Universität Trier – Fachbereich I – Psychobiologie

**Dissertation zur Erlangung des**

**Doktorgrades der Naturwissenschaften (Dr. rer. nat.)**

# The vocal fundamental frequency in psychophysiological research: affective F0 modulation induced by psychophysiological methods

Autor:

**Debora Elisabeth Plein**

**Eingereicht im August 2014**

Gutachter:

**Prof. Dr. med. Hartmut Schächinger**

**Dr. Ewald Naumann**

This dissertation thesis and the presented research were performed at the

**Department of Clinical Psychophysiology**

**Institute of Psychobiology – University of Trier**

**Affiliation of the Supervisors:**

**Prof. Dr. med. Hartmut Schächinger**

Department of Clinical Psychophysiology – Institute of Psychobiology – University of Trier

**Dr. Ewald Naumann**

Department of Psychology – University of Trier

The reported research was in part supported by the:

**International Research Training Group "Psychoneuroendocrinology of Stress: From Molecules and Genes to Affect and Cognition".**

Funded by the German Research Foundation (Deutsche Forschungsgemeinschaft: DFG), project GRK 1389/1.

Für Omi

„Wer weiß, wofür es gut ist...“

Rosemarie Plein

Acknowledgments

Without the interest, help, inspiration and humour of other people, this thesis wouldn’t exist. My deep thanks are dedicated to every single person who contributed to this work.

I thank Prof. Dr. Hartmut Schächinger for his encouragement, scientific advice, personal support and creativity. Also, my thanks is dedicated to Dr. Ewald Nauman for his interest, advice and support.

I thank Prof. Dr. Angelika Braun for her enthusiasm and help, especially for supporting my interdisciplinary interests in my early university years during my studies of Phonetics at the University of Trier. Her encouragement set the roots for my interest in professional science.

Also I would like to thank my colleagues for their steady support and fruitful discussions, even at rather unusual times, especially Dr. rer. nat. André Schulz, Diana Ferreira de Sá, Dr. rer. nat. Linn Kühl, Dr. rer. nat. Johanna Lass-Hennemann and Dr. Dr. med. rer. nat. Steffen Richter. Carina Zech, Debbinski always appreciated your smart and creative contributions.

For proofreading and moral support, my deep thanks goes especially to Diana Feirrera de Sá, Carina Zech and my dear aunt Catherine Woodgold. Dr. Dr. med. Steffen Richter was the best medical support one can imagine for a scientific endeavour such as this work.

Without Immo Curio as “lab angel“ this work surely would have been impossible, so thank you so much Immo! Also I thank Anke Ehses for her reliable help in data collection, Carolin Reichert for her great work with E Prime computing and everyone else involved in making this thesis possible.

Last but not least, I thank my family. My deep love is dedicated to my father, my mother and my grandparents, who always supported me, each in his or her own way, as well as my sister and my uncles, aunts and cousins just for being like they are. Also Mama and Papa Schacht, spasibo! Marat: you know.

## General Abstract

Every event in our daily lives triggers an emotional response. We listen to a beautiful piece of music, we are stuck in a traffic jam or win a tennis match. One way of expressing emotions is via the acoustic channel. Darwin described the voice as an "iconic affect signaling device" (Darwin 1998/1872). The influence of affect on vocal parameters has been well investigated in speech portrayed by actors, but little is known about affect expression in more natural or authentic speech behavior. This is partly due to the difficulty of generating speech samples that represent authentic expression of speaker affect.

The present work investigates the influence of speaker affect on the vocal fundamental frequency (F0) in comparatively authentic speech samples. Three well-documented psychophysiological research methods were applied for the induction of affective states in German native speakers in order to obtain speech samples with authentic affect expression: the Cold Pressor Test (CPT), the Stroop Color-Word Test (SCWT) and the presentation of slides from the International Affective Pictures System (IAPS).

The first chapter of this work describes the basic psychophysiological influences on F0 level. It discusses current challenges in the research of vocal affect expression and provides a conceptual background for the presented studies. The end of this chapter gives a short summary of findings and concluding remarks.

In the next three chapters (II – IV), three studies are presented in the form of publication-oriented manuscripts. The references for all chapters are listed at the end of this thesis. In the study presented in chapter II, we found that Cold Pressor stress moderately increased F0 levels in read speech. In chapter III, a study on speech produced in a SCWT is reported where elevated F0 levels in color-word/-ink color incongruent stimuli were found. Chapter IV provides a study that compares spontaneous and standardized speech samples obtained during the presentation of emotional slides from the "International Affective Pictures System" (IAPS). The results of this study show that affect modulated F0 levels in both, standardized and spontaneous speech samples.

The overall aim of this thesis is to meet the challenge of obtaining speech samples that represent authentic expression of speaker affect in F0 by using established psychophysiological methods of affect induction. The results here reported also indicate that this approach may be useful for future research and further to gain a deeper understanding of authentic vocal affect expression. Moreover, F0 may constitute an additional non-

invasive, easy to obtain measure for the established psychophysiological research methodology.

Content

## Index of Studies/ Contributions of others

This doctoral thesis consists of three chapters (and one general introductory chapter) written in the form of scientific research manuscripts. The author of this dissertation is the first author of all three manuscripts. However, as usual in psychophysiological research, other persons (listed in the table as "contributors") also contributed to the work by helping with data acquisition and laboratory preparation of raw data for analyses and would, therefore, qualify to be co-authors if the manuscripts were to be accepted for publication in a scientific journal.

***Content***

| | |
|---|---|
| Chapter II | Study I: Stroop conflict increases fundamental frequency of the voice.<br>Contributors: Diana S. Ferreira de Sá, André Schulz, Hartmut Schächinger |
| Chapter III | Study II: Modulation of vocal fundamental frequency in spontaneous and standardized speech samples during affective picture presentation.<br>Contributors: André Schulz, Carina Zech, Diana S. Ferreira de Sá, Hartmut Schächinger |
| Chapter IV | Study III: Cold Pressor Stress increases vocal fundamental frequency.<br>Contributors: André Schulz, Linn Kühl, Steffen Richter, Hartmut Schächinger |

## Index of Abbreviations

| | |
|---|---|
| Ag/AgCl | Silver/Silver chloride |
| ANOVA | Analysis of variance |
| CPT | Cold Pressor Test |
| ECG | Electrocardiogram |
| e.g. | For example |
| F0 | Vocal fundamental frequency |
| Hz | Hertz |
| IAPS | International affective picture system |
| ms | Milliseconds |
| RT | Reaction time |
| SAP | Systolic Arterial pressure |
| mmHg | Millimeters mercury |
| SCWT | Stroop Color-Word Test |

# Chapter I – General Rationale

Why investigate the voice? The answer is simple: because the sound characteristics of the voice have a fundamental influence on our daily social interactions. Our voices are signal carriers for diverse information, such as linguistic, anatomical and psychophysiological. On the one hand, when hearing a voice, the speech receiver immediately forms impressions e.g. of the sex, age, language and emotional state of the speaker (Eckert & Laver, 1994). On the other hand, the speaker may express a real or feigned emotional state, or a mixture of both.

Amongst the acoustic features of the voice, the voice pitch is probably one of the most salient cues not only for linguistic information, but also for paralinguistic information, such as emotion (Eckert & Laver, 1994; Laver, 1994; Ohala, 1984, 1994). The vocal fundamental frequency (F0) is the quantifiable acoustic counterpart and major correlate of perceived voice pitch. It is also one of the most frequently investigated vocal parameters in studies on vocal expression of emotion, in both animal and human studies (Ohala, 1984; Scherer, 2013).

## 1.1 Basic determinants of F0 level

F0 is the acoustic correlate of vocal fold vibration rate and commonly measured in Hz (Laver, 1994; Ohala, 1984). F0 variation is related to the perceived speech melody, e.g. low F0 variation is usually perceived as monotone voice (Barrett & Paus, 2002; Titze, 1994). From a physiological point of view, F0 level depends on several parameters. One is the length of the vocal folds which can be intentionally modulated to a certain extent. It also depends on the muscular tension of the vocal folds (cricothyroid and thyroarytaenoid laryngeal muscles) and on the subglottal air pressure (thoracic and abdominal muscles involved in respiration) (Anders et al., 1988; Atkinson, 1978; Hoh, 2005; Laver, 1994). A shortening of the vocal folds results in increased F0. Similarly, increased tension of the thyroarytaenoid as well as increased subglottal air pressure lead to F0 increase.

## 1.2 Affective modulation of F0

F0 has been shown to play a key role in affect signalling within and across species (Ohala, 1984). Its function as an affect expressing signal was enhanced over the course of human evolution into a signal carrier for language. Thus, F0 is influenced by high-level cognitive and low-level primitive functions (Jürgens, 2009). Concerning affective F0 modulation, it is widely assumed that F0 is primarily influenced by the state of arousal rather than by affective valence (Cowie & Cornelius, 2003; Russell, Bachorowski, & Fernandez-Dols, 2003). Autonomic arousal is hypothesized to increase F0 by increasing subglottal air pressure (Briefer, 2012; Harrigan, Rosenthal & Scherer, 2005) and is known to increase, in general, tension of the skeletal muscles (Contrada, 2009). Hence, increased tension of the larynx muscles during arousal might also contribute to elevated F0 levels.

To resume, F0 is an important carrier of linguistic and paralinguistic information (Briefer, 2012; Eckert & Laver, 1994) shaped by emotional and socio-situational factors which interact in a complex manner (Barrett, Pike & Paus, 2004; Eckert & Laver, 1994; Ekman, 1992; Grandjean, Bänziger, & Scherer, 2006). Scherer conceptualizes this complex interaction of various factors as F0 standing at the crossroads between “push and pull effects”(Scherer, 1986, 2003): push effects comprise the physiological component involved in F0 production; e.g. an activation of the autonomous nervous system is assumed to increase muscle tension in the larynx resulting in an elevated F0 level. Pull effects refer to social display rules and cultural norms which may lead to a strategic F0 manipulation of the speaker in order to evoke a certain impression in the perceiver. For example, although a speaker might feel scared during a job interview, he would try to speak in a confident sounding voice. As such, pull effects can also involve a masking of emotional processes by hiding the actual emotion and displaying another. Consequently, research addressing the vocal expression of affect should also take into account the push and pull effects working on F0 (Grandjean, Bänziger, & Scherer, 2006).

### 1.3 Challenges to research on affective F0 modulation

Research on affective modulation of F0 faces several challenges. Partly due to the interdisciplinary interest in this phenomenon, there is no unitary methodological framework. There are ongoing debates on the nomenclature, definitions and concepts of phenomena under investigation and accordingly a great variety of definitions of the term “emotion” co-exist (Cowie & Cornelius, 2003). All approaches towards what might be broadly named research on emotional expression in F0 share common basic challenges (for an overview see Cowie & Cornelius, 2003; Scherer, 2013).

One major challenge lies in the controlled obtainment of authentic affective speech samples. Previous studies have attempted to investigate “naturalistic” speech material, such as from TV, radio shows or emergency call recording with the disadvantage that the underlying affect cannot be accurately controlled (Drolet, Schubotz & Fischer, 2014; Gregory & Webster, 1996; Protopapas & Liebermann, 1997). Partly due to the difficulty of generating speech samples that represent authentic expression of speaker affect, most research on affective F0 modulation has focused on actor portrayals of emotional speech. This approach has proven to be a good solution to some basic questions and is especially fruitful for research on emotion perception from the voice. Posed emotional speech has provided a lot of insight in the field of pull influences on the voice, such as e.g. on conventional vocal patterns.

However, although it can be argued that there is always a certain amount of pull influences such as acting or posing also in natural authentic speech behavior, it can be assumed that acted emotional expression by nature is still more extreme than non-acted, natural expression. Recently, the assumption that vocal measures in general differ between acted speech and authentic speech samples found empirical support (Drolet, Schubotz & Fischer, 2014; Scherer, 2013). Weaker differences between emotions were found in authentic speech samples than in speech portrayed by actors (Laukka et al., 2011). Play-acted speech samples are also characterized by more variable F0 contours (Drolet, Schubotz & Fischer, 2014; Jürgens, Hammerschmidt, & Fischer, 2011). Not surprisingly, it was demonstrated that emotion portrayals seem to be more stereotypical and over-emphasized than authentic ones (Laukka, Audibert, & Aubergé, 2007) and that listeners also rate emotions in speech portrayals as more extreme than in authentic speech samples (Jürgens et al., 2013).

