th/, /z/ and /zh/ and the voiced stops /b/, /d/ and /g/. The production of these consonants is generally more complicated than the production of the vowels. This is because the vowels are pure voiced sounds while the voiced consonants are combinations of voiced sounds with unvoiced or plosive sounds. For simplicity reasons only the production of the pure voiced sounds, the vowels, is explained here.

For the production of a vowel, the vocal cords are tensed and the airflow from the lungs builds up a pressure below them. Once this pressure is higher than a certain threshold, the vocal cords burst open and a pulse of air is released. After this, the vocal cords close again and the pressure builds up again, after which the cycle repeats itself. This opening and closing of the glottis produces a triangular, periodic sound wave. This waveform is very rich in harmonics: it has a lot of components at integer multiples of its fundamental frequency.

This periodic signal enters the vocal tract. This vocal tract, consisting of the throat, the oral cavity and the nasal cavity, functions like a very complex acoustical tube. It filters the triangular source wave, amplifying some of its harmonics and attenuating others. The resulting signal is the speech wave.

A frequency domain plot of a speech signal, like Figure 4.5, show peaks at the harmonics, which have been amplified by the acoustic filter of the vocal tract. These peaks are at the so-called formant frequencies, or just formants. These formants are numbered from one up, so they are called F1, F2, F3 and so on. In a speech spectrogram the formants show up as dark horizontal bands.

The positions of these formants depend on the shape of the vocal tract and are characteristic of the sound that is produced. The shape of the vocal tract is determined mainly by the position of the tongue, but is also influenced by the position of the jaw and the lips. As a result of this, different placements of tongue, jaw and lips produce different vowels from the same basic triangular wave. For some vowels, called diphthongs, the shape of the vocal tract changes even during the production of the vowel. Examples of diphthongs are /ell/ as in bay, /oU/ as in boat, /al/ as in buy, /all/ as in how, loll as in boy and /jut/ as in you. Diphthongs can be recognized in spectrograms by the gradually changing formant positions.

It must be noted, that the fundamental frequency of the resulting speech signal is the same as the frequency of the periodic, triangular source signal produced by the vocal cords. This fundamental frequency of these glottal pulses is determined by many factors, like the air pressure below the vocal cords. However, it is also strongly influenced by the length, thickness, mass and tension of the vocal cords. The resulting fundamental frequency is therefore highly person dependent and gender dependent; males generally have a fundamental frequency of 70 to 140 Hz and females of 190 to 240 Hz.

4.2.2 Unvoiced sounds or fricatives

Unvoiced sounds, also called fricatives, are non-periodic and noise-like sounds. For the production of unvoiced sounds, the source of excitation for the vocal tract is not the glottis, as is case with voiced sounds, but a constriction somewhere in the proper vocal tract. During the production of unvoiced sounds, the vocal cords are not tensed and the air, which is pressed up from the lungs, can pass the glottis freely. At the constriction in the vocal tract air turbulence occurs, which produces noise in the form of an a-periodic waveform. This constriction can be at

different points in the vocal tract. For /f/, for example, the constriction is near the lips and for /s/ it is near the middle of the oral tract.

This noise-like source waveform is filtered by the vocal tract, just like the glottal pulse waveform during the production of voiced sounds. In this case though, the spectrum of the source waveform does not consist of harmonics of a fundamental frequency. The spectrum of a noise-like waveform is flat: the total energy is more evenly distributed over the entire range of frequencies. This is also apparent in speech spectrograms: during unvoiced sounds no dark horizontal bands are seen.

Besides the unvoiced fricatives, there are also voiced fricatives like /v/ and /z/. The production of the voiced fricatives is essentially the same as of unvoiced fricatives but in addition the vocal cords vibrate. Therefore there now are two excitation sources for the vocal tract. Apart from the a-periodic noise sound, a periodic wave is also present in the resulting speech signal.

4.2.3 Plosives or stops

Like the fricatives, there are two different kinds of plosives or stops: voiced and unvoiced stops. Both voiced and unvoiced stops are produced by letting air pressure build up behind a constriction somewhere in the vocal tract, and then releasing the air. The difference between voiced and unvoiced stops is that for voiced stops the vocal cords vibrate as the pressure builds up. The stop consonant that is produced depends on the position of the constriction and whether the stop is voiced or unvoiced. If the constriction is at the lips, the /b/ (voiced) or /p/ (unvoiced) is produced. A constriction near the back of the teeth results in /d/ or /t/ and a constriction near the velum produces a /g/ or a /k/.

4.2.4 Conclusion

Considering the production of the different classes of speech sounds, it can be concluded that there are two main sources of influence that contribute to the resulting speech waveform. First, the glottal pulse, which is determined by the sub glottal pressure from the lungs and the length, thickness, mass and tension of the vocal cords. The second source of influence is the shape of the vocal tract, which determines its filter characteristics.

Both these two factors have their own effect on the speech waveform. The fundamental frequency is totally determined by the rate of opening and closing of the glottis, and therefore, indirectly by the size, shape and tension of the vocal cords. The formant positions are determined by the filter characteristics of the vocal tract.

For speech recognition, it is important to know exactly how all phonemes are produced and how they can be recognized from the speech waveform and its spectrogram. However, for the research presented in this report it is not relevant to be able to recognize all phonemes. For non-verbal voice analysis it is not necessary to know what is being said, the only goal is to extract the characteristic features from the speech signal. Since, for this report we are concentrating on the fundamental frequency and variations in the fundamental frequency, it can be concluded from this chapter that the voiced sounds are most important for this research. From voiced sounds and from vowels in particular, the extraction of these features is the easiest, because vowel sounds are influenced both by the glottal pulse and the vocal tract shape.

5 The polygraph

5.1 The instrument

In popular fancy, the polygraph or lie-detector is often thought of as an instrument that is supposed to ring a bell, flash a light, or produce some other quick and positive indication of a lie whenever one is told. Unfortunately, at the present there is no real instrument that will detect deception so simply and effectively.

A polygraph is basically a pneumatically operated mechanical recorder of changes in blood pressure, pulse and respiration, plus a unit for recording the galvanic skin resistance (GSR, basically a measure for sweat production). These instruments direct pens, which record the information on a band of paper that rolls by at a constant rate.



Figure 5.1 Test subject attached to polygraph

The polygraph, as illustrated in Figure 5.1, is attached to the subject in the following manner. The pneumograph's tubes are fastened around the person's chest and abdomen, the blood pressure cuff is fastened around the person's right arm, and the GSR-electrodes are attached to the index finger and the ring finger of the left hand. The pneumograph's tubes consist of rubber tubes which are sealed on one end, and connected to the instrument by smaller tubes. As the subject breathes, his chest and abdomen expand and contract. This stretches the tubes, creating pressure changes in the tube, which are transferred into movements of the recording pens.

5.2 The polygraph session

In this section a description is given of what a polygraph session would look like, as described by [Reid '77]. Where necessary, the proceedings will be demonstrated by the following (fictitious) example.

It is assumed for this example that the subject to be tested is Joe "Red" Blake, and that he is suspected of the murder of John Jones, during the course of an armed robbery at his house last Saturday night. In the robbery, Jones' watch was

taken from him. It is further assumed that Blake knows the identity of the victim as John Jones.

5.2.1 Pre-test interview

No test should be conducted without a pre-test interview, during which the subject is conditioned for the test, and the questions to be asked can be carefully formulated by the examiner. The pre-test interview also involves the asking of a series of questions which allow the examiner to get an impression of the subject's truthful or deceptive status without unnecessarily releasing his tension.

The examiner should take a moment to talk in all openness with the subject about the instrument, the recordings and the test procedure. This will serve to increase a lying subject's concern over possible detection, which is an important requirement for an effective polygraph examination. The examiner should discourage any discussion about the matter under investigation though, because this could give the subject the opportunity to relieve some of the tension of lying. It is also important that the examiner remains completely objective throughout the test. If the subject feels that he is already considered responsible, his feelings of resentment would disturb the test results.

When the subject enters the room, the examiner first attaches him to the machine (see Figure 5.1). This gives the subject a chance to get used to the attachments, and the examiner a chance to calibrate the polygraph to the subject's settings. When this is done, the examiner will start asking a number of questions. The subject is lead to believe that this is to help formulate the questions that will be asked during the examination, and to evaluate the recording afterward. While asking these questions however, the examiner closely observes how the subject formulates his answers, in order to get a general idea about the subject's truthful or deceptive status and his willingness to co-operate. Questions that could be asked here (with respect to the John Jones murder) are:

- "Is there any reason why your fingerprints would be on a glass at John Jones's house?"
A truthful subject will immediately answer something like "No, I can not think of any reason", except of course when the subject might have a valid reason for being in the house.
A lying subject on the other hand would be inclined to come up with a reason like "Well, I was in his house a couple of days ago and I may have touched a glass somehow". In other words, a lying subject may feel impelled to make up an explanation, while a truthful person would have no such concern.
- "Did you ever think about killing anyone, even though you didn't do it?"
A truthful subject would deny this without any restraint, for even though he may have had a fleeting thought along that line, he interprets the examiners question to mean a serious, deliberate thought of killing someone.
On the other hand, a lying subject might say something like "Sure I've thought about things like that, everybody does."
- "What do you think they should do to the guy who did this?"
Here a truthful subject might say something like "String the SOB up, he certainly got me in a lot of trouble". A lying subject would shift around in his chair and say something like "Well, it all depends of what made him do it"

Some other questions that could be asked are:

- Have you ever taken a polygraph test before?
- Did anybody tell you how to behave on this test?

- Did you take any medication or drugs within the last twelve hours?
- Did you tell anybody that you would be taking this test today?

Again, the answers themselves are less important than the other observations that can be made. For instance, the delay in answering, bodily movements, movements of the eye and the general attitude of the subject while answering questions are of considerable value. A lying subject's delay in answering a question results from indecision on his part as to whether a half-truth might permit him to evade detection better than a full lie. Generally, a truthful subject has good eye-to-eye contact, moves very little, talks directly and casually and appears to be at ease during the interview.

No final conclusions can be drawn from the subject's answers or reactions, but they are very helpful as factors to be considered in the ultimate decision about truthfulness or deception.

5.2.2 The test questions

The phraseology of the test questions is an extremely important aspect of the examination. The questions must be as simple as possible and without such double inquiries as "Did you shoot the victim and then run away". This would combine two questions, one of which can be truthfully answered by "yes" and the other by "no". Words like "murder", "rape" and "embezzlement" should not be used because they do not have a precise enough meaning. Instead of "murder", words like "stab" or "shoot" should be used. Usually the perpetrator will have rationalised the event so that he sees the act of killing no longer as murder. The questions should also be about factual information. For example, asking a question like "Were you drunk on the evening of August 16?" would require the subject's personal interpretation, and this may be completely different from the examiner's interpretation.

After the pre-test interview, the examiner begins composing the list of questions that will be asked during the actual test. This list generally consists of about ten questions. Of course, the list will contain a number of questions that are relevant to the incident under investigation. Besides that, there must also be some control questions. A control question should be a question to which the subject will almost certainly lie, or to which his answer will at least be of dubious validity in his own mind. There will also be a number of irrelevant questions, which will enable the examiner to measure the subject's normal reactions, and to separate the reactions of the relevant questions.

In our hypothetical case the questions might be:

1. Do they call you 'Red'?
2. Are you over 21 years of age?
3. Last Saturday night did you shoot John Jones?
4. Are you in Chicago now?
5. Did you kill John Jones?
6. Besides what you told me about, did you ever steal anything?
7. Did you ever go to school?
8. Did you steal John Jones' watch last Saturday night?
9. Do you know who shot John Jones?
10. Did you ever steal anything from a place you worked?

The questions three, five, eight and nine obviously are relevant questions. Questions six and ten are control questions. Questions one, two, four and seven are irrelevant.

The test starts with two irrelevant questions to allow the subject to acclimate to the test. The most important and relevant question "Did you kill John Jones" is on number five because at this time the subject is best conditioned for a response or lack of response. Immediately after this comes the first control question, so that if the subject reacts to the control question but not to the relevant question, this can be seen as a sign of truthfulness.

5.2.3 The first test

After formulating the questions, the instruments get a final adjustment. Now before starting the actual test, the examiner tells the subject "If you're telling the truth, the lie-detector will show. If you're not, it will show that too", as a final stimulation to the subject. After this, the questions are asked with intervals of about fifteen seconds. After each question is asked, the number is placed on the recording chart. Also, at the point on the chart when the subject gives his answer, the examiner should place a symbol to mark this, and to indicate what the answer was.

If the subject sighs or coughs at the time, or near the time, a relevant question is asked, an irrelevant question should be asked next. This will assist in stabilising the subject's responses, so that the same relevant question can be asked again before proceeding with the rest of the questions.

During the test, the examiner observes the subject to see if he is trying to influence his responses. Things he should pay attention to are:

- Is the subject moving during the test?
- Does he close his eyes?
- Does he hesitate before answering?
- Is he trying to control his breathing?
- Is he purposefully tensioning his muscles?

If any of these observations are made, the examiner should not mention this until after question six. Only then should he warn the subject that if he does not stop this behaviour, it will be taken as an indication of deception. The test can then be taken again. The reason why the examiner doesn't react before question six, is so that he can see if the subject does his things only at the relevant, only at the irrelevant, at the control questions or at all the questions.