Few studies have applied mood induction techniques to investigate emotional speech in random participants (not actors), such as the memorization of idiosyncratic emotional life events (Barrett, Pike & Paus, 2004), computer games (Bachorowski & Owren, 1995) or the Velten procedure (Scherer, 2013). However, explicit instructions to enter a particular mood or emotion are associated with induction effects that are stronger in negative than positive mood, as shown in a meta-analysis of mood induction procedures by Westermann and colleagues (Westermann et al., 1996). Furthermore, in many cases, the actual emotional responses elicited cannot be controlled as accurately as desired.

### 1.4 Authentic speech in laboratory context: affective F0 modulation induced by psychophysiological methods

This work addresses the challenge of controlled obtainment of authentic affective speech samples. For that purpose, the selection of affect induction methodology is crucial. The aim of this thesis is to test well-proven affect induction methods of psychophysiology for their eligibility in generating comparatively authentic affective speech samples. The term "authentic" is here understood as opposed to acted, posed or explicitly demanded affective speech behavior as in the case of explicitly or implicitly instructed vocal affect expression.

Along with the challenge of finding appropriate affect induction methodologies goes a variety of approaches, definitions and ongoing debates on the phenomenon of "emotion" (Cowie & Cornelius, 2003; Scherer, 2013). It is beyond the scope of this work to attempt a resolution of these issues. Empirical evidence suggests that speech variables are correlated with dimensions rather than with discrete emotion categories (Cowie & Cornelius, 2003). In line with previous research, this thesis is grounded on the hypothesis that speech variables such as F0 are correlated best with activation-valence dimensions and reflect rather affective than emotional experience (Bachorowski & Owren, 2003a; Cowie & Cornelius, 2003; Greasley, Sherrard, & Waterman, 2000; Lang et al., 1993; Russell, 2009). Throughout this work, the modulation of F0 by speaker affect is thus termed "affective F0 modulation".

Accordingly, each of the selected methods presented in this thesis allows an investigation of affective valence and/ or arousal dimensions. In each of the following three studies, 1) no explicit or implicit instruction to express affect was given, 2) one well-established psychobiological paradigm for affect induction was applied, and 3) F0 modulation in a state of negative affect was compared to either a neutral or positive affect state. Additionally, participants were not informed about the real purpose of the respective study until after their participation. After the experiments, participants were asked what they assumed to be the real purpose of the study, and were excluded if they successfully reported it.

The overall aim of all studies was to test established psychophysiological methods for their eligibility in studying affect expression in authentic speech behavior. The CPT and SCWT were considered appropriate for the study of affect, because they are well-established stressors known to evoke negative affective responses (Dahlquist et al., 2007; Deuter et al., 2012; Dreisbach & Fischer, 2012; Fritz & Dreisbach, 2013; Kleinow & Smith, 2006; Lovallo, 1975; Mc Leod, 2005; Peckerman et al., 1994; Renaud & Blondin, 1997; Schwabe, Haddad, & Schächinger, 2008; Yamamoto, Iwase, & Mano, 1992) and the IAPS provides well-proven affective pictures categorized within an activation-valence space (Lang, Bradley, & Cuthbert, 1999).

#### *1.4.1 Study I: Stroop Color-Word Test*

In study I (chapter II), the SCWT was applied as a cognitive stressor. Incongruent stimuli are usually perceived as negative as compared to congruent stimuli, (Dreisbach & Fischer, 2012; Dreisbach & Fischer, 2012; Fritz & Dreisbach, 2013) and have been found to be associated with elevated autonomic arousal (Critchley et al., 2005; Kobayashi et al., 2007; Siegle, Steinhauer, & Thase, 2004). This study found increased F0 levels in incongruent stimuli.

#### *1.4.2 Study II: International Affective Picture System*

In study II, affective slides taken from the IAPS (Lang, Bradley, & Cuthbert, 1999) were presented and F0 levels in affective speech samples of standardized linguistic content were compared to spontaneous speech samples. The standard phrase was a predefined, semantically neutral sentence spoken at the beginning of each slide presentation, whereas the spontaneous speech sample consisted of spontaneously uttered picture descriptions. F0 levels and F0 variation were lowered in speech during negative slides compared to positive slides in both types of speech samples.

#### *1.4.3 Study III: Cold Pressor Test*

Study III (reported in chapter IV) employs the CPT a classic physiological stressor with well-documented sympatico-excitatory and neuro-endocrinological effects. It is known that CPT intervention is generally perceived as negative or aversive, as shown by a large corpus of empirical evidence (Deuter et al., 2012; Lovallo, 1975; Peckerman et al., 1994). Although we did not assess subjective ratings from our participants on pleasantness, pain, or stress levels, it can be safely assumed that our CPT intervention was perceived as aversive by our participants. F0 obtained from speech during reading aloud a short text was found to be moderately increased during CPT stress as compared to a baseline reading condition.

**1.5 Study II: Summary of findings and concluding remarks**

In both studies I and III, negative affect was associated with elevated F0 levels. These results are in line with some previous findings on elevated F0 levels during stress (Benson, 1995; Briefer, 2012; Giddens et al., 2013; Hollien, 1980; Jessen, 2006; Wittels et al., 2002). Negative affect induced by IAPS slides as reported in study III was associated with lowered F0 levels and variation as compared to positive affect. At first glance, this finding may appear to contradict the findings of study I and III. However, the types of affect and arousal levels induced by the various methods cannot be compared across studies and moreover, this was also not the aim of this thesis. Each of the methods applied in the presented studies determined the specificities of linguistic speech content and type of affect to a certain extent. In study III, speech samples consisted of a short read sentence and in study I the speech samples consisted of four color words which were inherently predefined by the Stroop task itself. Study II compared a standardized read phrase with speech from spontaneous picture descriptions, which presumably allowed a view on more naturalistic speech behavior in the spontaneous, free descriptions as compared to read sentences. As Scherer (2013) points out, even subtle differences in task characteristic can lead to major differences in F0. Accordingly, those preliminary results must be interpreted with caution.

It is inherent in our approach that generating experimentally controlled affective speech in samples in laboratories is never a “real real-life situation”, which is a general problem recently addressed and discussed by several researchers (Jürgens et al., 2011; Scherer, 2013). It might thus be questioned to what extent our speech samples can be regarded as representative for authentic “real-life” speech behavior. I argue that the speech samples contained rather authentic expressions of affect as opposed to comparable laboratory controlled studies on affective speech modulation, since participants were students, not actors, and were not explicitly instructed to express affect. Furthermore, the subtlety of affective F0 modulation is in line with empirical findings on comparisons of authentic speech and acted speech (Drolet, Schubotz & Fischer, 2012; Jürgens et al., 2011;

Scherer, 2013) and might also be regarded as argument for the authenticity of affect expression.

The results of this work suggest a number of interesting avenues for future research, e.g. studies employing established psychophysiological methods of affect induction with systematic assessment of evaluative affect components, F0 and other psychophysiological measures. This might give us more detailed knowledge on the complex interplay of push and pull effects in authentic vocal affect expression. On a side note, investigating F0 in well-established paradigms may add an interesting, non-invasive dependent variable to the commonly investigated ones.

The overall aim of this thesis was to meet the challenge of obtaining speech samples that represent authentic expressions of speaker affect in F0 by using established psychophysiological methods of affect induction. The studies reported in the following chapters show that this approach can be fruitful for future research.

# Chapter II- Study I: Stroop conflict increases fundamental frequency of the voice.


### 2.0 Abstract

During incongruent trials of “classical” Stroop Color-Word Tests, automatized reading interferes with controlled ink-naming, resulting in increased speech onset latencies (reaction times; RT). However, other than latency-related findings, it is not known how speech characteristics are influenced by Stroop conflict. The fundamental frequency of the voice (F0) is an important vocal parameter, modulated by various rapidly changing factors, such as arousal, and emotional state. The present study investigated whether Stroop conflict influences F0. Nineteen participants (ten female) performed a verbal Stroop Color-Word Conflict Test. Verbal response onset (RT), duration, and F0 were assessed. Prolongation of RT was found in incongruent trials relative to congruent trials ($p<0.0001$), validating a cognitive conflict situation. F0 significantly increased in incongruent trials ($p<0.01$). Verbal response duration remained unchanged, indicating that the increase in F0 cannot be attributed to faster speaking. This is the first study indicating that cognitive conflict during Stroop Color-Word Testing influences voice frequency characteristics of verbal responses.


***Keywords:*** Stroop, Fundamental frequency F0, cognitive conflict

## 2.1 Introduction

The Stroop Paradigm has become a gold standard for the investigation of cognitive conflict (MacLeod, 1991). In the classic Stroop Color-Word Test, participants are shown written color-words in their mother tongue printed in various ink colors and are asked to name (speak aloud) the ink colors, while ignoring the color-word content (Stroop, 1935). In the congruent condition, word and ink color match, whereas in the incongruent condition they mismatch. The semantic incompatibility within incongruent trials, challenged selective attention, and inhibition of automatic reading process in favor of color-naming are factors contributing to cognitive conflict.

Cognitive conflict during incongruent trials induces effects most apparent on the temporal level of the verbal response, resulting in increased verbal reaction time (RT). Timing is a crucial factor in processes related to both cognition and action, and temporal measures such as RT are established, well-proven windows to assess these processes (Liotti et al. 2000). Kawamoto et al. (2008) describe RT in terms of ”acoustic latency” as a popular index of processing effects occurring prior to response execution. Traditionally, most research on verbal Stroop conflict focuses on the measurement of RTs, since RT difference is an established measure for increased cognitive conflict in color naming with incongruent words (MacLeod, 1991).

Speech can be described as audible movement, because the complex movements of muscles involved in speech production result in the generation of acoustic energy (Kawamoto et al, 2008). Speech production is influenced by low-level primitive and high-level cognitive functions (Jürgens, 2009). It conveys both linguistic and paralinguistic information (Briefer, 2012; Titze, 1994). Speech, such as a verbal response, unfolds over time and hence is measurable on a temporal level. Additionally, there is a qualitative level of acoustic speech properties. In recent decades, acoustic speech analysis methods have advanced allowing a quantification of acoustic speech properties such as voice pitch, intensity and spectral signal properties.

One important aspect of qualitative speech properties is the sound of the voice. As a major acoustic correlate of perceived voice pitch (Titze, 1994), the vocal fundamental frequency (F0) is one of the most prominent acoustic voice parameters. F0 is the acoustic correlate of the vocal fold vibrations (Laver, 1994; Titze, 1994) and an object of interest in various fields of research and applications. F0 analysis has been optimized in recent years and it is a comparably robust measure which is relatively insensitive to low recording quality as compared to other vocal parameters (e.g. slight changes in microphone-to-mouth distance may complicate vocal intensity analyses, but not F0 analyses).

According to Scherer (1986, 2003), bodily expression, including the voice, is driven by the nature of cognitive appraisal. He suggests that F0, as an acoustic parameter of vocal expression, stands at the crossroads of “push and pull effects” (Scherer, 1986, 2003). "Push effects" refer to physiological processes involved in producing speech and shaping the voice signal, such as for example muscle tension and respiration, which are “naturally” influenced by emotions. In contrast, "pull effects" include social norms and display rules, reflecting the fact that vocal expression is often closely monitored and regulated by cognitive processes. This is the case, for example, when speakers intentionally express a feigned emotional state. Thus, social processes, such as self-representation (for example, by trying to create a certain impression on the listener for strategic purposes), and “conventionalization”, which refers to stereotyping in order to enhance clarity in shared meanings of signals, may function as pull factors on F0.

In line with Scherer´s theory, it has been shown that F0 is modulated by push factors, such as age, sex, emotional state and physical health, as well as by pull factors, such as socio-cultural influences, gender or intentional paralinguistic display strategies (Bachorowski & Owren, 1995, 2003a; Bezooijen, 1995; Gregory & Webster, 1996). A study on paralinguistic use of F0 found that Japanese women increase F0 in order to achieve a socially desired stereotype of displaying submissive behavior (Bezooijen, 1995), whereas Japanese men tend to speak in a lowered F0 associated with dominance behavior (Bezooijen, 1995; Ohala, 1984). It is also known that speech directed to infants is generally realized in a higher F0 than adult directed speech (Uther et al., 2007). Infants have an impact on the emotional state and arousal level of caregivers. Thus, it likely that these factors contribute to the higher F0 of speech directed to infants. Cognitive conflict in Stroop Color-Word Conflict Tests is known to increase arousal; indeed, such tests were used as stress tests (MacLeod, 1991; Rothkranz et al. 2004; Tulen, 1989). Thus, it is possible that F0 is also affected by cognitive conflict during a Stroop Color-Word Test.