5.2.4 The card test

After the first test usually follows the card test. This test is meant to stimulate the subject's belief that the polygraph test is effective. This will lead an innocent subject to feel more at ease because he knows he won't be falsely accused, and a guilty subject to feel even more apprehension about taking this test. For this card test, the examiner shows seven differently numbered cards, all face down. He then tells the subject to select a card, look at it and put it back.

The cards are arranged and shown to the subject in such a way that the examiner knows which card was selected. After the selection is made, he instructs the subject to answer all questions regarding the card with "no", even when asked about the card he selected. He then calls out each card as part of the question "Did you pick number ... ?". After he has asked the question about the selected card, he asks about one more card, and then repeats the question about

the selected card. This will give the subject the feeling that a significant response was detected when he lied.

5.2.5 The third test

After the card test has been finished, the examiner tells the subject what the selected card was, and that apparently the machine is working properly. He then asks the subject to remember the questions he was asked in the first test, and if there is anything he wants to explain about his answers to them. If this is the case, the questions may have to be rewritten so that they fit the test again. For instance, if the subject says to have a suspicion about who murdered John Jones, question nine will have to be rewritten to "Do you know for sure who shot John Jones?".

After this, the third test is conducted in the same manner as the first one. At this moment, the examiner evaluates the recordings of tests one and three. Dependent on his evaluation, he has several options. If the recordings are clearly indicative of either deception or truthfulness, the test is hereby finished. This happens only twenty-five per cent of the time though. If the recordings are not that clear, a number of other tests are available.

5.2.6 The mixed question test

The responses to the different questions may be dependent on the position of the question in the list. For instance, it is possible that the subject is anticipating a question that makes him anxious, either a relevant or a control question, and responds early so that the reaction is seen at the wrong question. It is also possible that a significant response is seen at question number three, which may be because it is the first relevant question.

This test also gives the opportunity to test other combinations of relevant questions and control question combinations.

5.2.7 The yes-test

If the examiner expects that the subject is consciously influencing his responses, he can apply the yes-test. The yes-test is conducted by instructing the subject to answer "yes" to all the questions he will now be asked. In such a test, a lying subject will often try to distort his recordings to make his "yes" answers look like lies. During this test, no control questions are asked, because the subject might be concerned about them, and tempted to distort the recordings because of that.

5.2.8 The "Guilt complex" test

If the subject reacts heavily to both the relevant questions and the control questions, this may be because the subject feels some sort of guilt complex, or he feels otherwise agitated about answering to questions related to some crime. In this case, a "guilt complex" test should be applied.

For this test, a new list of questions is created about a fictitious crime of a similar nature. If possible, the subject should be lead to believe that the questions are about a real incident. If the subject responds to this test, to a degree comparable to that in the earlier tests, this is indicative of his truthfulness in the matter.

5.2.9 The peak of tension test

In cases where information is known about the crime, which could not be known to the subject if he was innocent, a peak of tension test can be applied. The test

consists of asking a series of questions, in which only one has to do with the incident under investigation.

For instance in our example, suppose that only the police and the perpetrator know that the object stolen is a watch. The list of questions would then be:

1. Was the object that was taken a wallet?
2. Was it a necklace?
3. Was it a purse?
4. Was it a watch?
5. Was it a ring?
6. Was it a briefcase?
7. Was it a necklace?

To stimulate the subject, the examiner might start the test by explaining the effects of this test on a guilty subject. If the subject knows what the object was, his blood pressure will rise up to the question about that object, and then start to drop again.

5.3 Evaluating a polygraph recording

Since what really is being measured is stress, it is not really possible to determine when a lie is being told, but it is possible to see which answer aroused the most stress. A person who is not guilty to the subject under investigation may still be stressed by the relevant questions. However, since the control questions are chosen so that the subject will possibly lie to them, or at least will be uncertain about his answer, this question will probably arouse more stress. On the other hand, a guilty subject would probably also respond to the control question, but since the relevant questions are at this moment much more important to him, these will render more stress than the control question.

In the evaluation of a polygraph recording, the examiner will look for certain deviations in the subject's normal respiration and blood pressure signals. In contrast to blood pressure-pulse tracing, respiration is less apt to be affected by such extraneous factors as a truthful person's nervousness or his concern about some other matter unrelated to the one under investigation. Also, efforts to evade detection by distorting the polygraph recordings are more frequently detectable or reflected in the respiration tracing. A deviation in blood pressure only is not considered enough indication of deception unless it is very consistent. A response in respiration alone can be considered a sufficient indication.

A rise in GSR level is also an indication of stress, but it is generally not considered reliable enough for a definite conclusion and is only used as an additional indication of deception. Still, a response to a relevant question can only be considered deception, if it is greater than the response to the accompanying control question.

In interpreting polygraph records an examiner will focus on the following questions:

- What is the subject's normal breathing pattern?
- What question shows the greatest response in respiration?
- What question shows the greatest response in blood pressure?
- Does the subject show any response in the control question?

- Which responses are greater and more consistent, the crucial questions or the control questions?

As said before the, examiner will look for deviations in the recorded signals. In the paragraphs below a description is given of how normal and deceptive patterns may be recognized.

5.3.1 Effects on respiration

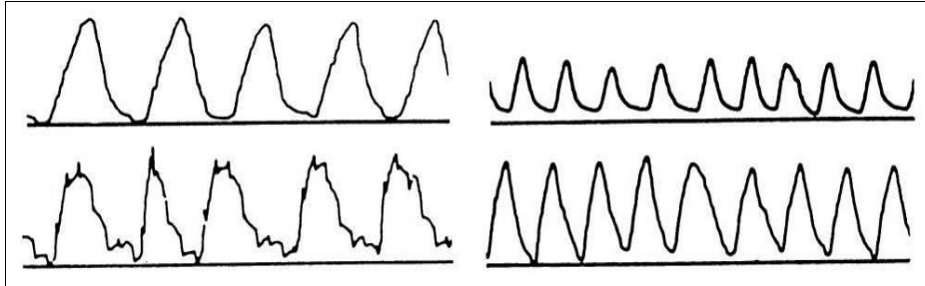


Figure 5.2 Normal respiration patterns

In order to facilitate an understanding of the deception criteria that may be disclosed in the respiration tracings, first an example is shown of normal respiration tracings (Figure 5.2). The serrated nature of the tracings in the lower bottom is no sign of deception but merely of nervousness (the subject is trembling).

Next, a number of illustrations are offered of the various types of responses in respiration that are considered reliable criteria for deception.

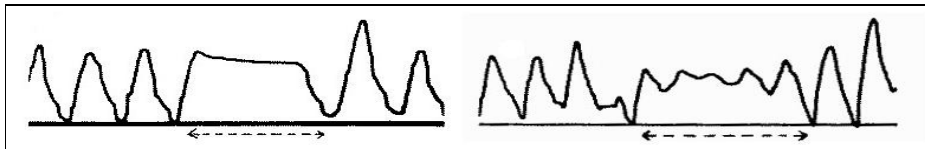


Figure 5.3 Respiratory blocks

A block or stoppage in respiration (Figure 5.3), occurring immediately after a subjects answer to a test question and lasting for several seconds is a very reliable symptom of deception. As seen in the tracing on the right a block doesn't have to occur at the beginning or end of n inhalation, but it can also occur in between.

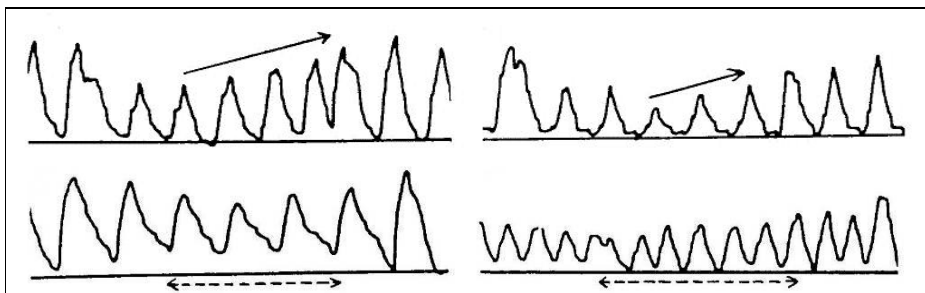


Figure 5.4 Suppression

The upper two tracings in Figure 5.4 show a staircase suppression, whereby the depth of the respiration gradually increases. The lower two tracings show

ordinary suppression. Although not as characteristic and as easily recognisable they are just as reliable criteria of deception.

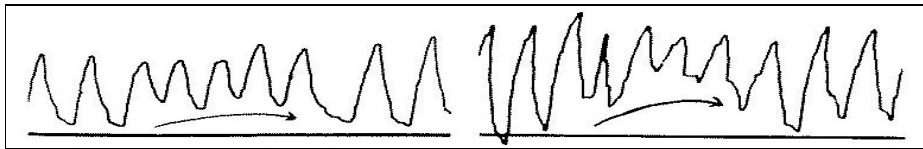


Figure 5.5 Baseline rise

A rise in the respiration baseline is another reliable symptom of deception. As shown in Figure 5.5, the rise usually lasts for 15 or 20 seconds, after which the baseline ordinarily returns to its normal level.

A change in respiration cycle after the asking of a relevant question is another dependable criterion of deception. A widening in cycles (as in the left of Figure 5.6) is indicative of slower breathing while shorter cycles suggest faster breathing.

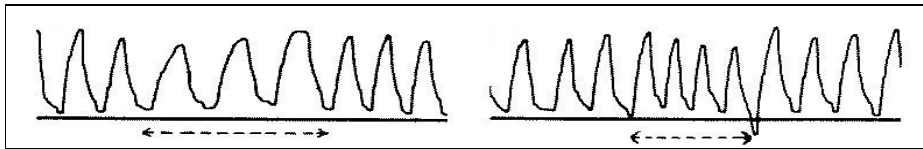


Figure 5.6 Cycle change

Heavier breathing (i.e. a sigh of relief) after a crucial question or even at the end of the test itself and without being preceded by any obvious suppression is another criterion of deception. In the case shown in the left of Figure 5.7 the subjects breathing throughout relevant question 3 appeared quite normal, but after the fourth question was asked the subject gave a sigh of relief in the form of three heavy breaths.

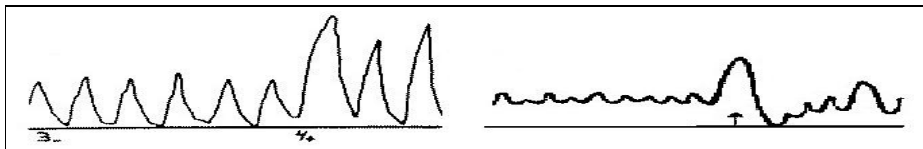


Figure 5.7 Post-deception respiration relief

In addition to the previously illustrated sets of specific responses, deception may be reflected by various forms of erratic breathing during the relevant question interval (as seen in Figure 5.8). Contrary to what is sometimes assumed, erratic breathing is not ordinarily due to deliberate efforts to sabotage the test. It apparently results solely from his own natural disturbance over the fact of lying.

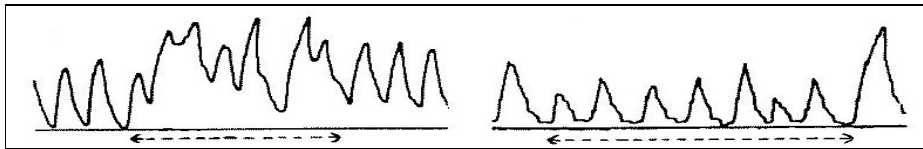


Figure 5.8 Erratic specific

The previously illustrated specific responses may be considered significant only when they appear after the answering of a question, or at most not earlier than a one-cycle interval before the question is answered. If the response occurs prior to

the answering of the question it may be it ordinarily is not indicative of deception and may be considered a pseudo-deception anticipatory response (Figure 5.9).

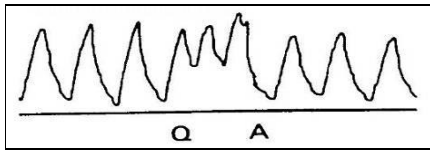


Figure 5.9 Pseudo-deception anticipatory response

5.3.2 Effects on blood pressure

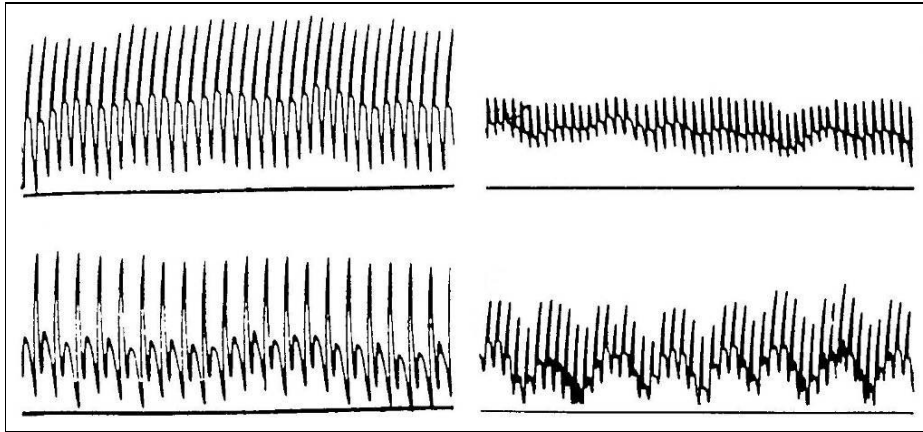


Figure 5.10 Normal blood pressure

Figure 5.10 shows some tracings of normal blood pressure / pulse patterns. Each individual stroke represents a heartbeat. Tracings like this, with the dicrotic notch near the centre, are ideal for deception pressure purposes. It indicates that the cuff pressure is approximately in between the systolic and the diastolic blood pressure.