In contrast to latency effects on the onset of the verbal response, acoustic speech characteristics of the verbal response itself, such as F0, have received little attention in studies applying Stroop paradigms. Kleinow et al (2006) found an increased articulation rate in the context of a Color-Word Stroop task (Kleinow & Smith, 2006). However, not the verbal response itself but standardized sentences preceded by single Stroop trials were analyzed. They concluded that Stroop conflict is associated with increased arousal levels and can influence speech motor control in children and adults. Kello et al. (2000) found that “naming durations” are not modulated within a standard Stroop task, but found prolonged verbal response durations when there was increased pressure for speedy responses. They interpret their findings as evidence for the flexibility of speech production mechanisms in adaptation to task demands. Damian, however, found no effect of conflict or increased conflict level on response duration. He concludes that central processes have no impact on articulatory execution after its initiation even under increased task demand (Damian, 2000). In an exploratory study, Rothkranz and colleagues (2004) applied the Stroop task as a cognitive stressor in order to investigate effects of increasing workload on F0. They report an increase of mean F0 and variable response duration patterns with increasing Stroop task demands. However, their study focuses on F0 in verbal responses to incongruent trials during stepwise decrease of inter-stimulus-interval. No congruent trials were assessed. Thus, until the present study it had not been shown, whether Stroop conflict per se leads to an increase of F0 as compared to the absence of conflict.

To our knowledge, no previous study has compared F0 of verbal responses in Stroop Test congruent and incongruent trials. In the present study, we investigated whether Stroop conflict modulates the voice pitch (F0) in a group of males and females. In line with other researchers (Melara & Algom, 2003; Sabri, 2001) we used a Stroop Color-Word Paradigm comparing congruent to incongruent stimuli. We hypothesized that cognitive conflict would increase the F0 level in conflict trials. We were also interested in possible effects of cognitive conflict on verbal response latency and duration.

## 2.2 Materials and Methods

### *2.2.1 Participants*

Nineteen healthy students (ten female; mean age 23 years; German native speakers) participated in the experiment. Participants received either course credit or monetary compensation. To increase motivation, they were told that an extra monetary reward would be given to the three best performers of their group. Physical health status was assessed by a health questionnaire. Exclusion criteria were color-blindness, confirmed somatic or psychiatric diseases within the last twelve months and medication other than occasional pain killers and oral contraceptives. All participants provided written informed consent and were made aware of their right to discontinue participation at any time.

### *2.2.2 Stimuli and Design*

The instructions to participants and the stimulus presentation were controlled by the computer program E-Prime (PST, Sharpsburg, USA). Participants were seated in front of a computer screen and were instructed to perform a verbal Stroop Color-Word Test (Mc Leod, 2005). The set of 160 stimuli was based on combinations of the four color words "grün", "blau", "rot" and "gelb" (German words for green, blue, red and yellow, respectively) and the four different colors (four color words × four colors × ten). It contained 40 congruent trials (10x 4 colors) and 120 incongruent trials. Each stimulus was preceded by a black fixation cross in the middle of the screen for 500 ms. The duration of the stimulus presentation was 1500 ms. The order of stimulus presentation was pseudo-randomized separately for each participant. Participants were instructed to quickly name the ink color of the word as soon as it was presented (e.g. for "GREEN" written in red, they had to say "red"). To minimize the influence of social pull effects, instructions were presented via computer screen and participants were alone in the room during speech recordings.

### *2.2.3 Speech Analysis*

For the speech recordings, a t-bone diaphragm studio microphone (SC450 large) was placed 10 cm from the participant's mouth. The speech signal from the microphone was amplified with an external USB audio interface (Edirol UA-25) and recorded in 24 bit/96 kHz quality using Adobe Audition.

All measurements of reaction times, F0, offsets and durations were conducted using the speech analysis software Praat (Boersma & Weenink, 2007). All speech samples were segmented semi-automatically into separate sound files. From each sound file, RT, response duration and mean F0 were detected semi-automatically and corrected manually. Detection errors (including octave jumps and detection of periodicity in unvoiced speech segments) were manually corrected. Wrong responses and other speech artifacts such as hesitation sounds were excluded from analysis. All analyses were carried out single blind.

With regard to the analysis of F0, phonetic studies have shown that especially for short speech samples, utterances standardized for linguistic speech content should be used in order to avoid influences on the fundamental frequency (F0) other than paralinguistic, e.g. due to intrinsic pitch differences in vowels (Ladd & Silverman, 1984). Our speech samples consisted of four one-syllable words ("rot", "grün", "blau", gelb") and thus were standardized for linguistic content. Mean F0 and duration for each of the 160 stimuli words was separately measured in order to compare congruent with incongruent trials for each color separately (e.g. ink color green combined with the word "red", "blue" or "yellow" versus ink color green with the word "green").

### *2.2.4 Statistical Analysis*

All statistical analyses were performed with SPSS 20.0. Testing was two-tailed and p-values smaller than 0.05 were considered to indicate statistical significance. Mean ± SD data are reported in text. Figures indicate mean ± SEM. Analysis was restricted to trials in which the ink color was named correctly. Other trials were discarded. Individual median values for verbal RT (ms), duration of the verbal response (ms), and F0 (Hz) were determined per person, ink color, and Stroop conflict level.

The dependent variables verbal RT, duration of the verbal response, and F0 were then statistically analyzed by univariate, three-factorial, mixed design ANOVAs with sex as a two-level between factor (female vs. male), and ink color as a four-level within factor (yellow vs. green vs. blue vs. red), as well as Stroop conflict as a two-level within factor (word/ink conflict vs. no conflict). Greenhouse-Geisser corrected p-values are reported for factors with more than two levels.

Furthermore, verbal RT and F0 values of conflict trials were z-scored per person and subject to be analyzed in a bivariate Pearson product-moment correlation. Individual Pearson correlation coefficients (r) were calculated and then tested in a two-sided, two groups, non-parametric Wilcoxon rank sum procedure for differences between females and males.

## 2.3 Results

### *2.3.1 Fundamental Frequency (F0)*

As expected, there was a strong main effect of sex on F0 ($F1:17=117.6$, $p<0.0001$), with higher F0 in females (223 ± 17 Hz) than males (127 ± 21 Hz). Similarly, and not surprisingly, a main effect of ink color on F0 was found ($F3:51=28.4$, $p<0.0001$), indicating that F0 of the color name “grün” (green) was highest, followed by “blau” (blue), “gelb” (yellow), and “rot” (red). A main effect of Stroop conflict on F0 ($F1:17=8.5$, $p<0.009$) was present (+1,8 Hz in conflict trials). There was no significant two-way interaction between ink color × sex ($F3:51=1.2$), nor Stroop conflict × sex ($F1:17=1.1$), nor Stroop conflict x ink color ($F3:51=0.1$). Similarly, the triple interaction between sex × Stroop conflict × ink color was not significant ($F3:51=0.5$).

### *2.3.2 Verbal reaction time (RT) responses*

No statistically significant main effect of sex ($F1:17=0.1$), nor ink color ($F3:51=1.7$) was detectable on verbal RT. As expected, a strong main effect of Stroop conflict on verbal RT was found ($F1:17=36.7$, $p<0.0001$), with overall longer verbal RT in conflict (737 ± 77 ms) than no-conflict (664 ± 96 ms) trials. There was no significant two-way interaction of ink color × sex ($F3:51=0.4$), nor Stroop conflict × sex ($F1:17=0.02$), nor Stroop conflict × ink color ($F3:51=0.4$). There was also no significant triple interaction of sex × Stroop conflict × ink color ($F3:51=0.2$).

### *2.3.3 Duration of the verbal response*

No statistically significant main effect of sex on verbal response duration ($F1:17=0.2$) was detectable. A main effect of ink color on verbal response duration was found ($F3:51=47.6$, $p<0.0001$), indicating that speech duration of the color-word “blau” (blue: 321 ± 53 ms) was longest, followed by “grün” (green: 301 ± 52 ms), “gelb” (yellow: 252 ± 54 ms), and “rot” (red: 235 ± 66 ms). There was no main effect of Stroop conflict ($F1:17=0.0$), and neither a significant two-way interaction between ink color × sex ($F3:51=1.2$), nor Stroop conflict × sex ($F1:17=1.0$), nor Stroop conflict × ink color ($F3:51=1.2$); nor was there any siginficant triple interaction between sex × Stroop conflict × ink color ($F3:51=1.7$).

### *2.3.4 Correlation analysis of RT and F0 values*

Individual Pearson product moment correlations (each based on all conflict trials of one person) of F0 with verbal RT ranged between r=-0.18 and r=0.43, indicating that on an individual level F0 and verbal RT share only a limited amount of common variance. However, the non-parametric Wilcoxon procedure indicates that correlation coefficients (r) are significantly higher in males than females (Z=2.16, p=0.03).

## 2.4 Discussion

Verbal RT, verbal response duration, and F0 of the spoken color-word were tested in 19 healthy subjects during classical Stroop Color-Word Testing. Incongruent trials induced delayed verbal RT in comparison to congruent trials, a finding indicating a conflict between automatized reading and controlled ink-color naming. F0 of verbal speech responses in incongruent trials was higher as compared to congruent trials. This new result indicates that the voice pitch increases in situations of cognitive conflict. Only weak correlations between individual F0 and RT data were found, suggesting that the effects of cognitive conflict on RT and F0 share only a limited portion of common variance.

Stroop cognitive conflict did not affect verbal response duration. The times used to pronounce the color-words were similar in congruent and incongruent trials. It is known that F0 increases when speech becomes faster (Gandour et al, 1999). Thus, this finding rules out the hypothesis that F0 increased as an artifact of shortened verbal response duration.

F0 is influenced by “push” and “pull” factors and their effects are not easy to disentangle (Scherer, 1986; 2003; 2013). Pull factors are related to social norms and interactions. However, participants in this study were alone in the experimental room and social interactions were reduced to a minimum by displaying all necessary instructions via PC on a screen. There was no direct social contact or influence of the experimenter or any other person. Therefore, it is unlikely that “pull effects”, in terms of social display behavior, played a major role in our results. Furthermore, the conflict effects on F0 reported here appear on an uncontrolled base and occur rapidly within the short RT of between 650 ms to 750 ms, which also suggests that F0 is primarily under the influence of rapidly changing push factors.

Our results raise the question of which rapid push factors might possibly contribute to rapid trial-by-trial F0 changes leading to increased F0 levels in conflict trials. Many researchers have adopted a two-dimensional approach to vocal expression of affective states focusing on valence (pleasant/unpleasant) and arousal (aroused-sleepy) dimensions (Harrigan, Rosenthal, & Scherer, 2005).

Cognitive conflict is associated with increases in autonomic arousal levels, as has been shown by studies on trial-by-trial changes in event-related skin conductance responses (SCR) and pupil diameter (Kobayashi, 2007). Stroop conflict elicits an autonomic pupil dilatation that is larger in incongruent trials than in congruent trials (Siegle et al., 2004). Similarly, SCRs have been found to be larger in incongruent trials than in congruent trials (Kobayashi et al., 2007).

Evidence from vocal affect expression in humans and mammals shows that high arousal emotions are associated with increased F0 (Briefer, 2012; Jürgens, 2009). Autonomic arousal is known to increase, in general, tension of the skeletal muscles (Contrada, 2009) and, hence, it stands to reason that conflict-induced arousal increases tension of the larynx muscles as well. Laryngeal muscles control F0 level by variation of vocal fold tension, mass, length and thickness (Anders et al., 1988; Atkinson, 1978; Laver, 1994) and subglottal pressure (Hoh, 2005). It has been suggested that increase of muscle tension induced by autonomic arousal may influence the breathing pattern and hence F0 by increase of subglottal pressure (Briefer, 2012; Harrigan et al., 2005). Further, autonomic activation might increase cricothyroid activity leading to an increase of vocal fold tension resulting in an increased rate of vocal fold vibration. Thus, autonomic arousal may contribute to F0 increase in response to cognitive conflict.