The fluctuations in the tracing on the lower right are again symptoms of nervousness with the subject. Such subjects are nevertheless suitable for test purposes. As a general rule, any irregularity in blood pressure or pulse that is consistently irregular will not prevent a polygraphic diagnosis.

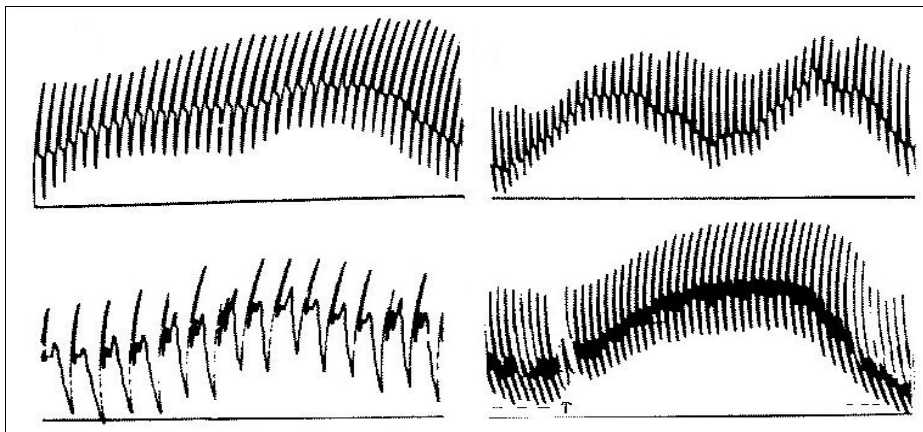


Figure 5.11 Blood pressure baseline rise

Figure 5.11 shows some examples of indications of deception in blood pressure tracings. A typical deception response is indicated by a rise in the base level of the tracing. It reflects an increase in the blood pressure as the crucial question was answered. In some of these tracings a reduction in the pulse amplitude can also be seen. Observe that in the example in the bottom right, the blood pressure was higher before the relevant question than after. This is probably due to the subject's tension in anticipation of the question and his relief thereafter.

In addition to the above mentioned specific response indications of deception, an examiner will sometimes encounter a general indication in the form of a gradual increase in blood pressure tracing up to the relevant question and a drop in blood pressure when the next question is asked. Care must be exercised however that a rise in blood pressure is not due to improper cuff pressure as seen in the lower left example in Figure 5.12.

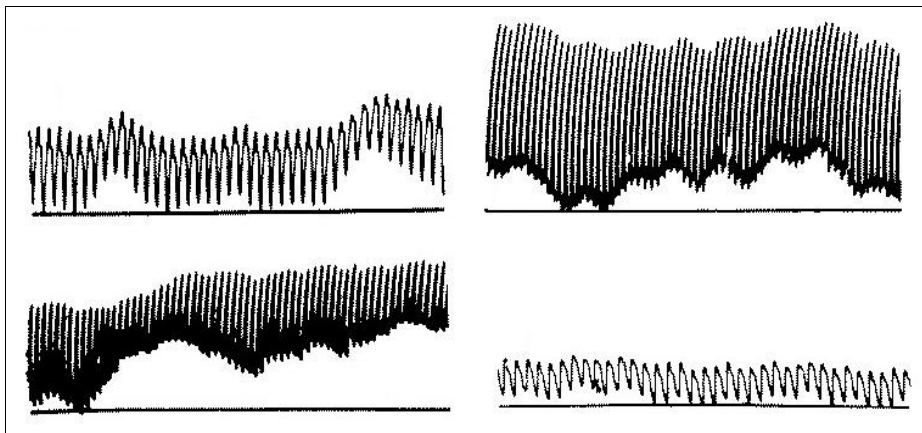


Figure 5.12 Improper cuff pressure

Care must be taken to set the pressure of the blood pressure cuff just right to get an optimal reading. If the pressure is too low, as seen in the upper left example of Figure 5.12, the diastolic notch is visible near the top of the pulse. In the next two examples the pressure is too high. This is uncomfortable for the subject and thus may increase a rise in blood pressure by itself (as in the lower left tracing).

5.4 Conclusions

From the previous paragraph's the following conclusions can be made:

- It is very important that the subject is adequately afraid of being detected as a liar.
- The examiner has a number of techniques to boost the subject's believe in the machine.
- There are more than one ways in how a subject can respond to his stress of lying.
- An important part of the polygraph test is to watch out if the subject is trying to distort the test by manipulation his respiration or by moving.
- The examiner will not only look for signals during the answering of the question but also directly after the asking of the question and after answering
- Deception is indicated when the reactions to the relevant questions are higher then the reactions to he control questions.

6 Experimental design

The first objective of this study was to determine whether evidence of lying could be found in a person's voice. Studies to measure vocal stress in subjects that were requested to lie have been done before ([Horvath '78], [Brenner '79]). Both used the PSE to record stress levels, and both were able to find statistically significant results, although only at high stress levels. [Wees '95] showed that stress may be indicated when pitch is relatively high and jitter is low.

It is very hard to come up with a test in which the subject is sufficiently involved to create a high stress level. The problem with making people lie in a controlled environment is that they know that you know that they are lying and therefore there is no fear of being caught.

The experiment that was finally decided to was a card guessing game in which the subjects were encouraged to lie in order to prevent the examiner from guessing which card they had picked. Afterward he would show the card he had drawn, so that it was known when he had lied.

The problem still exists however that the subject knew that we knew that he lied, and that there were no consequences for him if he got caught. For these reasons the lying might not induce enough stress. Our solution was to convince the subject that we were able to tell when he was lying, so that he would try to outsmart the computer. Actually, just because he did his best to lie as convincingly as possible, the most stress was induced. This is comparable to the techniques used by polygraph examiners.

It was important that the subject was in the right state of mind before the card guessing started. On the one hand he had to be relaxed so that truthful answers would show as truthful, and on the other hand he would have to be pressed to lie convincingly. For these reasons the introductory conversation and the calibration session were quite important.

6.1 Procedure

In order to have significant differences in stress-levels between a truthful and a deceptive answer, the subject had to be sufficiently relaxed at the start of the tasks. For this reason the first ten minutes of the session were spent in explaining to him exactly what the purpose of the experiment was, what was going to happen, and what was expected of him. Most subjects were immediately interested in the project, and they were very enthusiastic to co-operate.

Before the test could begin, it was important to set up the microphone just right, so that samples would be loud and clear enough for analysis. The subject got some time to practice speaking loud enough, and from close to the microphone. At the same time the subject had to be sitting comfortably. Usually the subject sat leaning forward with his elbows on the table, as is shown in Figure 6.1.

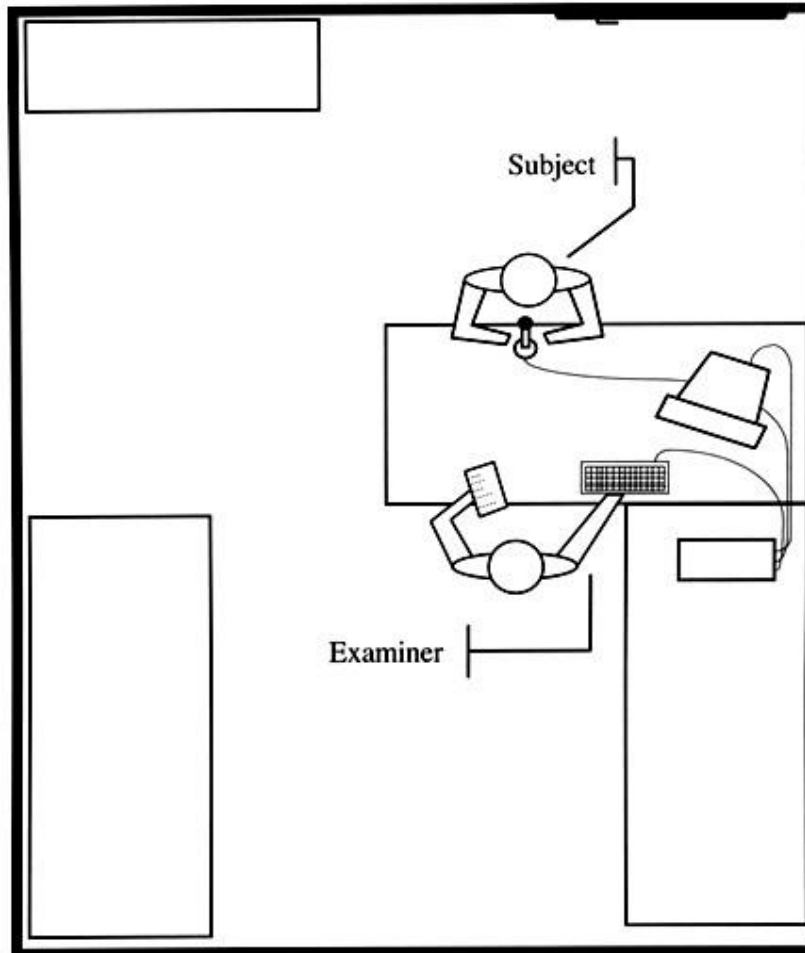


Figure 6.1 Setup experiment room

When all this was ready, the calibration task started, mainly to acquaint the subject with the setting. Then the card-guessing game was played several times, after which the results were discussed with the subject.

The sound samples were recorded as follows:

- When the examiner asked a question, he pressed a key on the computer to start the sampling.
- When the subject had answered the question, the examiner pressed a key to stop the sampling.
- The sample was automatically clipped so that only that part with energy over 1000 was left. This is the part of the sample with actual speech that can be analyzed.
- From what was left, the pitch, the jitter and the zero crossings were calculated and printed on the screen. At the same time the sample and these values were saved to disk.
- In the mean time the examiner marked on the question list what the answer had been, by what name the sample had been saved, and whether he believed the answer had been truthful. The latter in order to make a guess at the end of the task, so that the subject would view the task as a game.

6.2 Calibration questions

Preceding the card-guessing games, a number of calibration questions were asked. This was to obtain sound samples from the subjects at a time that they were not stressed (or at least not lying).

The questions were of a personal nature, and had to be answered truthfully with either "Yes" or "No" ("Ja" or "Nee"). If the subject found the question to be too personal, he was of course allowed not to answer it. This has not occurred though.

The questions that were asked are:

1. Your name is <subject's name>?
2. Your study number is <subject's number>?
3. You live in <subject's place of residence>?
4. Do you live in lodgings?
5. Are you a member of a fraternity?
6. Are you in any committee?
7. Did you get your propaedeutics in one year?
8. Is your study rate nominal?
9. Are you satisfied with your study rate?
10. Have you got a driver's license?
11. Have you ever failed you driver's exam?
12. Have you got al pilots license?
13. Have you ever been to Japan?

6.3 Card guessing

The main part of the experiment was the card-guessing game. Again, questions were asked, but this time the subjects were encouraged to lie. The subject had to pick a card from a deck. He was then asked some questions about that card, which could be answered by "Yes" or "No" (in Dutch respectively "Ja" or "Nee").

In order to maximize the stress at the moment the actual lie is committed, it is very important to give the subject the opportunity to "work up" his stress-level. This can be arranged by asking questions in a fixed pattern, so that the subject knows when the lie will come. When applying a polygraph test, this is a standard procedure. It is expected that stress levels will rise toward the moment of the actual lie, and sharply drop afterwards.

Early on in the experiment it was noticed that people would just randomly answer yes or no. This presents some problems. First of all the subject is apparently not very involved in the game. It is even possible that when he answers, he doesn't even know whether he is telling the truth or not. Furthermore, there was no build up of stress toward a certain point. For these reasons there was a minor change in the procedure. The subject was asked to draw a card, but answer the questions as if he had drawn a specific other card. This would resolve the previous problems, but introduced a new one. If the subject was answering the questions with respect to the "virtual" card, he might consider that one as the real

card, and feel as if he was telling the truth. To overcome that, he was asked to take a good look at the real card, each time before he answered.

This meant that there were now two moments when the subject lied, namely when he was asked about the card he had picked, and when he was asked about the card he had in his mind.

While asking the questions, the examiner noted the answer and the number of the sample, and if the analysis of the sample gave extreme values, he noted this as well. Of course, the examiner did not have to rely solely on the measurements. He could also mark an answer as a possible lie, on the basis of visible clues or even his intuition.

The questions that were asked about the card are:

- Is it a red card?
- Is it a black card?

- Is it of spades?
- Is it of clubs?
- Is it of diamonds?
- Is it of hearts?

- Is it over a seven?
- Is it below an eight?
- Is it over a four?
- Is it below a five?
- Is it over a ten?
- Is it below a jack?

- Is it a ...

| | | | |
|----------|----------|----------|----------|
| • Two? | • Five? | • Eight? | • Jack? |
| • Three? | • Six? | • Nine? | • Queen? |
| • Four? | • Seven? | • Ten? | • King? |

These questions were not always asked in the same order, mainly to prevent the subject from becoming bored, and from anticipating the next question. Usually the questions about the symbol of the card came first, followed by the higher/lower questions. After that came the list of numbers which was asked in order.

At the end of the list the examiner would try to determine which card the subject had picked, based on the recorded results from the analysis. He would then ask the list of numbers again, but in reversed order, and only the numbers around the suspected number.