Conflict trials also differ from non-conflict trials with regard to affective valence. It has been shown that conflict signals, such as in Stroop tasks, may exert an impact on emotional processes: incongruent stimuli tend to be perceived as negative, whereas congruent stimuli are comparatively more positive (Dreisbach & Fischer, 2012a; Dreisbach & Fischer, 2012; Fritz & Dreisbach, 2013). Previous studies have found increased F0 levels during aversive emotional states or negative arousing situations such as increased work load and/or stress (Hammerschmidt & Jürgens, 2007; Hansen & Patil, 2007; Hollien, 1980; Protopapas & Liebermann, 1997; Rothkranz et al., 2004; Russell et al., 2003; Scherer, 2003).

Such effects may be mediated by the anterior cingulate cortex (ACC), which is a crucial structure for the regulation of attention, motor control, the detection of aversive signals and detection of conflict between competing information streams (Paus, 2001; Ruff et al, 2001). It is activated in a conjoint manner by cognitive control, pain and negative affect (Botvinick, 2007; Dreisbach & Fischer, 2012a; Dreisbach & Fischer, 2012b). Moreover, it is also closely related to the state of arousal, negative affect (Paus, 2001) and F0 control (Jürgens, 2002). It has been suggested to play an important role in the volitional initiation of innate processes within vocal behavior such as F0 control during changes in affect and motivation (Jürgens, 2002, 2009). Barret and colleagues (2004) found that the ACC modulates paralinguistic features of speech during affective states, and describe it´s function as an "interface between actions and emotion" (Barrett, Pike & Paus 2004). Thus, it is possible that conflict-induced aversion triggers autonomic arousal via the ACC. However, in our study we did not measure trial-by-trial changes in ACC activity, autonomic arousal nor affective responses. Therefore, we are not able to identify the factors which mediated the observed F0 changes.

It is reasonable to assume that hard-wired, inborn, physiological factors are of crucial importance for the findings reported here, since previous studies suggest that increased arousal and negative emotional states tend to increase F0 even in non-human species (Briefer, 2012). However, it also has to be acknowledged that social factors may play a role, too. Possibly, learning experiences and sociocultural norms of how to express certain emotional states ("pull" factors) may predict the direction of F0 changes. Differences in the magnitude of emotion-related F0-changes exist between cultures, although apparently a positive association between arousal and F0 exists in all cultures studied so far (Scherer, 1986, 2003; Scherer et al. 2001; Scherer & Johnstone, 2000). We have tried to minimize the impact of such "pull" factors and followed an experimental setup which reduced the likelihood that social circumstances influence the voice pitch.

This study adds a further physiological parameter to the list of potential dependent variables used in Stroop tasks. We were able to demonstrate for the first time (to the best of our knowledge) that cognitive conflict, induced by color-word ink incongruency, raises F0. Importantly, this effect is not redundant to the observed RT changes. Correlation analysis within subjects showed only weak associations between RT and F0, albeit the associations were found to be somewhat higher in male than in female subjects. The median Pearson correlation coefficient in the male subject group was r=0.22, and the maximal r=0.44, corresponding to about 5 % and 19 % of commonly shared variance in RT and F0 within individual data sets. This does not point to a substantial overlap, and the respective values for females are even smaller.

### 2.4 Conclusion

This is the first study to show that vocal pitch (F0) is higher in incongruent trials than in congruent trials of Color-Word Conflict Tests. We conclude that F0 changes may be useful in future experiments to identify situations during which conflicting cognitive input is being processed. Acoustic speech analysis is non-invasive and cost-efficient. It may contribute to future investigations of Stroop conflict involving questions on the coordinated relationship between cognition and behavior.

# Chapter III - Study II: Modulation of vocal fundamental frequency in spontaneous and standardized speech samples during affective picture presentation.

### 3.0 Abstract

The influence of emotion on the vocal fundamental frequency (F0) has been well investigated in speech portrayed by actors, but little is known about authentic vocal expression of emotion. This is partly due to the difficulty of generating speech samples that represent authentic expression of speaker affect. This study investigates the influence of speaker affect on F0 in comparatively authentic speech samples obtained in a controlled laboratory setting. Participants (n=40, students, 20 female) were audio-recorded while describing a selection of negative and positive pictures from the International Affective Picture System (IAPS). A cover story was used and no explicit or implicit instruction on vocal affect expression was given. Two types of speech samples, a standard phrase and spontaneous speech, were recorded from each participant for each picture in order to investigate possible differences of affect modulation on F0 due to the type of speech sample. Mean F0 and F0 variance of all speech samples were analysed. Results showed a lowered mean F0 and reduced F0 variation for negative pictures compared to positive pictures, in both standard phrase and spontaneous speech samples. These results, based on relatively authentic speech behaviour, complement earlier findings on the vocal expression of emotions from explicit instructions to vocally portray or express induced emotions. Furthermore, affective F0 modulation is found in both spontaneous and standardized speech, suggesting that short standardized phrases are eligible speech samples in studies of non-portrayed affective vocal expression.

***Keywords:*** IAPS, Fundamental frequency F0, affect

### 3.1 Introduction

Communication of affective states is essential for social life. Disturbances of affect expression can contribute to severe dysregulation of social behaviour. One way to communicate affect is via the acoustic channel, which can be regarded as the phylogenetically most continuous way (Scherer, 1993). Darwin described how in many species, the voice is used as an ”iconic affect signalling device” (Darwin 1998/1872). The fundamental frequency (F0) of the voice has been shown to play a key role in affect signalling within and across species (Ohala, 1984). It can be regarded as the acoustic correlate of the vocal fold vibration rate during phonation and represents a crucial acoustic measure to our auditory perception of voice pitch. F0 variation is related to the perceived speech melody, e.g. low F0 variation is usually perceived as monotone voice (Barrett & Paus, 2002; Titze, 1994).

From a physiological point of view, F0 level depends on the length of the vocal folds, which can be intentionally modulated to a certain extent, on the muscular tension of the vocal folds (cricothyroid and thyroarytaenoid laryngeal muscles), and on the subglottal air pressure (thoracic and abdominal muscles involved in respiration) (Anders, 1988; Atkinson, 1978; Hoh, 2005; Laver, 1994). In the course of human evolution, the voice's function as a signal carrier for language was superimposed on its function as an affect signalling device. Thus, F0 is an important carrier of linguistic and paralinguistic information (Briefer, 2012; Eckert & Laver, 1994) shaped by emotional and socio-situational factors which interact in a complex manner (Barrett, Pike, & Paus, 2004; Eckert & Laver, 1994; Ekman, 1992; Grandjean, Bänziger, & Scherer, 2006).

On the one hand, this becomes evident in the fact that, when hearing someone speak, listeners immediately form impressions, e.g. about the age, sex and emotional state of the speaker (Eckert & Laver, 1994). On the other hand, speakers not only convey information about their actual emotional state but shape the vocal signal intentionally in order to achieve a certain impression in the listener. Some researchers argue that most paralinguistic vocalizations, e.g. laughter, are used to influence the affective state of the listener and consequently to affect his or her behavior rather than conveying representational cues about the actual emotional state of the speaker (Bachorowski & Owren, 2003a). Scherer introduced the concept of "push" and "pull" effects (Scherer, 1986):

"push effects" refer to the physiological component involved in producing speech and shaping the voice signal; e.g. an activation of the autonomous nervous system might lead to a higher muscle tension in the larynx resulting in an elevated F0 level. "Pull effects" include social norms and display rules which may lead to a strategic vocal signalling of certain paralinguistic information with the aim of evoking a certain impression in the perceiver. For example, although a speaker might feel bored by a listener, he might display interest towards that person. Furthermore, pull effects can also involve a masking of emotional processes by hiding the actual emotion and displaying another. Consequently, research addressing the vocal expression of emotion should also take into account the push and pull effects working on the voice signal and, ideally, aim to disentangle those kinds of influences (Grandjean, Bänziger, & Scherer, 2006).

But up to now, few studies addressing paralinguistic vocal communication of emotion have considered the interaction of push and pull effects. Instead, two major lines of research have evolved over the last few decades (Scherer, 2003). One focuses on the expression of emotion, whereas the other is more interested in the perception. However, research on perception and in particular on expression is generally limited by several methodological problems.

First, there is a trade-off problem with spontaneous in contrast to non-spontaneous speech material. Ideally, the verbal content of speech should be phonetically standardized in order to exclude influences, e.g. intrinsic pitch differences in vowels (Titze, 1994), other than those of paralinguistic relevance. As a consequence, spontaneous speech conveying natural emotional expression as we find it in most day-to-day interactions would be excluded leaving a corpus of less authentic speech material. Therefore, a great variety of speech material from spontaneous, semi-spontaneous and standardized (mostly read) speech has been used to understand  more about the nature of emotional expression in the voice, but due to this variation most findings can hardly be compared to one another. Recent clinical findings on effects of different tasks on speaking fundamental frequency suggests that continuous speech should be used in order to determine authentic speaking F0 values, rather than using single vowels and counting tasks (Barsties, 2013).

Secondly, there is no commonly established method of emotion induction, a lack which leads to a variety of approaches ranging from the application of computer games

(Bachorowski & Owren, 1995) to the recalling of idiosyncratic emotional life events (Barrett, Pike & Paus, 2004), but only rarely to the application of established psychological mood induction techniques, such as the Velten procedure (Scherer, 2013). The problem here is that in many cases, the actual emotional responses elicited cannot be controlled as accurately as desired. This is especially true for studies using "naturalistic" speech material, such as e.g. TV or show conversations or pilot voice box recordings (Drolet, Schubotz & Fischer, 2014; Gregory & Webster, 1996; Protopapas & Liebermann, 1997).

As a consequence of these research implications, most research controlling for verbal speech content uses actor portrayals of emotional speech and most of our knowledge on vocal affect expression is based on this research convention (Laukka, 2008; Scherer, 2013) which has proven to be a good solution to some basic problems in vocal emotion research; yet its application is somewhat limited to the study of emotion perception rather than authentic real-life expression. Although emotional speech portrayals provide information on pull effects such as conventional vocal patterns for expressing emotional valence, the authenticity of acted emotional speech is, by nature, fairly limited. Recent studies comparing authentic speech to speech produced by actors show that authentic speech samples differ in vocal measures from speech portrayed by actors (Drolet, Schubotz & Fischer, 2014; Scherer, 2013).

Recently, Laukka and colleagues found weaker differences between emotions for authentic speech samples than portrayed speech (Laukka et al., 2011). Studies comparing speech of actors with authentic spontaneous speech found that play-acted speech revealed more variable F0 contours (Drolet et al., 2014; Jürgens, Hammerschmidt, & Fischer, 2011). Other research has demonstrated that play-acted vocal emotion portrayals seem to be over-emphasized and more stereotypical than authentic ones (Laukka, Audibert, & Aubergé, 2007) and that listeners rated emotions in authentic speech samples as less extreme than portrayed speech samples (Jürgens et al., 2013).

Some researchers have tried to evoke more laboratory controlled natural speech from students imagining certain types of speech partners instead of using actors but, still being based on simulated interactions, such approaches have similar limitations and have been found to be in some cases even less useful than actor speech portrayals (Knoll, Scharrer, & Costall, 2009). Moreover, in a meta-analysis of mood induction procedures, Westermann and colleagues found that explicit instructions to enter a particular mood or emotion are associated with induction effects that are stronger in negative than positive moods (Westermann et al., 1996). The social context of speech recording has rarely been reported in most studies so far, although it can have a crucial impact on the expression of emotion (Buck et al., 1992; Laukka, 2008; Scherer, Banse, & Wallbott, 2001). Thus, it is possible that some contradictory findings on vocal expression of emotion might be due to different social settings leading to variations of push and pull influences, e.g. sex of the experimenter, the experimenter being present or not present in the room.