In the end the examiner told the subject which card he thought the subject had picked. The subject then showed the card, so that it was known to which questions he had lied.

6.4 Subjects and equipment

In total, 46 persons volunteered to be a subject for this experiment. They were all students at the TUD, and followed the course "A415: Technical Applications of Computer Science". This means that over all they were interested in the subject of this experiment, and they co-operated enthusiastically.

Most of them were native speakers of the Dutch language. Only one was a girl.

The equipment used for this experiment was:

- A 80486 based computer running at 66 MHz.
- A SoundBlaster™ 16 ASP soundcard. Sampling was done at 16 bits resolution and 44,1 kHz sampling rate.
- A Shure SD-588 uni-directional microphone.
- A modified version of the VoiceMaster program running under Windows 3.11.

7 VoiceMaster

Earlier in the 'Stress assessment through multimedia' project, it was determined that voice processing software had to be developed. This software should have the ability to record and play sound samples and to automatically extract important features like pitch and jitter. Because the algorithms used to extract these features are not yet very robust the program should also have some editing functions, so that the feature extractions can be manually checked and if necessary adjusted.

VoiceMaster was first developed by [Hoogerdijk '94] and later modified by [Wees '95]. It can be described as an acoustic workbench for recording, playback and off-line analysis. VoiceMaster is an off-line sample analyzing and editing environment. It is capable of recording and playing voice samples. Pre-recorded voice samples can be loaded from disk, viewed on screen and edited. Multiple samples can be loaded at once, making comparison of several samples possible.

The next paragraph offers a small review of the original VoiceMaster program. The paragraph after that describes the changes that were made to VoiceMaster in order to facilitate the analysis of the data gathered in the experiment.

7.1 The original VoiceMaster

7.1.1 Analysis functions

The analysis part is, for voice stress assessment purposes, the most important capability of VoiceMaster. It consists of several feature extraction routines and can be used to construct graphs showing the contour of the features over the length of the sample. This way the variation of the feature values can be examined on screen.

The computed values from the analysis function will be displayed in a graph directly beneath the sample signal. Most features have a time-scale as the horizontal axis just like the sample signal. This makes it possible for the sample signal and the analysis graph to use the same horizontal axis so they can be directly compared. This is a great advantage in visually examining the feature graph. It is also possible to zoom in on any part of the sample, including the analysis graph, which greatly enhances the accuracy.

In order to be usable for stress-related research, the program needed some functions to determine the fundamental frequency and jitter of a speech sample. These functions were built into VoiceMaster 1.2.

Jitter is the perturbation in the vibration of the vocal cords. This results in a cycle to cycle variation of the fundamental frequency. The analysis of this perturbation has gone through some refining in recent years. One has to understand that there are actually three factors that determine the perturbation measures obtained in a laboratory:

1. The physical perturbation created by the vocal folds
2. The perturbation introduced by hardware due to additive noise and signal distortion
3. The perturbation introduced by the algorithms used in the software

The objective is to find the perturbation due to the first factor. Unfortunately the additional undesired perturbations introduced by the second and third factors are often unavoidable and can only be minimized.

The jitter percentage in an utterance is usually very low (< 2%); therefore one has to use a high sampling frequency. [Titze '87] stated that, to extract jitter accurately, the fundamental frequency of the signal must not exceed 0.5% of the sample rate. In other words, there should be about 200 samples per cycle to measure jitter satisfactorily. The technology available to perform the digitization seems adequate; a sampling rate of 44.1 kHz would be sufficient to measure jitter in voices where the fundamental frequency is as high as 220 Hz.

7.1.2 Pitch determination using the Excursion Cycle method

VoiceMaster contains two methods for extracting the fundamental frequency, namely the autocorrelation (AC) and excursion cycle (EC) method. Of these, the Excursion Cycle method is best because it is faster and more precise. Because this algorithm is capable of exactly determining the length of each period in a speech signal, it can also be used to calculate the perturbation of the fundamental frequency period length. This is not the case with the used autocorrelation method as this method averages the number of fundamental periods over a certain window size.

In Figure 4.1 a small segment of a periodic speech sample is shown. With the bare eye the periods are easily distinguished. Some features of the speech signal may attract attention. First of all each period starts with a zero crossing. And secondly, the first peak of each period is the biggest. More precisely put: the energy between the first and the second zero crossing of each period is the highest. This property of voiced speech is used in the Excursion Cycle method to determine the period lengths.

An excursion cycle is defined to consist of all the samples between two consecutive zero crossings ([Hess '83]).

The excursion cycle pitch determination method consists of three steps:

1. Find all excursion cycles and store their starting points and energy within those cycles in a data structure. Also record the positive and negative maximum energy values.
2. Eliminate all excursion cycles that are not the first excursion cycle of a fundamental period, which is called the significant excursion cycle. Elimination of the non-significant excursion cycles is done using the following properties:
 - All significant excursion cycles have the same polarity. So, using the absolute maximum of the positive and negative maximum energy values, all EC's with the opposite sign of this absolute maximum are eliminated.
 - Unvoiced speech segments have significantly lower energy values. Now all EC's of unvoiced speech segments are eliminated using a threshold value. This threshold is chosen at one-tenth of the previous found maximum.
 - Since fundamental frequency generally is lower than 500 Hz, the minimum length of the fundamental period is 2 ms. This means there can never be two significant EC's within a 2 ms range.

Now all EC's that have a higher energy EC in a 2 ms range are eliminated.

- The amplitude of an EC gets lower near the end of a fundamental period and the lower limit of the fundamental frequency is about 50 Hz. So, after dividing the sample in blocks of 20 ms, each non-empty block must contain at least one significant EC. Thus after, in each block, finding the EC with the maximum energy and deleting all other EC's with an energy value beneath 90% of this maximum, it is assumed that only the significant EC's remain.
3. Count the samples between two consecutive EC markers. This number, after checking if it falls in the 50 to 500 Hz fundamental frequency boundary range, can then be converted to either fundamental frequency or fundamental period.

7.1.3 Jitter determination using the Excursion Cycle method

The excursion cycle routine can simultaneously be used to determine the perturbation in the fundamental frequency because the length of each period is known exactly. As with the fundamental frequency there are a number of ways that can be used to determine the jitter. A review of a number of methods is given in [Pinto '90]. In this project the choice has been for the formal definition of perturbation.

Formally the term perturbation implies a deviation from steadiness or regularity. Let a_i be any cyclic parameter (amplitude, pitch period, etc.) in the i^{th} cycle of the waveform. Then, the steady value of this parameter over a span of N cycles can be estimated from its arithmetic mean:

$$(1) \quad \bar{a} = \frac{1}{N} \sum_{i=1}^N a_i$$

and the zeroth-order perturbation function as the arithmetic difference:

$$(2) \quad p_i^0 = a_i - \bar{a}, \quad i = 1, \dots, N,$$

where the superscript gives the order of the perturbation function. Higher-order perturbation function can be obtained by alternately taking backward and forward differences of lower order functions (backward for functions of odd-number order, forward for functions of even order).

For example, the first-order perturbation function is:

$$(3) \quad p_i^1 = p_i^0 - p_{i-1}^0 = a_i - a_{i-1}, \quad i = 2, \dots, N.$$

This first order perturbation function can be used to determine the fundamental frequency perturbation if in (3) a_i is the fundamental frequency.

7.2 Analysing voice samples with VoiceMaster

For this specific study, a number of additional functions were added to VoiceMaster. Because we wanted to measure jitter, we needed the Excursion Cycle method to determine the fundamental frequency and the variations therein. The existing method for determining EC's was not considered very robust, so a couple of functions were developed to visualize and manually edit the EC's.

Furthermore, there were some functions added to write the found fundamental periods and jitter to a file, so that they could be analyzed with other programs.

7.2.1 Clipping samples

During the experiment background noise and non-voice sound made by the subject was recorded. In order to analyse the answer only, the data produced by this sound prior and after the answer should be removed. Therefore the option 'clip on energy' is added. With this option, all sound with energy below a certain threshold is removed.

Sound made during the answer, which was not relevant for the analysis, should be removed as well. In such a case the sample can be clipped manually.

7.2.2 Editing excursion cycles

To determine the average pitch of a sample the exact placement of zero crossings in the sample is less important. The influence of measuring a cycle larger than it should be is counteracted because the next cycle will be measured shorter. However, in order to determine average jitter, the exact placement of the zero crossings is important. This is because jitter is determined by comparing each cycle with its neighbour.

Using VoiceMaster we sometimes found unlikely values for jitter. This may happen because of the way VoiceMaster determines the start of a cycle. This is demonstrated in Figure 7.1.

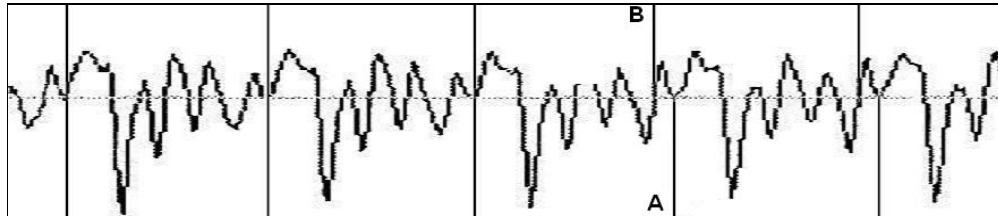


Figure 7.1 Wrongful determination of excursion cycles

At first an EC is chosen as the start of each cycle. After three cycles it can be seen that point A should have been chosen as the starting point for the next cycle. However there is no zero-crossing at this point and therefore point B is selected. The difference in pitch is easily averaged out over the rest of the sample but the effect on jitter is far larger.

As can be seen in Figure 7.2, sometimes the result is much better when the opposite polarisation is used. When this observation was made, a button was added to VoiceMaster to force it to analyse the sample again but this time with the opposite polarisation from the one used last time.

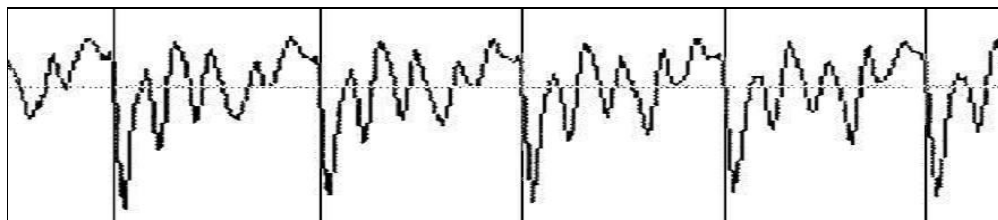


Figure 7.2 Using the opposite polarisation

7.2.3 Input/Output functions

The original version of VoiceMaster could not be used to analyse batches of samples. Only one sample could be opened and analysed at the time. In order to facilitate the processing of batches some extra features were added. The data obtained by the analysis is automatically written to an output file and the next file from the same directory is automatically opened and analysed.

When the sample is clipped, the old file can be overwritten with the adjusted sample.

8 Praat

8.1 Backgrounds

Praat is a widely used general purpose tool to analyze and manipulate digital speech data. It was (and still is) developed by Paul Boersma and David Weenink at the Institute of Phonetic Sciences, University of Amsterdam. Although the first aim of Praat was to give students and scientists of Phonetics a handy tool for manipulating speech data and for creating stimuli for perception experiments, the Praat tool very quickly evolved into a general purpose speech tool.

Praat is a very flexible tool to do speech analysis. It offers a wide range of standard and non-standard procedures, including spectrographic analysis, articulatory synthesis, and neural networks. Praat allows multi-tier labelling, labelling of synchronized multi-channel signals, it may be used to label segments or points in time, and it contains a vast number of analysis tools and algorithms. Praat also contains a script language, numerical tools for optimization, manipulation tools, statistics, a graphical processor to render results into a graphical format, a speech synthesis component and learning algorithms.

8.2 An introduction to Praat

The first thing you'll notice upon starting up Praat is that it is not your typical Microsoft-clone program. There is no "File" pull-down menu with options like "Save" and "Open". This chapter is an adaptation from Katherine Crosswhite's 'Introduction to Praat' ([Crosswhite '02]) and will walk you through some of the most basic aspects of Praat.

8.2.1 Absolute Fundamentals: Opening and Closing Files

Opening a file

To open a file in Praat, go to the "Read" pull-down menu and read in a file. All that will happen is that the name of the file will appear over on the left hand side of the Praat display (in the "Object List"), as a Sound object. If something is listed in the Object List, it means that it is currently stored in Praat's active, working memory. This is what you want. If you have read in a sound file, then you can get almost all the usual functionality of other sound-editing programs simply by clicking the "Edit" button.

Saving a file

Note that the things listed in the Object List are not necessarily files. They are just things that are currently in Praat's memory. To "save" a file, make sure the appropriate object in the Object List is selected (if not, click on it), then go to the "Write" pull-down menu and choose the appropriate option.

Closing a file

In Praat, the equivalent of closing a file is to simply remove it from the Object List. Make sure that the no longer-needed object is selected and click the "Remove" button.

8.2.2 Basic Phonetics Functions: The Edit Window

Basics

If you want to run a quick spectrogram, look at a sound file, etc., then the easiest thing to do is look at it in the Edit window. Use the "Read" pull-down menu to read

in your file, then click the “Edit” button over on the right hand side of the Object List. You will get your typical waveform display. It may or may not also include a spectrogram drawn underneath the waveform. In fact, via the Edit window, you can have access to five different types of phonetic displays: waveform, spectrogram, formant tracking, pitch, and intensity.