Another challenge in the research of vocal emotion expression is the use of diverse emotion concepts, or in other words, there is no unitary, generally accepted concept of emotion in vocal expression research (Cowie & Cornelius, 2003), instead, a variety of approaches to, definitions of and ongoing debates on the phenomenon of “emotion” exist. Cowie and colleagues argue that the only reliable relationships between voice and emotion are those that reflect rather basic emotion-related changes in physiology, despite the fact that either evolution or social factors have ensured that the voice communicates more about an organism’s general orientation to the world than physiological changes require it to (Cowie & Cornelius, 2003). They suggest disregarding traditional emotion categories by using a simpler approach towards this topic, acknowledging that speech variables could be correlated with activation-valence dimensions rather than with discrete categories. Similarly, Bachorowski points out that there is remarkable support of the notion that using dimensions of non-specific arousal and affective valence is the best way to describe features of “emotional speech” and further states that most vocal productions reflect rather affective than emotional experience (Bachorowski & Owren, 2003b). Lang and colleagues argue that the consistency of physiological response and the dimensional relationships between evaluative judgments on pleasure and arousal means that emotion is fundamentally organized by these motivational parameters (Lang et al., 1993). Other researchers indirectly support this approach by arguing that it is rarely the case that only one emotion is encoded in spontaneous speech (Greasley, Sherrard, & Waterman, 2000; Jürgens et al., 2011).

Recently, interest in investigating authentic speech samples has grown (Drolet, Schubotz & Fischer, 2012, 2014; Greasley, Sherrard, & Waterman, 2000; Jürgens et al., 2013; Jürgens et al., 2011). In consideration of the aforementioned challenges to vocal emotion research, this study investigates affective modulation of F0 in non-instructed and non-acted affective speech, which in the following will be referred to as “authentic”. Instead of using real-life media samples or explicit emotion induction techniques, we chose to investigate speech elicited by students, not actors. We used a cover story and avoided any explicit or implicit instruction towards participants to vocally express affective states. As method of emotion induction, we presented affective pictures selected from the International Affective Picture System (IAPS), which provides a well-established, validated corpus of affective pictures allowing controlled induction of affect (Lang, Bradley, & Cuthbert, 1999). A selection of positive and negative pictures was used for the induction of negative and positive affect. We collected two types of speech samples per picture from each participant (one standard phrase, one spontaneous description) to investigate if and to what extent affective modulation of F0 would depend on the type of speech sample. We were interested in modulatory effects of positive and negative affect on F0 mean and F0 variation.

## 3.2 Methods

### *3.2.1 Participants*

The sample started with 44 healthy students (20 female). All participants (mean age: 24 years) were native speakers of German without voice or speech impairment (e.g. cold), color blindness or current or recent (less than 3 years) psychological treatment and with a normal body mass index (BMI<25). Before the start of the experiment, every subject was interviewed according to the BASE health questionnaire and underwent a short visual test for color-blindness (Ishihara, 1917). The participants were told that they would help with the categorization of pictures for a scientific databank, and received an allowance after finishing the experiment. In order to keep their focus on the linguistic content of the description, they were instructed to describe the presented pictures as precisely and in as much detail as possible. All participants underwent a small manipulation check questionnaire after the experiment and two were excluded, because they assumed that the purpose of the experiment was to analyse the voice and not the linguistic content of the description. Two participants were excluded because of speech errors (e.g. inconsistent rephrasing of standardized sentence) in more than one third of the speech samples, leaving a total of 40 participants for final analyses (20 female).

### *3.2.2 Picture Presentation*

Participants sat on a chair in front of a computer screen. The experimental instructions and the stimuli presentation were controlled by the computer program E-Prime (PST, Sharpsburg, USA). In order to familiarize the participants with the situation and the recording procedure, the experiment began with a practice trial where participants had to describe two neutral pictures. Each picture was presented for 50 seconds. During the initial 20 seconds, participants had to look at the picture intensively without speaking. Afterwards, a short acoustic signal (350 Hz sine wave) indicated the start of the description task and the initial phrase was displayed above the picture. Participants had to start each picture description with the initial phrase "Auf diesem Bild ist Folgendes zu erkennen" ("In this picture the following can be seen."). This initial phrase was used to generate standardized speech samples. The participants had a time frame of 30 seconds for the picture description, then a black screen appeared for 10 seconds and the next picture was presented. Speech from the picture descriptions represented spontaneous speech samples. After the two practice trials the participants had the chance to ask questions if necessary. Then the experimental picture description task started.

The picture sample consisted of 16 positive[1] and 16 negative[2] pictures of various social scenes (e.g. sports activities, violence) selected from the IAPS based on valence and arousal ratings obtained from previous validation studies (Lang et al., 1999). The two content groups, positive and negative, were distinct and representative of affect category (negative: 2,9; positive 7,03), mean arousal ratings for both categories were similar and did not differ significantly (negative: 5,7; positive: 5,3). The order of picture presentation was pseudo-randomized.

### *3.2.3 Speech Recording and Analyses*

For the voice recordings, a t-bone SC450 large diaphragm studio microphone was placed 10 cm from the subject's mouth. The speech signal was amplified with an external USB audio interface (Edirol UA-25) and recorded in 24 bit/96 kHz quality using Adobe

[1] IAPS picture numbers used in this study were 2304, 2310, 2311, 2320, 2485, 2500, 2510, 5410, 5621, 8030, 8034, 8160, 8180, 8185, 8370, 8490

[2] IAPS picture numbers used in this study were 2752, 2753, 2800, 2900, 3220, 3230, 6243, 6250, 6350, 6360, 6550, 6530, 9000, 9001, 9530, 9520

Audition. From each of the participants, 32 speech recordings of each the initial standard phrase and the spontaneous picture description were analysed. Durations of the speech samples were limited to 10 seconds for the initial standard phrase and to 30 seconds for the spontaneous picture descriptions.

Concerning the duration of short speech utterances, it is agreed that 10 s is long enough to allow precise interpersonal perception from paralinguistic speech behavior (Ambady & Rosenthal, 1992). With regard to fundamental frequency analysis, phonetic studies have shown that especially for short speech samples, utterances standardized for linguistic speech content should be used in order to avoid influences on F0 other than paralinguistic ,e.g. due to intrinsic pitch differences in vowels (Ladd & Silverman, 1984; Ohala & Eukel, 1987).

All measurements were conducted using the speech analysis software Praat (Boersma & Weenink, 2007). The mean and standard deviation of the fundamental frequency were extracted using an autocorrelation algorithm. Detection errors (including octave jumps and detection of periodicity in unvoiced speech segments) were manually corrected. The coefficient of F0 variance (F0 varco) was calculated to index F0 variation independent from absolute F0 level (F0 varco = F0 standard deviation/ F0 mean) (Quinto, Thompson, & Keating, 2013). All measurements were carried out blind.

### *3.2.4 Statistical Analysis*

All statistical analyses were performed with SPSS 20.0. Testing was two-tailed and p-values smaller than 0.05 were considered to indicate statistical significance. Mean ± SD data are reported in text. Individual mean values for mean F0 (Hz) and F0 variation (varco) were determined per person, type of speech (standard phrase, spontaneous), and affect (negative, positive).

The dependent variables mean F0 and F0 variation were then statistically analyzed by univariate, three-factorial, mixed design ANOVAs with sex as a two-level between factor (female vs. male), and speech type as two-level within factor (standard phrase vs. spontaneous speech), as well as valence as a two-level within factor (negative vs. positive). Greenhouse-Geisser corrected p-values are reported for factors with more than two levels.

## 3.3 Results

### *3.3.1 Mean F0*

As expected, there was a strong main effect of sex on mean F0 ($F_{1:38}=253.21$, $p<0.00$), with higher mean F0 levels in females (224 ± 22 Hz) than males (102 ± 17 Hz). A main effect of speech type on F0 was found ($F_{1:38}=28.15$, $p<0.00$), indicating that mean F0 in speech from the standardized phrase (179 ± 57 Hz) was 11 Hz higher than in spontaneous speech (168 ± 51 Hz), and also a strong interaction between sex × speech type was found ($F_{1:38}=7.50$, $p<0.01$). Separate ANOVA for male and female speech show a significantly higher F0 in both male speech (+ 4,5 Hz $F_{1:19}=6.59$, $p<0.01$) and female speech (+ 14 Hz; $F_{1:19}=21.57$, $p<0.00$) in the standard phrase as compared to spontaneous speech.

A main effect of valence was present showing a significant difference of 1,3 Hz between speech during presentation of negative (171,9± 53,9 Hz) and positive (173,3 ± 54,7 Hz) pictures ($F_{1:38}=11.65$, $p<0.01$). There was no significant two-way interaction between valence × sex ($F_{1:38}=2.17$), speech × valence ($F_{1:38}=0.16$), and no significant three-way interaction between speech × valence × sex ($F_{1:38}=3.65$).

### *3.3.2 F0 variation*

No main effect of sex was found for F0 variation (Mean F0 varco= 12.94; $F_{1:38}=0.44$). A main effect of valence was present showing a significant difference of 0.80 varco between speech during presentation of negative (12,54± 0.4) and positive (13,34 ±0.5) pictures ($F_{1:38}=47.95$, $p<0.00$), but a strong interaction between sex × valence was found ($F_{1:38}=7.31$, $p<0.01$). Separate ANOVA for male and female speech showed a significantly higher F0 variation in both male (+ 0.48; $F_{1:19}=6.59$, $p<0.01$) and female speech (+ 1.12; $F_{1:19}=21.57$, $p<0.00$) for positive pictures than for negative pictures. No significant interaction of speech × valence was found ($F_{1:38}=3.41$). No main effect for speech type ($F_{1:38}=1.79$) and no significant three-way interaction between speech × valence × sex ($F_{1:38}=0.07$) was found.

## 3.4 Discussion

Our results show typical sex differences in mean F0 reflecting the well-investigated fact that men typically speak in a lower F0 than women (Ohala, 1984). No overall sex differences were found for F0 variance. More importantly, our results show subtle effects of affective modulation in both mean F0 and F0 variance.

We found a small decrease of mean F0 during negative affect compared to positive affect. The sex x valence interaction shows that F0 variation decreases during negative affect in both, men and women. In women, this affective modulation of F0 variance appears to be more pronounced than in men. It is known that gender can play an important role in paralinguistic modulation of F0 (Bachorowski & Owren, 2003a; Bezooijen, 1995; Cartei, Cowles, & Reby, 2012; Eckert & Laver, 1994). However, it is not clear whether this is a valid gender effect or an artefact of the principle that F0 variance is dependent on absolute F0 level. For example, to achieve an increase of F0 by one octave, F0 needs to be doubled. A male speaker with a mean F0 of 100 Hz would have to increase F0 by 100 Hz, whereas a female speaker with a mean F0 of 200 Hz would have to increase her voice pitch by 200 Hz. Accordingly, as it comes to perceived speech melody, women need to apply also a higher F0 variance than men to achieve equal auditory effects on the level of pitch perception. The sex x valence interaction found in F0 variance might be due to this principle. Further studies are needed to clarify influences of gender on affective modulation of F0 variance.

We found that mean F0 in the standard phrase and spontaneous speech differed significantly. This is in line with previous studies, which have found that F0 varies significantly with the type of speech task (Baker et al., 2008; Barsties, 2013; Fitch, 1990; Zraick, Skaggs, & Montague, 2000). Mean F0 during the standard phrase was higher than in the spontaneous picture description. Several factors might contribute to the direction of this difference. We assume that it is largely based on an uneven distribution of voiced sounds in the standard phrase. The standard phrase contains a comparatively larger percentage of high intrinsic pitch sounds than a supposedly more even distribution of sounds in spontaneous speech. This is somewhat inherent in the nature of speech tasks, since the standard phrase contains always the same semantic content, whereas the spontaneous picture description varies semantically within-subjects and between pictures allowing a more even distribution of sounds and their respective intrinsic pitches.

Apart from this, the standard phrase was always produced initially. We cannot exclude that the constant initial production of the standard phrase had an influence on F0. For example, it is possible that respiratory and laryngeal adjustments made in anticipation of the initiation or termination of speaking directly influenced intensity and, indirectly, mean F0 (Zraick, Birdwell, & Smith-Olinde, 2005). The two speech sample types also differed in duration, but based on empirical findings that found no significant difference in mean F0 between spontaneous speech samples of 10 and 30 seconds duration, we assume that the difference in duration play a rather minor role for our findings (Zraick et al., 2005).