Playing

If you look at the bottom of the Edit window, you’ll see either two or three tiers of grey rectangles stretching along the entire bottom edge of the waveform/spectrogram display area. These are playback buttons. There will be one for playing the whole Sound object, one for playing whatever is currently displayed in the window, and others for playing little parts of whatever is displayed in the window. That last type of playback button only shows up if you have clicked somewhere in the Edit display. In that case, there will be two playback buttons located in that tier: one for the portion preceding the cursor and another for the one after it. If you select a portion of the display using the mouse, you will get yet another playback button for that.

Other Edit Window Stuff

You’ll notice a number of other pull-down menus in the Edit window. The “File” menu is pretty straightforward. The “Query” menu is also pretty easy to figure out. The “Select” menu is pretty self-explanatory, but is really mostly useful for script writing – the only thing that you would probably use there is the command to go to the nearest zero-crossing, which is useful if you are cross-splicing something.

8.2.3 Running and Implementing Scripts

The greatest thing about Praat is its scriptability. Praat scripts can do just about anything you would ever want to do. Unfortunately, I am not going to teach you how to write them. The scripting tutorial that comes with Praat is really good, so I refer you to that. What I am going to do is tell you how to run scripts written by someone else, and how to make minor modifications to them.

Running a script

Just download the script that you want to use and put it in some directory that is convenient for you. In Praat, go to the “Control” pull-down menu and select “Open script” to open the file you’ve downloaded. Push control-r to run it.

Modifying a script

You may have to make minor adjustments to a script. One of the most common changes will be specifying what directory your files are in. In connection with this, one must note that there are different ways of referring to directories on different computer platforms. On Mac for example, you use colons wherever you would normally use backslashes in Windows.

Adding a script to the dynamic menu

Sometimes, you would like to be able to run a script more quickly. A good option is to make a button for it in the dynamic menu. The dynamic menu is the official Praat name given to the collection of action buttons arrayed down the right hand edge of the Object List window. It is called “dynamic” because the selection of buttons thus arrayed depends on what object(s) is/are selected at the moment. For example, if a Sound object is selected, you will get a bunch of buttons that are relevant to Sound objects. So if you want to add your script to the dynamic menu, you will first have to consider what type of object is relevant for your script. Then, open the script, following the instructions for *Running a script* given above.

Once the script is open, look at the top of the script window and find the “File” pull-down menu, and choose “Add to dynamic menu” from it. On the following dialogue window, you will have a bunch of fields to fill in.

8.2.4 Doing Resynthesis

Resynthesis is really easy on Praat. To start, just select a Sound object, then find the button labelled “To Manipulation...” from the dynamic menu and click it. Unless you have some reason not to, you can probably just accept the defaults on the ensuing dialog box. You will then see a pitch analysis window flash across your screen, and then you will see a new object appear in your Object List. It will be of the type Manipulation. Then click on the Edit button.

In the Manipulation Editor window, you will see three regions. The first one contains a waveform –the blue marks in that window indicate voicing cycles. You do not change anything in that window. The second region is where you make modifications to pitch, and the third region is where you make modifications to duration. (For changes to intensity or formants, see notes at end of this section.)

Modifying pitch

The pitch manipulation area is pretty straightforward. The x-axis is time, and the y-axis is pitch. The green dots are “pitch points”, and you can drag them around as you will along either axis. Furthermore, you can remove pitch points, or add new ones. A useful tool here is the “Stylize Pitch” commands, found in the “Pitch” pull-down menu. Stylizing the pitch means removing as many pitch points as possible while still maintaining the overall shape; you can specify how many you want to keep. Note that pitch between pitch points is interpolated. This means that if you want a constant pitch at, say, 150 Hz, you will have to have *two* pitch points both set at 150: one indicating where the 150 Hz region should begin and another where it should end, with none in between. Pitch *before* that region will be determined by interpolating between the first 150 Hz point and whatever pitch point immediately precedes it; likewise for the pitch after the 150 Hz region.

Modifying duration

The duration manipulation area is a little less straightforward. The x-axis is time. The y-axis represents what *ratio* you want to use to contract or expand the original sound. For example, the value 0.5 means that you want the new sound to take 50% as much time as the original (i.e., it will be faster). The value 1.5 means that the new sound will take 150% as much times as the original (i.e., it will be slower). To change the duration of something, you will have to add *duration points* in the duration manipulation area. Note that duration between duration points is *also* interpolated, just like for pitch. This means that if you want to speed up or slow down some very specific part of your sound file, you may have to add one or two “extra” duration points.

Playing the resynthesis

The playback buttons at the bottom of the Manipulation Editor window work the same as the playback buttons at the bottom of the regular Edit window, but the resynthesized sound will be heard. Note that there are two resynthesis methods available in Praat: PSOLA and LPC. PSOLA will probably be better for most things, so you should probably use that one (it is the default). However, you can change to LPC using the “Synth” pull-down menu in the Manipulation Editor window.

Using the same menu, you can also set Praat to play back in hums or in pulses.

Playing the original file

To compare your resynthesis with the original sound file, use the playback buttons while holding down shift.

Saving the resynthesized sound file

To get save/export the resynthesized sound, go to the “File” menu of the Manipulation Editor and choose “Publish Resynthesis”. The result of this will be to create a brand new Sound object in the Object List. If you want to save that Sound object, for example, as a .wav file to be used as a stimulus sound file in an experiment, you must write the Sound object to a file

Resynthesis of formants

This can be done in Praat, but the only really easy types of changes are global shifts or scalings, like raising F1 by 50 Hertz. If you want to make more picky changes to formants, you should use Bob McMurray’s KlattWorks program, which works with Praat. You can also use KlattWorks to manipulate the full panoply of usual Klatt synthesis parameters.

8.2.5 Doing Other Types of Analyses

You will notice that Praat has a number of other buttons, commands, and menus not covered here. Many of these are useful in scripting, producing pictures, etc. Many of them have the effect of producing some sort of analysis object. You can make Pitch objects, Formant objects, etc., etc. The basic types of analyses can also be gotten at through the Edit window so you may wonder why you would ever need or want to do the same thing via creation of a separate analysis object.

There are two basic reasons for this. The first is that analysis objects are easier to use in scripting: you create the analysis object, and then you simply query it for the values you want. Faster, easier, and less drain on Praat’s memory. The second reason is that the Edit window only gives you access to the most common and useful analysis algorithms. You may want to find formants using a method other than the Burg algorithm available through the Edit window, for example. A third reason might be that you want to create a specific type of graphic for inclusion in a publication – you can do this by creating a separate analysis object, then drawing it in the Picture window.

9 VoiceBase

The way the experiment was designed resulted in 5447 samples. Per sample information had to be stored about the person, the question and the answer, as well as the results of the analysis of the sample. For this purpose a database was developed. In this chapter the functionality of the database is explained using the database interface as guide. The database can be found on the CD-Rom attached to this report.

9.1 Functions and features

9.1.1 Import samples

Per sample the following information is stored: subject number, session number, question number and the sound file of the sample. To automate this, each sound file was named using the following format: subject#_session#_question#.wav. Per subject a folder was used for all the sound files.

Using the function 'import data' (see Figure 9.1) the database scans a folder, including all it's subfolders, and imports all .wav files that conform to this naming format.

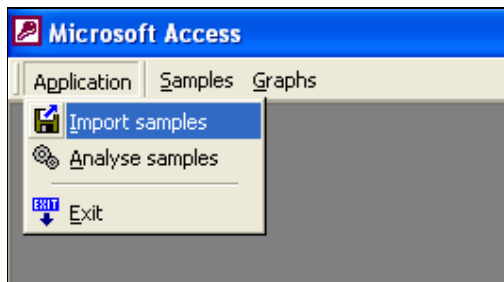


Figure 9.1 VoiceBase Menu

9.1.2 Analysing samples using Praat

The function 'Analyse samples' (Figure 9.1) runs through all samples in the database that have not yet been analysed and submits them to the Praat script one by one.

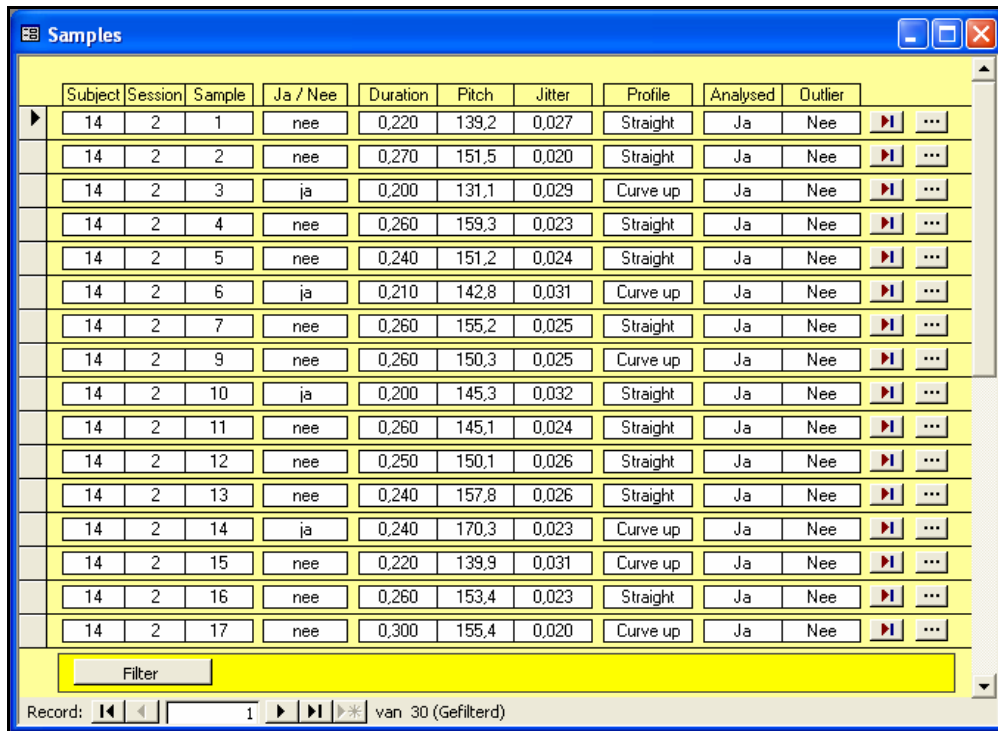
The function works as follows. First the sound sample is written to a temporary file. The program 'praatcon' is then executed with as command line parameters the name of the 'Analyse.praat' script (see appendix A), the path to sound file and the path to the expected output file.

Praat then analyses the file as defined in the script and writes the required data to the specified output file. Subsequently the database imports the output and stores it with the original sample.

Next the script 'CreateGraph.praat' is called with the same sound file as parameter. The script generates the required graphs and writes it to an Figure file. This Figure is subsequently stored in the database together with the other data.

9.1.3 Browse data

The menu function 'Samples' opens the screen that is shown in Figure 9.2. This screen can be used to browse through all the samples in the database.

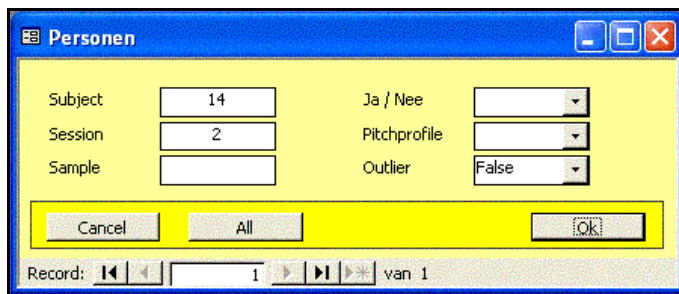


The 'Samples' window displays a table with the following columns: Subject, Session, Sample, Ja / Nee, Duration, Pitch, Jitter, Profile, Analysed, and Outlier. The table contains 17 rows of data for Subject 14, Session 2. Each row includes a play button and a menu icon (three dots).

| Subject | Session | Sample | Ja / Nee | Duration | Pitch | Jitter | Profile | Analysed | Outlier |
|---------|---------|--------|----------|----------|-------|--------|----------|----------|---------|
| 14 | 2 | 1 | nee | 0,220 | 139,2 | 0,027 | Straight | Ja | Nee |
| 14 | 2 | 2 | nee | 0,270 | 151,5 | 0,020 | Straight | Ja | Nee |
| 14 | 2 | 3 | ja | 0,200 | 131,1 | 0,029 | Curve up | Ja | Nee |
| 14 | 2 | 4 | nee | 0,260 | 159,3 | 0,023 | Straight | Ja | Nee |
| 14 | 2 | 5 | nee | 0,240 | 151,2 | 0,024 | Straight | Ja | Nee |
| 14 | 2 | 6 | ja | 0,210 | 142,8 | 0,031 | Curve up | Ja | Nee |
| 14 | 2 | 7 | nee | 0,260 | 155,2 | 0,025 | Straight | Ja | Nee |
| 14 | 2 | 9 | nee | 0,260 | 150,3 | 0,025 | Curve up | Ja | Nee |
| 14 | 2 | 10 | ja | 0,200 | 145,3 | 0,032 | Straight | Ja | Nee |
| 14 | 2 | 11 | nee | 0,260 | 145,1 | 0,024 | Straight | Ja | Nee |
| 14 | 2 | 12 | nee | 0,250 | 150,1 | 0,026 | Straight | Ja | Nee |
| 14 | 2 | 13 | nee | 0,240 | 157,8 | 0,026 | Straight | Ja | Nee |
| 14 | 2 | 14 | ja | 0,240 | 170,3 | 0,023 | Curve up | Ja | Nee |
| 14 | 2 | 15 | nee | 0,220 | 139,9 | 0,031 | Curve up | Ja | Nee |
| 14 | 2 | 16 | nee | 0,260 | 153,4 | 0,023 | Straight | Ja | Nee |
| 14 | 2 | 17 | nee | 0,300 | 155,4 | 0,020 | Curve up | Ja | Nee |

Figure 9.2 Browse sample screen


It is possible to show only a selection of all samples in the previous screen by pressing the 'filter' button and entering the criteria in the screen shown in Figure 9.3



The 'Personen' filter screen contains the following fields and controls:

- Subject: text input field with value '14'
- Session: text input field with value '2'
- Sample: empty text input field
- Ja / Nee: dropdown menu
- Pitchprofile: dropdown menu
- Outlier: dropdown menu with value 'False'
- Buttons: Cancel, All, and Ok
- Record indicator: Record: 1 van 1

Figure 9.3 Filter screen

By double clicking a sample or by pressing the  button the detail screen is opened (shown in Figure 9.4). This screen shows all available data for the selected sample including the graph. The graph shows the sample, the pink vertical lines show the individual excursion cycles, the blue line shows the pitch contour and the brown line the intensity of the speech.

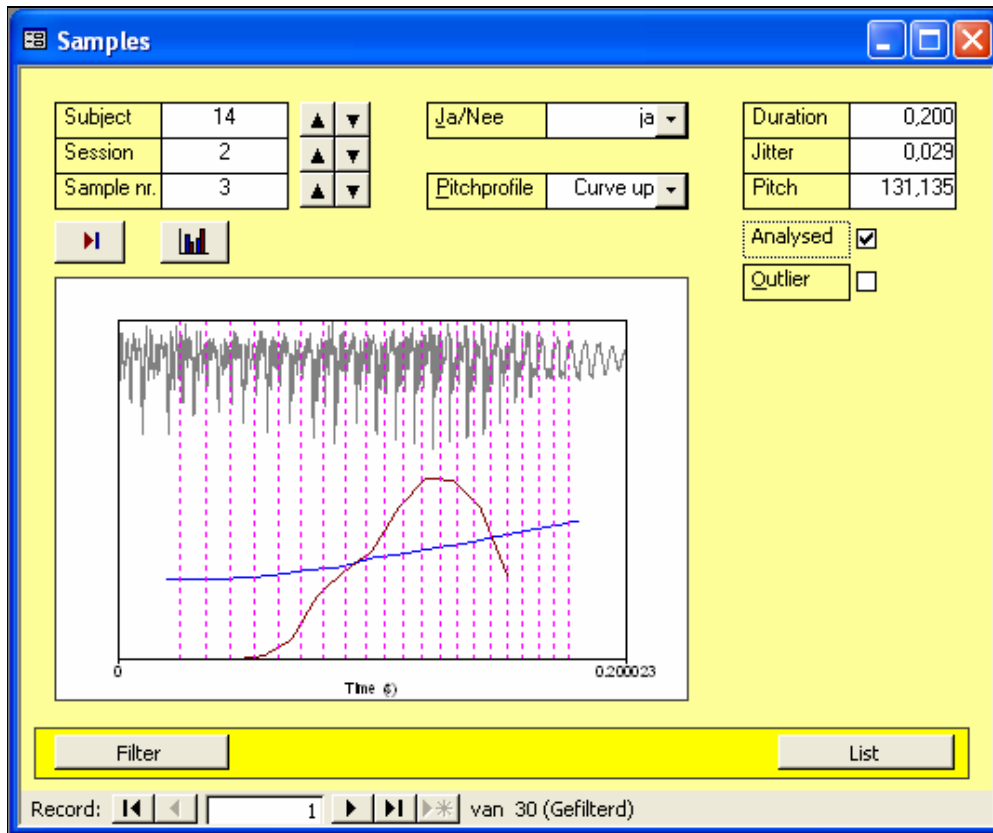


Figure 9.4 Sample detail screen

9.1.4 Generating graphs

VoiceBase has the possibility to generate graphs of selected data in order to perform the appropriate analysis. First the type of graph can be chosen as in Figure 9.5.

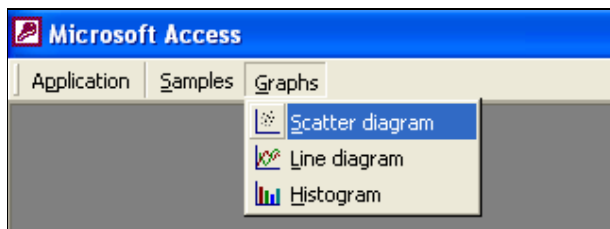


Figure 9.5 Select graph menu

When a scatter diagram is chosen the following screen appears:

Scatter graph

Graph

Graph title: Scatter plot

X - Axis: Pitch

Y - Axis: Jitter

Nr. of series: 2

Series 1

Series name: Series 1

Subject: 17

Session: 2

Sample:

Ja/Nee:

Pitchprofile:

Outlier: False

Series 2

Series name: Series 2

Subject: 24

Session: 2

Sample:

Ja/Nee:

Pitchprofile:

Outlier: False

Series 3

Series name: Series 3

Subject:

Session:

Sample:

Ja/Nee:

Pitchprofile:

Outlier:

Cancel Create graph

Figure 9.6 Generate scatter graph

Up to three data series can be shown in the graph. A series consists of a collection of samples. In this example it's subjects numbers 17 and 24 and for both only session 2.

An example of a scatter graph is Figure 10.3.
An example of a line diagram is Figure 10.4

10 Experimental results

In this chapter the results of the analysis of the voice samples are given. For this analysis the representative data is used. Samples that applied to the following criteria were removed from the set:

- Voice samples with a duration longer than 1 second
In these cases the subject did not give a simple yes or no answer.
- Voice samples with a duration shorter than 0.2 seconds
More than 20 cycles are needed to accurately determine the jitter. Therefore these short samples will no lead to a valid result.
- All test before subject nr. eight
After the first criteria were applied on the data, the data set for these subject were too small for relevant analysis. Apparently these first test were not taken in a consistent manner.

10.1 Difference in analytical tools

As described in chapters 7 and 8, two different analytical tools were available to analyse the data. The first thing we are interested in is whether both tools get the same results for pitch and jitter.

In Figure 10.1 the pitch calculated in Praat is plotted against the pitch calculated in VoiceMaster. All data is centred around a straight line at an angle of 45 degrees, which shows that the results from both tools are nearly identical.

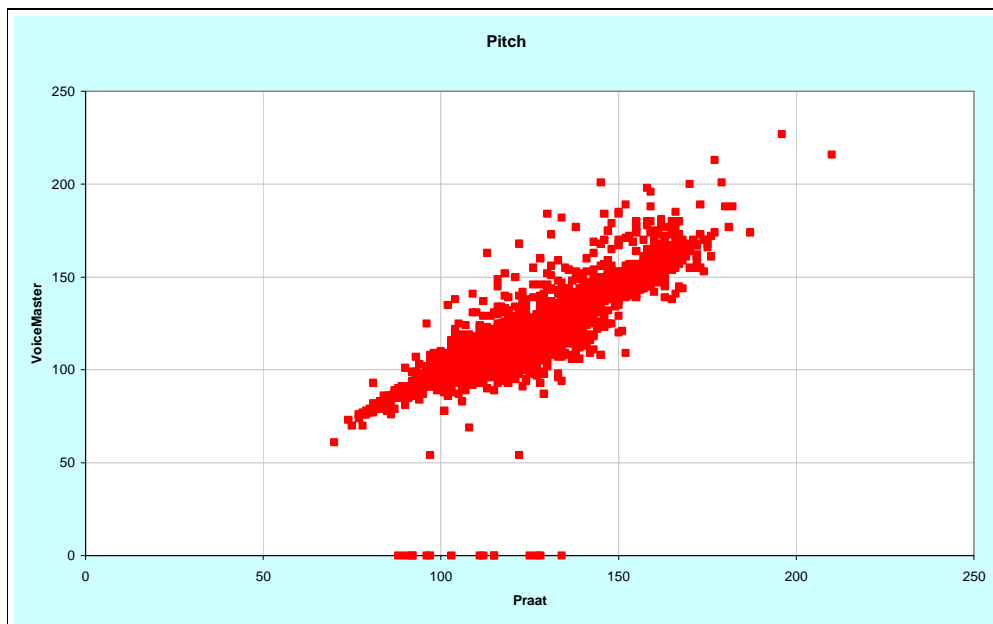


Figure 10.1 Correlation between pitch in Praat and in VoiceMaster

It must be mentioned that Praat allows for five different methods for determining jitter of a sample. The method used, called jitter (local), is defined as: the average absolute difference between consecutive periods, divided by the average period. This is the same definition as the one used in VoiceMaster.

Figure 10.2 shows a similar plot for jitter calculated in Praat and VoiceMaster. In this case there is no evidence of such a straight line. The conclusion is that the results from Praat and VoiceMaster are quite different.

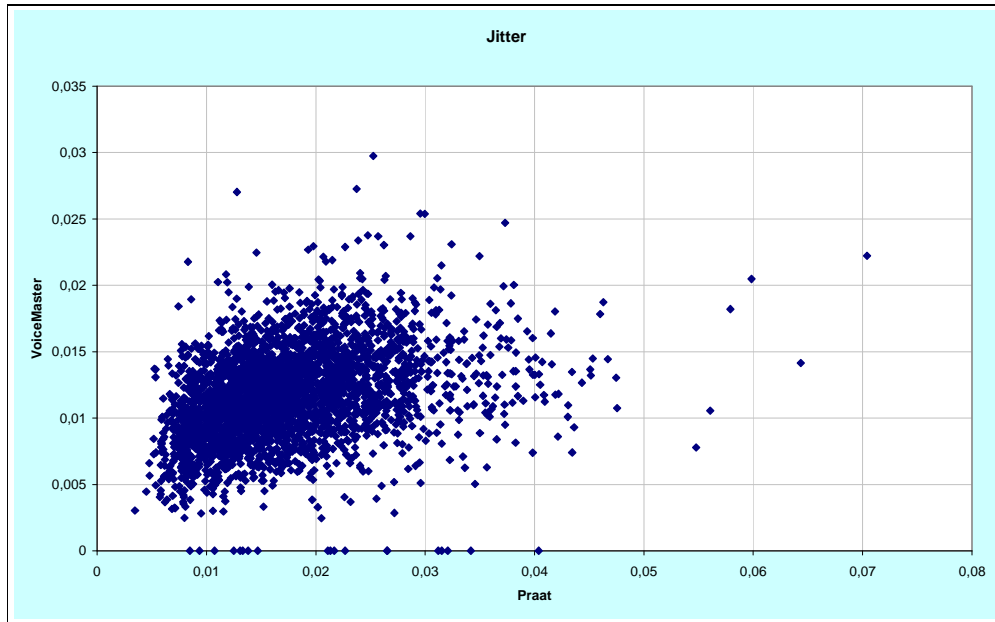


Figure 10.2 Correlation between jitter in Praat and VoiceMaster

Since from these comparisons we cannot determine which of these analyses is more valid, our choice for the rest of the analyses is Praat. This is because Praat is now a generally accepted tool with which a lot of scientific research has been and will be done, while the development of VoiceMaster will probably not be continued.

10.2 General analysis

In order to determine the influence of stress on the voice, samples were taken during a baseline session, session 1. In this baseline session general questions were asked which were all answered truthfully.

In all subsequent sessions the card guessing test was performed. Since in these sessions some of the answers are lies, the average stress levels should be higher than during the baseline session.

Figure 10.3 shows the results of the baseline session against the first card guessing session. It can be seen that the pitch and jitter regions of both sessions cover the same region of the chart.

| | Nr. of samples | Avg. Pitch | St.dev. Pitch | Avg. Jitter | St.dev. Jitter |
|-----------------------------|----------------|------------|---------------|-------------|----------------|
| Session 1 (baseline) | 334 | 118,79 | 20,18 | 0,01964 | 0,00721 |
| Session 2 | 888 | 120,68 | 20,06 | 0,01725 | 0,00670 |
| Session 3 | 982 | 118,18 | 19,24 | 0,01802 | 0,00847 |

Table 10.1 Statistical analysis of sessions one to three

Table 10.1 shows the statistical analysis of both pitch and jitter from the different sessions. Using the T-test it can be determined that the pitch and jitter of the card guessing sessions is not significantly different from the baseline session.