More importantly, affective modulation of F0 was found in both types, spontaneous and standardized speech samples. This result suggests that short standardized phrases are eligible speech samples in studies on authentic affective vocal expression in F0. However, this preliminary conclusion is limited to modulation of mean F0 and F0 variation and given the above discussed point needs further empirical verification.

It is inherent in our approach that generating controlled “authentic” affective speech samples in a laboratory setting is never a “real real-life situation”, which is a general problem recently addressed and discussed by several researchers (Jürgens et al., 2011; Scherer, 2013). It might thus be questioned whether the obtained speech samples can be regarded as authentic speech or not. We argue that the speech samples contained rather authentic expression of affect, since participants were normal students and not explicitly instructed to express affect. Furthermore, the subtlety of F0 modulation by positive and negative affect is in line with empirical findings on comparisons of authentic speech and acted speech (Drolet, Schubotz & Fischer, 2012; Jürgens et al., 2011; Scherer, 2013) and can also be regarded as evidence for authenticity of affect expression in that particular laboratory setting. We cannot draw conclusions from our data on how relevant the reported subtle F0 differences are on the preceptional level, but surely this would be a very interesting question to address in a future study.

The influence of diverse pull effects on the voice, such as social norms and display rules of expressing and maybe also masking of affect in the vocal expression, might contribute to the subtlety of affect modulation in F0. It is not clear to what extent affective modulation of F0 was influenced by the laboratory environment, for example if tendencies to hide affect expression in the voice existed. As mentioned earlier, push and pull influences on F0 are always present and pull influences might mask or overshadow affect induced push influences on F0 in our study, as in any social situation. As Scherer points out, even subtle changes in task characteristic can lead to major changes in the vocal expression (Scherer, 2013). Another explanation for the subtlety might lie in the possibility that picture induced affect was possibly not strong enough to lead to a more pronounced F0 modulation. However, these are just speculations and further studies are needed to deepen our understanding on why authentic speech samples appear to contain more subtle affect expressions than acted speech.

Due to some trade-offs within our study design, there are some limitations when it comes to discussing underlying psychophysiological mechanisms leading to affective F0 modulation. By definition, the valence dimension cannot be viewed independently from the arousal dimension (Lang et al., 1993, 1999; Russell, 2009). Our selection of IAPS pictures was chosen to contain valid representations of valence groups and it was well-balanced for arousal ratings reported by Lang and colleagues (Lang et al., 1999), since we aimed to compare positive and negative affect expression in F0. Nevertheless, it is possible that not affective valence but differences in actual autonomic arousal levels between negative and positive pictures influenced F0 level and F0 variation.

Russell argues that changes in F0-related measures in affective speech samples of both acted portrayals and natural speech rather reflect changes in arousal level of a speaker than affective state (Russell 2003). He states that vocal cues unique to a valence or pleasure dimension are elusive. Similarly, Scherer describes F0 mainly as an indicator of arousal and not affective valence, but he also points out that conclusion is drawn mostly from research based on actor portrayals including the aforementioned limitations (Scherer, 2013). Briefer found that vocal correlates of arousal have been investigated considerably more than correlates of valence (Briefer, 2012).

We neither measured autonomic responses to affective pictures, nor assessed valence or arousal ratings from our participants. This was due to a trade-off between not revealing the study purpose in order to obtain authentic affect expression and explicitly assessing valence/arousal ratings and measures which would have possibly raised awareness in participants of the study purpose. Notably, our study shows that positive and negative affect modulates F0 in authentic speech samples, which predominantly might be interpreted as an effect of affective valence, indeed, and/or as a result biased by underlying arousal differences between positive and negative picture groups. To clarify this question, the detailed relationship of individual arousal levels, valence ratings and affective F0 modulation is to be addressed in future studies.

### 3.4 Conclusion

As Scherer points out, one of the great difficulties in vocal emotion research is to obtain access to speech samples that represent authentic expressions of a speaker´s emotion, and even subtle changes in task characteristic can lead to major changes in the vocal expression (Scherer, 2013). Our study investigated the influence of positive and negative affect on mean F0 and F0 variance in speech from students in a controlled laboratory setting, in order to obtain more authentic speech samples than previous research using acted or instructed emotional speech tokens. Subtle effects of positive and negative affect on F0 measures were found. Furthermore, affective modulation of F0 was found in both spontaneous and standardized speech samples, which suggests that short standardized phrases are eligible speech samples in studies on authentic affective vocal expression in F0.

# Chapter IV- Study III: Cold Pressor Stress increases vocal fundamental frequency.

**4.0 Abstract**

Speech conveys a lot more than linguistic information. From the tone of voice, listeners immediately draw conclusions on diverse speaker characteristics, e.g. stress level. It is assumed that stress increases muscle tension in the larynx leading to an increase of vocal fundamental frequency (F0). However, there are no unitary empirical findings. This might be partially due to the application of non-validated tests and/ or stress provocations. In this study, we used the Cold Pressor Test (CPT) to investigate the influence of acute stress on mean F0. The CPT is an established instrument for the physiological induction of stress with well-documented sympathico-excitatory and neuroendocrine effects. Forty female volunteers were randomly assigned to either a control or a stress group. Both groups were asked to read aloud a short text in two conditions: once in neutral a condition to assess baseline F0 level (baseline measurement), and once while immersing the left hand for two minutes into ice-cold water (stress group), or luke-warm water (control group), respectively. Speech, heart rate (HR) and systolic arterial blood pressure (SAP) data were recorded during both reading conditions. Mean F0 levels of speech were analyzed. Cardiovascular data indicated a successful CPT intervention. The stress group showed a significant moderate increase of mean F0 during CPT intervention relative to baseline F0 level, while no differences were found in the control group. Our results support earlier findings of increased mean F0 during acute stress. Our study is the first to investigate the influence of CPT stress on F0.

***Keywords:*** Cold Pressor Test, Fundamental frequency F0, stress

## 4.1 Introduction

The voice is an important carrier of paralinguistic information (Eckert & Laver, 1994). Listeners immediately draw conclusions from the sound of a speaker´s voice on diverse biologically determined speaker characteristics, e.g. sex and age, but also on short-term psychophysiological characteristics, such as stress level (Giddens et al., 2013).

The influence of stress on the voice is a phylogenetically old heritage. Intuitively we might think of the elicitation of a high-pitched voice as a warning signal to same-species members in face of predators, an adaptive behavior not only in regard to the lives of our human ancestors, but also commonly observed in mammals, generally (Briefer, 2012; Contrada, 2009). Stress can be regarded as having originally been an adaptive bodily response to a (potential) life-threat redirecting biological and physiological mechanisms towards survival (Contrada, 2009). Nowadays, human speech is still influenced by this old cascade of physiological changes during stress. A great corpus of research on the influence of stress on vocal parameters reflects an interdisciplinary interest in this phenomenon (Juslin & Laukka, 2003; Scherer, 2003).

The vocal fundamental frequency (F0) is one of the most prominent acoustic voice parameters and correlates with the vocal fold vibrations (Laver, 1994; Titze, 1994). Laryngeal muscles control F0 level by variation of vocal fold tension, mass, length and thickness and subglottal pressure (Anderson & Cooper, 1986; Atkinson, 1978; Eckert & Laver, 1994). It stands to reason that increased autonomic arousal as part of the physiological stress response may play an important role in modulating F0 during stress. Firstly, autonomic arousal is known to increase, in general, tension of the skeletal muscles (Contrada, 2009) and, hence, it seems a reasonable assumption that stress-induced arousal increases tension of the larynx muscles as well. Secondly, it is known that increased tension of the cricothyroid, which is in control of vocal fold tension, leads to an increase in F0 (Titze, 1994).

Evidence from vocal affect expression in humans and mammals shows that stress and/or arousing situations are often associated with increased F0 levels (Giddens et al., 2013; Hammerschmidt & Jürgens, 2007a; Hansen & Patil, 2007; Protopapas & Liebermann, 1997; Scherer, 2003). However, empirical findings on the influence of stress on F0 are divergent (Hansen & Patil, 2007; Juslin & Laukka, 2003; Scherer, 2003). This might be partially due to the application of non-validated tests and/ or stress provocations. Also, most research has investigated simulated emotion portrayals from actors (Scherer, 2013) and few studies have attempted to use speech from normal speakers (Bachorowski & Owren, 1995). However, findings from both approaches seem to support the idea of this link between F0 level and the general arousal of a speaker (Russell, Bachorowski, Fernandez-Dols, 2003; Scherer, 2013).

To further investigate this assumption, we decided to apply a well-established stress test in Psychophysiology, the Cold Pressor Test (CPT) (Hines, 1940; Schwabe, Haddad, & Schächinger, 2008). The CPT has well-documented sympathico-excitatory and neuroendocrine effects (Allman et al., 2001; Lovallo, 1975; Schwabe, Haddad & Schächinger, 2008; Streff et al., 2010). Throughout the last few decades, it has found wide-spread applications in various fields, such as research on stress, pain, psychosomatic illness and basal autonomic function (Streff et al., 2010). The CPT has been shown to lead to a general sympathetic activation involving two somatosensory systems, thermo- and nociception (Lovallo, 1975). It has been suggested that heart rate regulation during CPT is strongly influenced by the sensation of pain (Mourot, Bouhaddi, & Regnard, 2009). Generally heart rate peaks after the first 60 seconds of the CPT (Moriyama & Hirotoshi, 2007; Victor et al., 1987). Also, an increase of muscle sympathetic nerve activity during the first 30-90 seconds of CPT has been found (Victor et al., 1987; Yamamoto, Iwase, & Mano, 1992) and the strongest blood pressure increases occur during the second minute of CPT stimulation (Deuter et al., 2012). Victor and colleagues (1987) found a positive correlation between the increases in muscle sympathetic nerve activity and the increases in arterial pressure during the CPT. They interpret their findings in terms of direct evidence for increased sympathetic drive to skeletal muscle during the CPT.

In this study, we applied the CPT paradigm as a method of stress induction to investigate the influence of acute stress on F0. We hypothesized that in speech produced during CPT-induced stress, F0 would increase due to sympathetic activation relative to F0 level in a non-stress baseline speech sample of the same speaker. Also, we wanted to obtain “natural” speech from normal participants, not actors, without explicit induction of affective states or instruction to vocally express them.

## 4.2 Methods

### *4.2.1 Participants*

Forty healthy female students (mean age 22 years, German native speakers) participated to receive either course credit or monetary reward. Physical health status was assessed by a health questionnaire. Exclusion criteria were confirmed somatic or psychiatric diseases within the last 12 months, other than banal infections or minor injuries and medication other than occasional pain killers and oral contraceptives. All participants were informed that they would have to immerse their left hands up to the wrist into either cold or luke-warm water, that their speech would be recorded to document the experiment and that they would be constantly monitored via video- and audiotransimission (small webcam). Participants from the stress group were additionally informed that they also would be videotaped in order to analyze their facial expressions during the cold water test. After the experiment, participants were asked what they assumed to be the purpose of the study. No participant reported relevant assumptions to justify exclusion from analysis. All participants provided written informed consent and were made aware of their right to discontinue participation at any time.

### *4.2.2 Physiological Recordings*

Participants were seated in a comfortable chair at a viewing distance of approximately 60 cm in front of an LCD computer monitor. Electrodes for ECG measurement (ECG Tyco healthcare H34SG Ag/AgCl electrodes of 45 mm diameter) were placed according to a standard lead II configuration. The ECG signal was high-pass filtered (Biopac ECG100; HPF:0.5 Hz) and stored to disk (1 kHz), and heart rate was derived from the ECG measure. Beat detection and artifact control was performed offline with WinCPRS (Absolute Aliens, Oy,Turku, Finland). Continuous blood pressure was recorded using the Finapres System (Ohmeda,Englewood, CO; USA). Beat-to-beat systolic blood pressure was determined offline by WinCPRS software. Mean values over minute two of hand immersion in water and mean values over minute two of baseline speech recording for heart rate and systolic blood pressure were calculated in order to determine whether the CPT intervention was successful.

Participants were asked to sit in a comfortable upright position, not to move and to follow the instructions displayed on the computer screen. Before the experiment started, a five-minute resting period allowed participants to adapt to the laboratory setting. Then, participants were instructed to read aloud a short, neutral text ("Der Nordwind und die Sonne", reading time between 30-50 s) in order to become familiarized with the recording settings, to practice reading and to allow for adjustment of microphone settings if necessary. After this reading practice, the intervention started and a bowl filled with either luke-warm water or ice-cold water (below 3 degree Celsius) filled with crushed ice was positioned on the left side next to the participant.