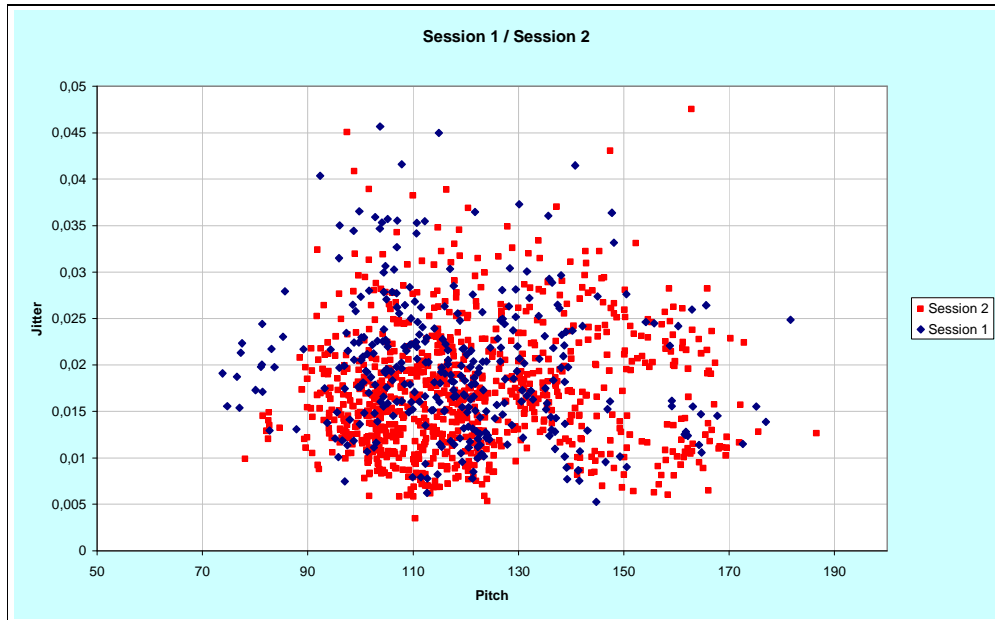


Figure 10.3 Pitch against jitter, sessions one and two

10.3 Voice characteristics for different persons

In the previous paragraph the comparison between the sessions was done using all subjects together. In this paragraph we do the this same analysis for each individual subject.

Figure 10.4 and Figure 10.5 show the pitch and jitter for the different questions in the baseline session, for three different subjects. It is clear that for each subject the baseline level of pitch and jitter is different. This holds that there is no general absolute value which can be taken as a high pitch or low jitter level to indicate stress.

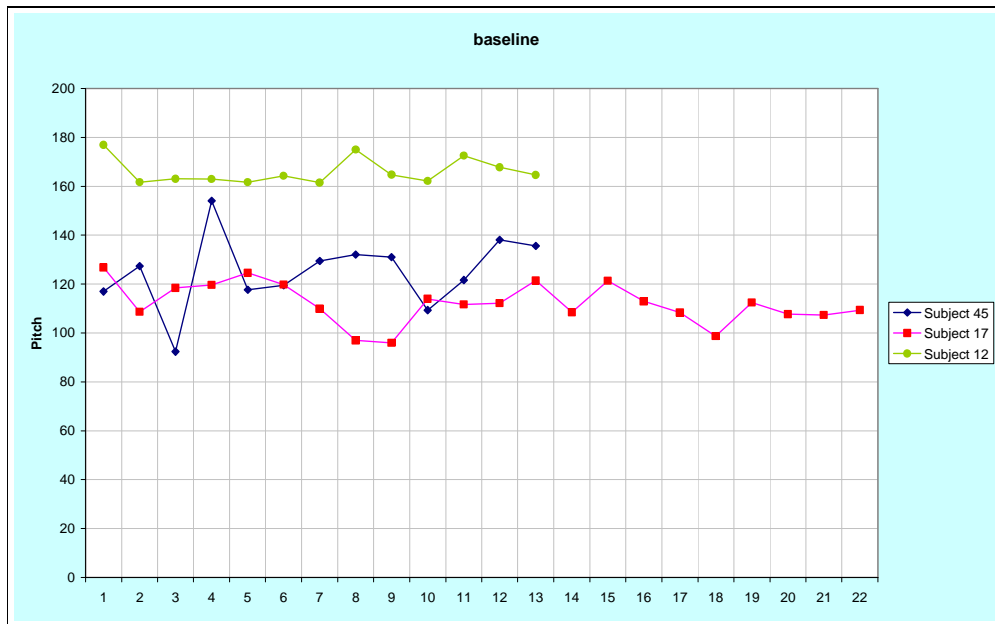


Figure 10.4 Pitch in session one, for three different subjects

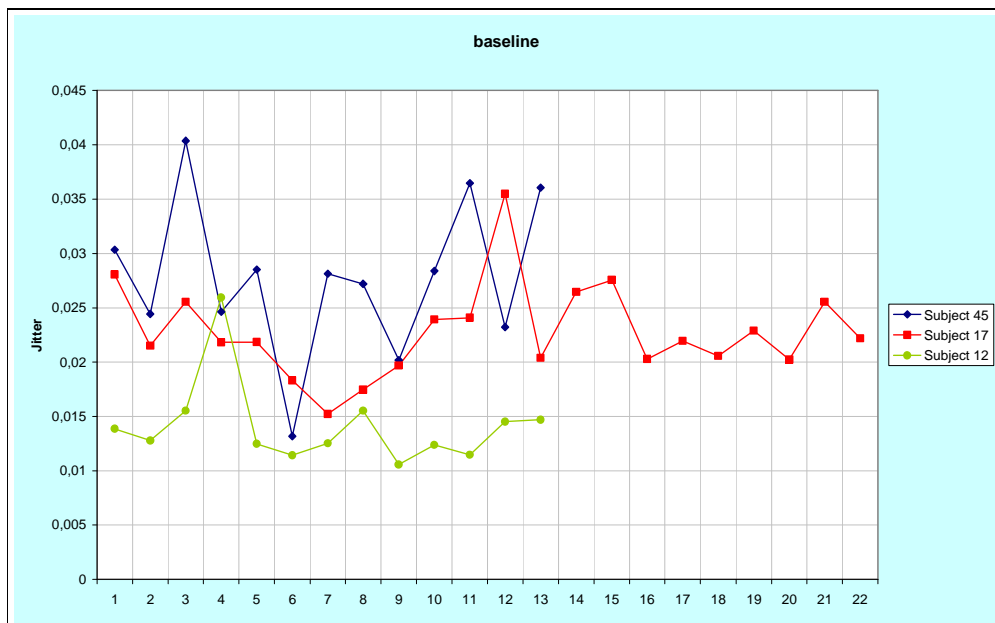


Figure 10.5 Jitter in session one, for three different subjects

10.4 Stress detection per person

In the previous paragraph it is shown that baseline parameters differ per subject. This lead to the assumption that evidence of stress should be examined per subject independently.

Table 10.2 summarises the table from appendix B, which shows the results of a statistical analysis per subject. In this table, an X means that the T-test indicates that the pitch or jitter of sessions two and three are significantly different from the baseline session.

For 12 subjects there is a significant difference for both pitch and jitter for one of the sessions. These sessions are indicated in the table by the coloured background. However, for 7 of these cases, the pitch of the card guessing sessions is lower than for the baseline session. Therefore the difference is not caused by stress. For 1 case, the jitter of the card guessing session is higher than for the baseline session. This difference is also not caused by stress.

For only 4 of the 72 sessions, we get the expected results of a higher pitch and lower jitter.

| Subject | Sessions 1 / 2 | | Sessions 1 / 3 | | |
|---------|----------------|--------|----------------|--------|-------------------|
| | Pitch | Jitter | Pitch | Jitter | |
| 8 | | | X | X | Pitch descending |
| 9 | | | | | |
| 10 | | | | X | |
| 11 | | | X | X | Jitter increasing |
| 12 | | X | | | |
| 13 | X | | X | | |
| 14 | X | X | | X | |
| 15 | X | | | | |
| 16 | | X | X | | |
| 17 | X | X | X | | Pitch descending |
| 19 | | | | | |
| 20 | X | | X | X | Pitch descending |
| 21 | | | | X | |
| 22 | | X | | X | |
| 23 | | | | | |
| 24 | X | X | X | X | Pitch descending |
| 25 | | X | | | |
| 26 | | | X | | |
| 27 | | | | | |
| 28 | | X | X | X | Pitch descending |
| 29 | | | X | X | Pitch descending |
| 31 | | | | X | |
| 32 | X | | | | |
| 33 | | | | | |
| 34 | X | | X | | |
| 35 | | | X | | |
| 36 | | | | | |
| 37 | X | | | | |
| 38 | | X | X | X | |
| 39 | X | | | | |
| 40 | X | X | X | X | |
| 41 | X | | X | | |
| 42 | | | X | X | Pitch descending |
| 43 | X | X | | | |
| 45 | X | | X | | |
| 46 | | | X | | |

Table 10.2 T-test for sessions 1-2 and sessions 1-3 per subject

11 Conclusions and recommendations

As stated in the introduction, this project had the following goals.

- Perform a literature study on stress and the use of the polygraph
- Design and perform an experiment for stress assessment
- Develop a system for automatic stress assessment based on voice analysis
- Analyse the test results
- Draw conclusions based on these results

In this chapter we will check for each of these goals whether they have been made, and what conclusions and recommendations can be made from the results.

11.1 Literature study on stress and the polygraph

Chapter 3 states that a number of parameters can be extracted from the voice. Of these, the parameters most likely to qualify as indicators of stress are mean pitch, pitch perturbation, high frequency energy and speech rate. For this study the choice has been made to research the effects of stress on mean pitch and pitch perturbation (jitter).

From chapter 4 we learn that only the voiced part of the sounds are useful for this kind of analysis.

Chapter 5 shows the use of the polygraph in practise. The most important conclusion that can be drawn here is that in order of having any chance of telling whether a subject is lying the subject has to be adequately afraid to be caught.

Also, there is not one specific reaction for which the machine tests. The interpreter looks just as much for signs that the subject is trying to beat the machine as he looks for physiological effects of stress. The conclusion seems to be that the polygraph is more of a tool for interrogating subjects of a crime rather than a truly scientific analysis of a subjects stress level due to his efforts of deception.

11.2 The experimental setup

As can be read in chapter 10, the experiment that is described in chapter 6 did not result in significantly different stress levels from the baseline readings.

Since other previous studies [Wees '95] have shown that sufficiently high stress levels should be detectable in the voice, the conclusion must be that the lying in our experiment did not lead to high enough stress levels.

A recommendation for further research into this subject is that in subsequent experiments the setup should be such that there is something on the line for the subject. Possibly money or status.

The subject should never consider the experiment as 'just an experiment'. He may not feel that it is ok to be caught lying, as would be the case in a number of games. Lying should feel like cheating. Therefore there may be no encouragement toward the subject to lie. The problem now remains of course: how do we get the subject to lie and how do we know when he lied.

During a polygraph session it is sometimes decided to do a cad-test. Although the literature does not say how it's done, the polygraph interpreter does make sure he knows in advance which card the subject has drawn. All this is done to convince the subject that the polygraph will really work.

In order to distinguish between stress caused by deceptive behavior and stress caused by the investigation, distinction should be made between relevant questions and control questions. A subject may only be labeled as deceptive when the reaction to the control questions is greater than the reaction to the control questions.

11.3 Stress detection through voice analysis

Praat gebruiken ipv VoiceMaster want praat wordt volgens de meest recente inzichten aangepast en heeft ene gebruikers vriendelijk interface. Verder heeft Praat script taal waardoor analyses te automatiseren zijn.

VoiceBase is een goede tool voor voice analysis van grote batches omdat het gekoppeld is aan praat, omdat de grafieken makkelijk inzicht geven in de resultaten en omdat er statistische analyses mogelijk mee zijn.

11.4 Analysis of test results

When this experiment was first setup, the idea existed that we would find certain levels of pitch and jitter that would indicate stress. These levels could be either absolute or relative to the subject's baseline. As can be seen in Figure 10.4 and Figure 10.5, the variation in a person's pitch and jitter is so high that it is unlikely that such a level exists.

Because the baseline levels for pitch and jitter are so different per person, it is not possible to set an absolute value above or below which it can be assumed the subject is experiencing stress. This leaves us two possibilities.

The first option is to determine the baseline levels for pitch and jitter for each subject. This way it is possible to determine stress levels by comparing the subject's pitch and jitter to his baseline levels.

The other possibility is to evaluate pitch and jitter levels continuously and determine the stress level by watching for sudden variations in these levels. This is comparable to the way polygraph sessions are evaluated.

In the evaluation a polygraph session the interpreter not only looks at the stress levels at the time a question is answered, but also right before and after that time. In order for this to be possible for deception of deception through voice analysis, pitch and jitter would have to be measured continuously throughout the session.

Literature

- [Brenner '79] M. Brenner, H.H. Branscomb, G.E. Schwarz, "Psychological Stress evaluator: Two tests of vocal measure", *Psychophysiology*, 1979
- [Cannon '15] W.B. Cannon, "*Bodily changes in panic, hunger, fear and rage*", Appleton-Century-Crofts, 1915
- [Crosswhite '02] K. Crosswhite, "*Introduction To Praat*", December 2002
<http://www.ling.rochester.edu/people/cross/intro-to-praat.pdf>
- [Gray '71] J.A. Gray, "*The psychology of fear and stress*", McGraw-Hill, 1971
- [Griffin '87] G.R. Griffin, C.E. Williams, "*The effects of different levels of task complexity on three vocal measures*", *Aviation, Space and Environmental Medicine*, 1987
- [Hess '83] W. Hess, "*Pitch Determination of Speech Signals*", Springer Verlag, 1983
- [Hollien '90] H. Hollien, "*The acoustics of crime, the new science of forensic phonetics*", Plenum Press, 1990
- [Horvath '78] F. Horvath, "*An experimental comparison of psychological stress evaluator and the galvanic skin response in detection of deception*", *Journal of Applied Psychology*, 1978
- [Hoogerdijk '94] J.W. Hoogerdijk, "*Non-verbal voice analysis*", TU Delft TWAIO report, 1994
- [Lazarus '79] R.S. Lazarus, "*Psychological stress and coping processes*", Raven Press, 1979
- [Poulton '83] A.S. Poulton, "*Microcomputer Speech Synthesis and Recognition*", Sigma Technical Press, 1983
- [Pinto '90] N.B. Pinto, I.R. Titze, "*Unification of perturbation measures in speech signals*", *Journal of the Acoustical Society of America*, 1990
- [Reid '77] J.E. Reid, F.E. Inbau, "*Truth and Deception: the polygraph ("lie-detector") technique*" 2nd ed., Williams & Wilkins, 1977
- [Schachter '75] S. Schachter, "*Cognition and peripheralist-centralist controversies in motivation and emotion*" in "*Handbook of Psychology*", Academic Press, 1975
- [Scherer '82] K.R. Scherer, "*Handbook of methods in non-verbal behaviour research*", Cambridge University Press, 1982
- [Scherer '89] K.R. Scherer, "*Vocal measurement of emotion*", "Emotion, theory, research and experience", Academic Press, 1989

- [Selye '56] H. Selye, "*The stress of life*", 1956
- [Titze '87] I.R. Titze, Y. Horii, R.C. Scherer, "*Some technical considerations in voice perturbation measurements*", *Journal of speech and hearing research* Vol. 30, 1987
- [Vark '93] R.J. van Vark, "*Knowledge based behaviour feedback system using physiological data*", TU Delft graduation thesis, 1993
- [Wees '95] J.W.A. van Wees, "*Voice stress analysis: Using non-verbal voice analysis in automatic stress assessment*", TU Delft graduation thesis, 1995

Appendix A: Praat scripts called from VoiceBase

File: Analyse.praat

```
form Analyseer Soundfile
  sentence input
  sentence output
endform

call analyseer 'input$' 'output$'

procedure analyseer input$ output$

  if fileReadable (output$)
    filedelete 'output$'
  endif

  Read from file... 'input$'
  aSound$ = selected$("Sound")
  To PointProcess (periodic, cc)... 60 300
  aPoint$ = selected$("PointProcess")

  aWaarde = Get duration
  fileappend 'output$' Duration: 'aWaarde' 'newline$'
  aWaarde = Get jitter (local)... 0 0 0.003 0.02 1.3
  fileappend 'output$' Jitter: 'aWaarde' 'newline$'

  select PointProcess 'aPoint$'
  Remove
  select Sound 'aSound$'
  To Pitch... 0 60 300
  aPitch$ = selected$("Pitch")

  aWaarde = Get mean... 0 0 Hertz
  fileappend 'output$' Pitch: 'aWaarde' 'newline$'

  aNum = Get number of frames
  aStart = Get time from frame number... 1
  aEind = Get time from frame number... 'aNum'

  for aTeller from 0 to 4
    aTijd = aStart + ((aTeller * 0.25) * (aEind - aStart))
    aWaarde = Get value at time... 'aTijd' Hertz Linear
    aLabel = 25 * aTeller
    fileappend 'output$' Pitch_'aLabel': 'aWaarde' 'newline$'
  endfor

  select Pitch 'aPitch$'
  Remove
  select Sound 'aSound$'
  Remove

Endproc
```

File: CreateGraph.praat

```
form Maak Grafiek uit Soundfile
  sentence input
  sentence output
endform

call maakplaatje 'input$' 'output$'

procedure maakplaatje input$ output$

  Erase all
  Read from file... 'input$'
  mpSound$ = selected$("Sound")
  To Pitch... 0 60 300
  mpPitch$ = selected$("Pitch")
  To PointProcess
  mpPoint$ = selected$("PointProcess")
  select Sound 'mpSound$'
  To Intensity... 75 0
  mpIntensity$ = selected$("Intensity")

  select Sound 'mpSound$'
  Viewport... 0 6 0 2
  Grey
  Draw... 0 0 0 0 no

  select Pitch 'mpPitch$'
  Viewport... 0 6 1 4
  Blue
  Draw... 0 0 0 300 no

  select Intensity 'mpIntensity$'
  Viewport... 0 6 1.5 4
  Maroon
  Draw... 0 0 0 0 no

  select PointProcess 'mpPoint$'
  Viewport... 0 6 0 4
  Magenta
  Draw... 0 0 yes

  Write to Windows metafile... 'output$'

  select Sound 'mpSound$'
  Remove
  select Pitch 'mpPitch$'
  Remove
  select PointProcess 'mpPoint$'
  Remove
  select Intensity 'mpIntensity$'
  Remove

Endproc
```

Appendix B: T-test for significant difference between sessions 1, 2 and 3

| Subject | Session 1 | | | | | | Session 2 | | | | | | Session 3 | | | | | | T-Test Sessions 1 / 2 | | T-Test Sessions 1 / 3 | | | |
|---------|------------|-----------|-------------|------------|--------------|--|------------|-----------|-------------|------------|--------------|--|------------|-----------|-------------|------------|--------------|--|-----------------------|--|-----------------------|--|------------------|-----------------|
| | Nr. of Smp | Avg Pitch | StDev Pitch | Avg Jitter | StDev Jitter | | Nr. of Smp | Avg Pitch | StDev Pitch | Avg Jitter | StDev Jitter | | Nr. of Smp | Avg Pitch | StDev Pitch | Avg Jitter | StDev Jitter | | | | | | | |
| 8 | 12 | 108 | 4.310 | 1.83% | 0.00447 | | 22 | 107 | 5.945 | 1.80% | 0.00344 | | 11 | 100 | 7.778 | 1.44% | 0.00234 | | | | | | Pitch descending | |
| 9 | 12 | 143 | 9.036 | 1.81% | 0.00411 | | 32 | 139 | 5.087 | 1.44% | 0.00344 | | 26 | 146 | 4.705 | 1.42% | 0.00320 | | | | | | | |
| 10 | 5 | 102 | 2.421 | 2.11% | 0.00480 | | 23 | 107 | 7.217 | 1.88% | 0.00770 | | 26 | 104 | 3.391 | 1.36% | 0.00433 | | | | | | | |
| 11 | 2 | 95 | 0.876 | 1.83% | 0.00477 | | | | | | | | 15 | 114 | 7.654 | 4.07% | 0.00858 | | | | | | | Jitter (3) > 4% |
| 12 | 13 | 166 | 5.365 | 1.41% | 0.00389 | | 26 | 165 | 6.605 | 1.19% | 0.00247 | | 31 | 162 | 7.970 | 1.23% | 0.00236 | | | | | | | |
| 13 | 11 | 116 | 8.286 | 2.03% | 0.00352 | | 29 | 125 | 9.508 | 2.07% | 0.00443 | | 32 | 133 | 7.313 | 2.01% | 0.00387 | | | | | | | |
| 14 | 7 | 126 | 20.051 | 3.28% | 0.00499 | | 30 | 149 | 8.687 | 2.56% | 0.00330 | | 29 | 131 | 8.181 | 2.65% | 0.00373 | | | | | | | |
| 15 | 4 | 134 | 4.797 | 1.95% | 0.00352 | | 14 | 147 | 10.361 | 2.06% | 0.01016 | | 18 | 141 | 9.214 | 1.91% | 0.00889 | | | | | | | |
| 16 | 9 | 119 | 12.553 | 2.26% | 0.00773 | | 22 | 114 | 6.188 | 1.81% | 0.00343 | | 17 | 108 | 5.630 | 2.32% | 0.00393 | | | | | | | |
| 17 | 22 | 112 | 8.364 | 2.28% | 0.00429 | | 25 | 102 | 7.433 | 1.77% | 0.00600 | | 24 | 100 | 5.013 | 2.45% | 0.00527 | | | | | | | |
| 19 | 4 | 119 | 2.067 | 1.41% | 0.00290 | | 5 | 122 | 2.564 | 1.38% | 0.00233 | | 15 | 122 | 8.041 | 1.88% | 0.00588 | | | | | | | |
| 20 | 11 | 108 | 8.649 | 2.15% | 0.00401 | | 38 | 104 | 4.794 | 2.01% | 0.00739 | | 37 | 104 | 2.274 | 1.85% | 0.00377 | | | | | | | |
| 21 | 11 | 140 | 12.718 | 2.51% | 0.00346 | | | | | | | | 32 | 140 | 12.340 | 2.29% | 0.00384 | | | | | | | |
| 22 | 7 | 105 | 5.087 | 1.41% | 0.00284 | | 26 | 104 | 1.882 | 1.04% | 0.00233 | | 32 | 104 | 3.028 | 0.97% | 0.00241 | | | | | | | |
| 23 | 8 | 120 | 12.403 | 2.28% | 0.00209 | | 18 | 113 | 15.903 | 2.34% | 0.00558 | | | | | | | | | | | | | |
| 24 | 11 | 100 | 5.875 | 2.35% | 0.00691 | | 34 | 96 | 4.206 | 1.79% | 0.00327 | | 35 | 93 | 3.920 | 1.80% | 0.00398 | | | | | | | |
| 25 | 10 | 123 | 1.452 | 1.11% | 0.00202 | | 25 | 121 | 2.773 | 0.96% | 0.00192 | | 29 | 122 | 2.381 | 1.10% | 0.00318 | | | | | | | |
| 26 | 4 | 113 | 0.958 | 0.97% | 0.00389 | | 26 | 111 | 1.519 | 0.84% | 0.00246 | | 38 | 108 | 1.744 | 0.68% | 0.00347 | | | | | | | |
| 27 | 4 | 103 | 3.334 | 2.55% | 0.00697 | | 10 | 101 | 4.273 | 1.82% | 0.00673 | | 15 | 105 | 4.912 | 1.93% | 0.01088 | | | | | | | |
| 28 | 3 | 121 | 13.387 | 2.50% | 0.01169 | | 31 | 117 | 4.976 | 1.72% | 0.00338 | | 32 | 112 | 3.124 | 1.60% | 0.00427 | | | | | | | |
| 29 | 12 | 131 | 4.812 | 1.95% | 0.00443 | | 33 | 131 | 5.210 | 1.81% | 0.00817 | | 34 | 128 | 3.502 | 1.42% | 0.00200 | | | | | | | |
| 31 | 7 | 129 | 35.899 | 2.34% | 0.00561 | | 24 | 121 | 10.692 | 1.99% | 0.00671 | | 27 | 122 | 12.175 | 1.62% | 0.00406 | | | | | | | |
| 32 | 13 | 111 | 4.637 | 2.49% | 0.01022 | | 26 | 114 | 2.759 | 2.25% | 0.00640 | | 32 | 111 | 5.460 | 2.25% | 0.00847 | | | | | | | |
| 33 | 8 | 139 | 16.306 | 1.63% | 0.00516 | | | | | | | | 4 | 125 | 2.182 | 1.12% | 0.00298 | | | | | | | |
| 34 | 2 | 117 | 8.543 | 2.18% | 0.00206 | | 31 | 95 | 4.676 | 1.81% | 0.00651 | | 35 | 93 | 3.010 | 1.41% | 0.00578 | | | | | | | |
| 35 | 4 | 104 | 5.945 | 1.63% | 0.01052 | | 26 | 102 | 2.967 | 1.45% | 0.00662 | | 27 | 97 | 3.449 | 1.72% | 0.00756 | | | | | | | |
| 36 | 3 | 120 | 8.438 | 2.61% | 0.00434 | | 15 | 115 | 5.827 | 2.50% | 0.00454 | | 37 | 120 | 7.766 | 3.20% | 0.00860 | | | | | | | |
| 37 | 10 | 107 | 7.213 | 1.88% | 0.00329 | | 74 | 138 | 20.349 | 1.97% | 0.00426 | | | | | | | | | | | | | |
| 38 | 14 | 101 | 6.022 | 2.12% | 0.01093 | | 15 | 100 | 2.710 | 1.31% | 0.00215 | | 21 | 111 | 3.429 | 1.31% | 0.00376 | | | | | | | |
| 39 | 12 | 119 | 5.886 | 1.41% | 0.00465 | | 23 | 123 | 4.928 | 1.45% | 0.00389 | | 29 | 118 | 4.702 | 1.67% | 0.00594 | | | | | | | |
| 40 | 12 | 123 | 6.853 | 1.87% | 0.00345 | | 35 | 112 | 3.447 | 1.11% | 0.00335 | | 36 | 117 | 3.445 | 1.35% | 0.00539 | | | | | | | |
| 41 | 13 | 145 | 12.014 | 1.22% | 0.00689 | | 34 | 153 | 8.064 | 1.00% | 0.00244 | | 32 | 157 | 7.274 | 1.27% | 0.00280 | | | | | | | |
| 42 | 9 | 100 | 5.058 | 1.79% | 0.00522 | | 4 | 96 | 0.426 | 1.81% | 0.00648 | | 30 | 95 | 4.053 | 1.32% | 0.00376 | | | | | | | |
| 43 | 16 | 80 | 3.628 | 1.98% | 0.00374 | | 10 | 84 | 3.575 | 1.23% | 0.00146 | | 6 | 81 | 2.737 | 2.73% | 0.01655 | | | | | | | |
| 45 | 13 | 125 | 15.047 | 2.78% | 0.00721 | | 41 | 137 | 9.689 | 2.38% | 0.00617 | | 40 | 142 | 14.307 | 2.51% | 0.00737 | | | | | | | |
| 46 | 14 | 126 | 5.288 | 1.27% | 0.00189 | | 14 | 124 | 6.137 | 1.17% | 0.00285 | | 24 | 122 | 5.127 | 1.34% | 0.00495 | | | | | | | |