In the stress group, the experimenter pretended to activate a video camera that was placed in front of the participant. Then, the experimenter counted down from three to zero and participants had to immerse the left hand up to the wrist in the ice-cold water. The experimenter remained present in the experiment room without speaking to the participant. It has been shown that participants have a longer tolerance of Cold Pressor pain, if the experimenter is of opposite sex (Kallai, Barke, & Voss, 2004), so we chose opposite-sex experimenters to decrease the probability of experiment abortion by the participant. Furthermore, the presence of an experimenter during the CPT and videotaping of the participants are social-evaluative elements increasing HPA axis activity compared to standard CPT (Schwabe, Haddad & Schächinger, 2008) and fostering a stronger stress reaction than the standard paradigm. One minute after hand immersion in the water, the short neutral text "Der Nordwind und die Sonne" was presented on the screen again and participants had to read it aloud. This point in time was selected for the recording of speech during the intervention because it has been shown that cardiovascular effects of Cold Pressor stress reach their maximum after 60 seconds (Peckerman et al., 1994; Victor et al., 1987).

Each participant had been told that they had to leave their hand in the water for 3 minutes, since 3 minutes is the standard time of CPT intervention and participants should expect this duration accordingly prior to the intervention. But after the reading task was finished by the end of minute two, participants were allowed to remove their hands from the water. Since we were interested in in the speech during minute two, it was not necessary to have participants endure the cold water intervention for one more minute after the reading. Furthermore, we had decided to record the baseline speech samples not prior to but after the intervention, because it has been shown that anticipation of a stressor or pain influences cardiovascular and psychological variables (Conrod, 2006; Pieper et al., 2007; Tomaka, 1997) and we wanted to avoid possible influences of anticipatory stress on the speech signal. Since autonomic cardiovascular measures are rapidly responding within seconds and return to baseline within minutes (Linden et al., 1997), baseline speech samples were recorded 10 minutes after the end of the water intervention and were used to determine baseline F0 levels.

Participants were allowed to remove their hands from the water after having finished reading aloud the text, which was at the end of minute two. They were handed a towel to dry their hands and were given a break of 10 minutes to recover while remaining seated. Then the baseline speech recording was performed while participants were alone in the room.

In the control group, the experimenter counted down to three and then the participant had to immerse the left hand up to the wrist into the luke-warm water, then the experimenter left the room. Similar to the stress group, participants had to start reading aloud the text at the onset of minute two of hand immersion in the water. Throughout the whole experimental session, heart rate and systolic blood pressure were continuously measured. Participants were monitored via video and audio transmission.

### *4.2.3 Recording and analysis of speech samples*

For the speech recordings, a t-bone SC450 large diaphragm studio microphone was placed 10 cm from the participant’s mouth. The speech signal from the SC450 studio microphone was amplified with an external USB audio interface (Edirol UA-25) and recorded in 24 bit/96 kHz quality using Adobe Audition. Overall, 80 speech samples (40 baseline samples, 40 intervention samples) from all 40 participants each of less than one minute duration were obtained. From each speech sample, the first sentence was used for mean F0 analysis.

With regard to the analysis of the fundamental frequency (F0), phonetic studies have shown that especially for short speech samples, utterances standardized for linguistic speech content should be used in order to avoid influences on the fundamental frequency (F0) other than paralinguistic, e.g. due to intrinsic pitch differences in vowels (Ladd & Silverman, 1984).

The first sentence of the text was chosen for analysis: “Einst stritten sich Nordwind und Sonne, wer von ihnen beiden der Stärkere wäre, als ein Wanderer, der in einen warmen Mantel gehüllt war, des Weges daherkam”. All of our speech samples consisted of the same short text and were thus standardized for linguistic content. All measurements were conducted using the speech analysis software Praat (Boersma & Weenink, 2007). The mean of the fundamental frequency was extracted using an autocorrelation algorithm. Detection errors (including octave jumps and detection of periodicity in unvoiced speech segments) were manually corrected. All analysis was carried out blind in regard to knowledge of whether speech samples were recorded within the stress or control group.

### *4.2.4 Statistical Analyses*

All statistical analyses were performed with SPSS 20.0. Testing was two-tailed and p-values smaller than 0.05 were considered to indicate statistical significance. Mean ± SD data are reported in text.

The dependent variables mean F0, mean heart rate (HR) and mean systolic arterial pressure (SAP) were analyzed in a repeated measures 2 x 2 ANOVA, with the factors “group” as a two-level between-subjects factor (control vs. CPT stress), and “condition” as a two-level within-subjects factor (baseline vs. water) for each dependent variable.

Statistical analysis of heart rate and blood pressure data was used to assess whether the CPT intervention was successful. Because we were interested in the difference between control and stress intervention, we were focused on studying interactional effects of condition x group. If condition x group interaction was significant, separate ANOVAs for control and stress group were carried out and results from these within-group comparisons were considered to reflect effects of the CPT.

## 4.3 Results

### *4.3.1 Cardiovascular Data*

Our results showed a significant condition x group interaction for mean HR ($F_{(1,38)}=7{,}145$, $p<.05$, $\eta 2=.158$) and SAP ($F_{(1,38)}=13{,}995$, $p<.001$, $\eta 2=.269$). Further ANOVAs performed for control and stress group, separately, showed a significant increase of HR by 5 bpm ($F(1,19)=12,43$, $p<.00$, $\eta 2=.392$) and SAP by 20 mmHg ($F(1,19)=31,57$, $p<.00$, $\eta 2=.624$) in the stress group during water intervention relative to baseline speech condition. No significant differences were found in the control group. The CPT intervention was considered successful.

#### *4.3.2 Vocal fundamental frequency*

No significant differences between control and stress group in absolute mean F0 values were found (control group: 211 +/- 21 Hz; stress group: 218 +/- 12 Hz). Analysis of Mean F0 revealed a significant group x condition interaction ($F(1,38)=13,84$, $p<.00$, $\eta 2=.267$). Further application of ANOVAs for stress and control group, separately, show a significant moderate increase of mean F0 (+3,7 Hz) during water intervention (216 Hz, +/-3) in the stress group ($F(1,19)=13,30$, $p<.00$, $\eta 2=.412$), but no significant differences in the control group.

### 4.4 Discussion

The aim of this study was to investigate the influence of acute stress on F0 level as tested by the influence of the CPT. The cardiovascular response pattern showed significant increases in HR and SAP within typical response ranges (Lovallo, 1975) indicating that CPT was successful in eliciting a physiological stress response including sympathetic activation. Our results showed a moderate increase of mean F0 during CPT and support earlier findings of a stress-induced acute rise of F0.

Our findings present further evidence for the linkage of autonomic arousal and F0 level. However, it is known that F0 is always shaped by a great variety of psychobiological influences at the same time. Besides psychobiological mechanisms, F0 is also influenced by external factors such as socio-cultural influences (Bachorowski & Owren, 2003a; Juslin & Laukka, 2003), a fact that Scherer conceptualizes as “push” and “pull” effects on the voice (Scherer, 1986, 2003). "Push effects" refer to physiological processes involved in producing speech and shaping the voice signal, such as for example muscle tension and respiration, which are “naturally” influenced by emotions. In contrast, "pull effects" include social norms and display rules, reflecting the fact that vocal expression is often closely monitored and regulated by cognitive processes. This is the case, for example, when speakers intentionally express a feigned emotional state.

Intuitively, it stands to reason that our findings on CPT-stress-induced F0 changes could be modulated by social pull factors. Our CPT paradigm included social-evaluative elements to increase the psychophysiological stress reaction in comparison to a standard CPT (Schwabe, Haddad, & Schächinger, 2008), but by doing so it also added a social component, which was not applied in the warm water control intervention. It can be argued that the presence of a running camera and of an opposite sex experimenter in the room may have influenced our F0 results not only on the push but also on the pull level, e.g. if participants were signalling a certain attitude towards the experimenter and/or trying to create a certain impression on the listener through F0 modulation (Bachorowski & Owren, 2003). One way to shed more light on the actual nature of those pull effects would be to investigate the mere presence effects of an opposite-sex experimenter in combination with an obviously running camera during a warm water control procedure. Since our main objective was to use the CPT and the respective control condition as a validated laboratory stress procedure, we did not add this social-evaluative component to our control procedure.

Moreover, there are reasons to assume that our F0 results are not primarily due to differential pull influences in control and stress group. All participants in both stress and control group were aware that their speech was recorded and that they were steadily monitored via video- and audio transmission by a group of experimenters and their laboratory helpers throughout the whole experiment. This social setting was the same for control and stress group. Only in the stress group, an activation of a second camera and physical presence of the experimenter was added. Therefore, the physical presence of the experimenter in the stress group compared to the "psychological" presence of the experimenter in the control group should evoke comparatively similar pull influences on F0, e.g. in both groups, the female participant knew that a male experimenter was listening and hypothetically adapted F0 level following similar social display rules. Thus, we assume that the reported differences in F0 levels between control and stress group are primarily due to psychophysiological push effects of CPT treatment and that hypothetical differences in social pull factors between control and stress group played a rather minor role in modulating push effects on F0. Still, this assumption needs further research.

Generally, we wanted to obtain speech samples from speech behavior as "naturalistic" as possible in a laboratory environment. This implied that the attention of the participants should be kept away from the real purpose of our study. For this reason, we did not assess subjective ratings of affective valence. We did not want the participants to explicitly reflect and report their affective states during the experiment to avoid possible influences of affective awareness on F0 levels. There was no instruction to the participants to express their affective state or stress level with their voice. There are comparatively few studies that use standardized speech samples with non-instructed emotional expression and, as Scherer points out, more research investigating more naturalistic speech samples other than portrayed emotion is required (Scherer, 2013). One advantage of using a well-documented stress induction method as in the case of the CPT, is that subjective ratings are not necessarily needed to assess affective valence of the cold water treatment, allowing speech samples which are more naturalistic to be obtained.

It is known that CPT intervention is generally perceived as negative or aversive, as shown by a large corpus of empirical evidence (Deuter et al., 2012; Lovallo, 1975; Peckerman et al., 1994). Although we did not assess subjective ratings from our participants on pleasantness, pain or stress levels, it can be assumed that our CPT intervention was perceived as aversive by our participants. Concerning the influence of affective valence on F0 level, it is often argued that positive or negative affective valence alone does not lead to changes in F0 levels and that affective modulation of F0 is rather linked to the general arousal levels of a speaker (Hammerschmidt & Jürgens, 2007b; Jürgens, 2009; Russell, Bachorowski, & Fernandez-Dols, 2003; Scherer, 2013).

However, as a side note, it might be interesting in a future study to include the dimension of cognitive appraisal and investigate the relationship of valence ratings, arousal and F0 level, e.g. to investigate individual differences in stress perception and F0 level. Moreover, it is possible that the reading task itself modulates affective valence ratings on CPT intervention. As a fact, the CPT of the present study reliably shows a typical cardiovascular response pattern during the reading task, but it has been shown that distraction from the pain stimulus (Hodes et al., 1990), especially interactive distraction (Dahlquist et al., 2007), influences subjective pain ratings and increases pain tolerance during CPT. Possibly, the reading task functions as distractor and helps coping with CPT stress, which adds an interesting notion to our experiment from a more general methodological perspective.

## 4.4 Conclusion

Our study shows how the CPT may be applied to investigate effects of acute stress on speech and discusses some preliminary results and thoughts on F0 increase during CPT stress. Replications and further developments of this experiment are needed to verify our results and deepen our understanding about which psychophysiological mechanisms are primarily involved in shaping speech during stress. The cardiovascular pattern during CPT intervention supports the idea that autonomic activation may play an important a role in the rise of F0 during acute stress. Still, more research and empirical evidence is needed. To the best of our knowledge, our study is the first one to investigate the influence of CPT stress on F0. Our results support earlier findings of elevated F0 levels during acute stress.

## References


Allman, J. M., Hakeem, A., Erwin, J. M., Nimchinsky, E., & Hof, P. (2001). The Anterior Cingulate Cortex. *Annals of the New York Academy of Sciences, 935*(1), 107-117.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*(2), 256-274.

Anders, L., Thomas, B., Nancy, S. M., & Robin Seider, S. (1988). The crico-thyroid muscle in voicing control. *The Journal of the Acoustical Society of America, 83*(S1), S111-S112.

Anderson, S. W., & Cooper, W. E. (1986). Fundamental frequency patterns during spontaneous picture description. *The Journal of the Acoustical Society of America, 79*(4), 1172-1174.

Atkinson, J. (1978). Correlation analysis of the physiological features controlling fundamental voice frequency. *The Journal of the Acoustical Society of America, 63*, 211-222.

Bachorowski, J.-A., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science, 6*, 219-224.

Bachorowski, J.-A., & Owren, M. J. (2003a). Reconsidering the Evolution of Nonlinguistic Communication: The Case of Laughter. *Journal of Nonverbal Behavior, 27*(3), 183-200.

Bachorowski, J.-A., & Owren, M. J. (2003b). Sounds of Emotion. *Annals of the New York Academy of Sciences, 1000,* (Emotions Inside Out: 130 Years after Darwin's The Expression of the Emotions in Man and Animals), 244-265.

Baker, S., Weinrich, B., Bevington, M., Schroth, K., & Schroeder, E. (2008). The effect of task type on fundamental frequency in children. *International Journal of Pediatric Otorhinolaryngology, 72*(6), 885-889.

Barrett, J., & Paus, T. (2002). Affect-induced changes in speech production. Experimental Brain Research, 146(4), 531-537.

Barrett, J., Pike, G. B., & Paus, T. (2004). The role of the anterior cingulate cortex in pitch variation during sad affect. *European Journal of Neuroscience, 19*(2), 458-464.

Barsties, B. (2013). Effects of different tasks on determination of the speaking fundamental frequency. *HNO, 61*(7), 609-616.

Benson, P. (1995). Analysis of the acoustic correlates of stress from an operational aviation emergency, ESCA/NATO Tutorial Research Workshop on Speech Under Stress. Lisbon: SUS-1995, 61-64.

Bezooijen, R. v. (1995). Sociocultural Aspects of Pitch Differences between Japanese and Dutch Women. *Language and Speech, 38*(3), 253-265 .

Boersma, P., & Weenink, D. (2007a). *Praat.* Dallas, TX.

Boersma, P., & Weenink, D. (2007b). *Praat, a system for doing Phonetics by computer.* Dallas, TX.

Botvinick, M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective, & Behavioral Neuroscience, 7(*4), 356-366.

Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *Journal of Zoology, 288*(1), 1-20.

Buck, R., Losow, J. I., Murphy, M. M., & Costanzo, P. (1992). Social facilitation and inhibition of emotional expression and communication. *Journal of Personality and Social Psychology, 63*(6), 962-968.

Cartei, V., Cowles, H. W., & Reby, D. (2012). Spontaneous voice gender imitation abilities in adult speakers. *Plos One, 7*(2), 17.

Conrod, P. J. (2006). The role of anxiety sensitivity in subjective and physiological responses to social and physical stressors. *Cognitive Behavior Therapy, 35*(4), 216-225.

Contrada, R. J. (2009). Stress, Adaptation and Health. In R. J. Contrada & A. Baum (Eds.), *The Handbook of Stress Science: Biology, Psychology and Health* (pp. 1-10). New York.

Cowie, R., & Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech Communication, 40*(1–2), 5-32.

Critchley, H. D., Tang, J., Glaser, D., Butterworth, B., & Dolan, R. J. (2005). Anterior cingulate activity during error and autonomic response. *NeuroImage, 27*(4), 885-895.

Dahlquist, L. M., McKenna, K. D., Jones, K. K., Dillinger, L., Weiss, K. E., & Ackerman, C. S. (2007). Active and Passive Distraction Using a Head-Mounted Display Helmet: Effects on Cold Pressor Pain in Children. *Health Psychology, 26*(6), 794-801.

Damian, M. F. (2000). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(3), 416-431.

Darwin, C. (1998). *The Expression of the Emotion in Man and Animals.* John Murray, London. (Reprinted with introduction, afterword and commentary by P. Ekman, Ed.) Oxford University Press, New York (Original work published 1872).

Deuter, C. E., Kuehl, L. K., Blumenthal, T. D., Schulz, A., Oitzl, M. S., & Schächinger, H. (2012). Effects of Cold Pressor Stress on the Human Startle Response. *Plos One, 7*(11).

Dreisbach, G., & Fischer, R. (2012). Conflicts as aversive signals. Brain and Cognition, 78(2), 94-98.

Dreisbach, G., & Fischer, R. (2012). The role of affect and reward in the conflict-triggered adjustment of cognitive control. *Frontiers in Human Neuroscience, 6*(342), 31.

Drolet, M., Schubotz, R., & Fischer, J. (2012). Authenticity affects the recognition of emotions in speech: behavioral and fMRI evidence. *Cognitive, Affective, & Behavioral Neuroscience, 12*(1), 140-150.

Drolet, M., Schubotz, R. I., & Fischer, J. (2014). Recognizing the authenticity of emotional expression: F0 contour matters when you need to know. Frontiers in Human Neuroscience, 8.

Eckert, H., & Laver, J. (1994). *Menschen und ihre Stimmen - Aspekte der vokalen Kommunikation.* Weihnheim: Beltz.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*(3), 169 - 200.

Fitch, J. L. (1990). Consistency of fundamental frequency and perturbation in repeated phonations of sustained vowels, reading, and connected speech. *Journal of Speech and Hearing Disorders, 55(*2), 360-363.

Fritz, J., & Dreisbach, G. (2013). Conflicts as aversive signals: Conflict priming increases negative judgments for neutral stimuli. *Cognitive, Affective, & Behavioral Neuroscience*, 1-7.

Gandour, J., Tumtavitikul, A., & Satthamnuwong, N. (1999). Effects of Speaking Rate on Thai Tones. *Phonetica, 56*(3-4), 123-134.

Giddens, C. L., Barron, K. W., Byrd-Craven, J., Clark, K. F., & Winter, A. S. (2013). Vocal indices of stress: a review. *Journal of Voice, 27*(3), 23.

Grandjean, D., Bänziger, T., & Scherer, K. (Eds.). (2006). Intonation as an interface between language and affect. *Progress in Brain Research, 156,* 235-47.

Greasley, P., Sherrard, C., & Waterman, M. (2000). Emotion in language and speech: methodological issues in naturalistic approaches. *Language and Speech, 43*(Pt 4), 355-375.

Gregory, S. W. J., & Webster, S. (1996). A Nonverbal Signal in Voices of Interview Partners Effectively Predicts Communication Accommodation and Social Status Perceptions. *Journal of Personality and Social Psychology, 70*(6), 1231-1240.

Hammerschmidt, K., & Jürgens, U. (2007). Acoustical Correlates of Affective Prosody. *Journal of Voice, 21*(5), 531-540.

Hansen, J. H. L., & Patil, S. (2007). Speech under Stress: Analysis, Modeling, recognition. In C. Müller (Ed.), *Speaker Classification I* (pp. 108-137). Heidelberg.

Harrigan, J., Rosenthal, R., & Scherer, K. (2005). *The New Handbook of Methods in Nonverbal Behavior Research (Series in Affective Science)*. Oxford University Press, USA.

Hines, E. A., Jr. (1940). Significance of vascular hyperreaction as measured by cold-pressor test. *American Heart Journal, 19,* 408-416.

Hodes, R. L., Howland, E. W., Lightfoot, N., & Cleeland, C. S. (1990). The effects of distraction on responses to cold pressor pain. *Pain, 41*(1), 109-114.

Hoh, J. F. (2005). Laryngeal muscle fibre types. *Acta Physiologica Scandinavica, 183*(2), 133-149.

Hollien, H. (1980). Vocal Indicators of Psychological Stress. *Annals of the New York Academy of Sciences, 347*(1), 47-72.

Ishihara, S. (1917). *Tests for colour-blindness.* Tokyo: Handaya.

Jessen, M. (2006). *Einfluss von Stress auf Sprache und Stimme.* Universität Bielefeld, Bielefeld.

Jürgens, R., Drolet, M., Pirow, R., Scheiner, E., & Fischer, J. (2013). Encoding conditions affect recognition of vocally expressed emotions across cultures. *Frontiers in Psychology, 4*(111).

Jürgens, R., Hammerschmidt, K., & Fischer, J. (2011). Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Frontiers in Psychology, 2*(180).

Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews 26*(2), 235-258.

Jürgens, U. (2009). The Neural Control of Vocalization in Mammals: A Review. *Journal of voice: Official Journal of the Voice Foundation, 23*(1), 1-10.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770-814.

Kallai, I., Barke, A., & Voss, U. (2004). The effects of experimenter characteristics on pain reports in women and men. *Pain, 112*(1), 142-147.

Kawamoto, A. H., Liu, Q., Mura, K., & Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *Journal of Memory and Language, 58*(2), 347-365.

Kleinow, J., & Smith, A. (2006). Potential interactions among linguistic, autonomic, and motor factors in speech. *Developmental Psychobiology, 48*(4), 275-287.

Knoll, M., Scharrer, N., & Costall, A. (2009). Are actresses better simulators than female students? The effects of simulation on prosodic modifications of infant- and foreigner-directed speech. *Speech Communication 51*(3), 296-305.

Kobayashi, N., Yoshino, A., Takahashi, Y., & Nomura, S. (2007). Autonomic arousal in cognitive conflict resolution. *Autonomic Neuroscience, 132*(1–2), 70-75.

Ladd, D., & Silverman, K. (1984). Vowel Intrinsic Pitch in Connected Speech. *Phonetica, 41*(1), 31-40.

Lang, P. J., Bradley, M., & Cuthbert, B. (1999). International affective picture system (IAPS): Technical manual and affective ratings. Center for Research in Psychophysiology: University of Florida.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30*, 261.

Laukka, P. (2008). In a Nervous Voice: Acoustic Analysis and Perception of Anxiety in Social Phobics' Speech. *Journal of Nonverbal Behavior, 32*(4), 195-214.

Laukka, P., Audibert, N., & Aubergé, V. (2007). Graded structure in vocal expression of emotion: what is meant by 'prototypical expression'?," 1st International Workshop on Paralinguistic and Speech – Between Models and Data. Saarbrücken.

Laukka, P., Neiberg, D., Forsell, M., Karlsson, I., & Elenius, K. (2011). Expression of affect in spontaneous speech: Acoustic correlates and automatic detection of irritation and resignation. *Computer, Speech and Language,* 25(1), 84-104.

Laver, J. (1994). *Phonation, Principles of Phonetics* (pp. 184-204). Cambridge: Cambridge University Press.

Liotti, M., Woldorff, M. G., Perez III, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologica (38)*, 701-711.

Löfqvist, A. (1990). Speech as Audible Gestures. In W. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 289-322): Springer Netherlands.

Lovallo, W. (1975). The Cold Pressor Test and Autonomic Function: A Review and Integration. *Psychophysiology, 12*(3), 268-282.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin, 109*(2), 163-203.

Mc Leod, C. M. (2005). The Stroop task in cognitive research. In A. W. D. C. Rubin (Ed.), *Cognitive methods and their application to clinical research* (pp. 17-40). Washington, DC: American Psychological Association.

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review, 110*(3), 422-471.

Moriyama, K., & Hirotoshi, I. (2007). Assessment of cardiac contractility during a cold pressor test by using (dP/dt)/P of carotid artery pulses. *European Journal of Applied Physiology, 100*(2), 185-191-191.

Mourot, L., Bouhaddi, M., & Regnard, J. (2009). Effects of the Cold Pressor Test on Cardiac Autonomic Control in Normal Subjects. Physiological Research, 58, 83-91.

Ohala, J. (1984). An Ethological Perspective on Common Cross-Language Utilization of F0 of Voice. *Phonetica, 41*, 1-16.

Ohala, J. (Ed.). (1994). *The biological bases of sound symbolism*. Cambridge: Cambridge University Press.

Ohala, J., & Eukel B. W. (1987). Explaining the Intrinsic Pitch of Vowels. In C. R & S. L (Eds.),*In honor of Ilse Lehiste*. Dordrecht: Foris.

Paus, T. (2001). Primate anterior cingulate cortex: where motor control, drive and cognition interface. *Nature, 2*(6), 417-424.

Peckerman, A., Hurwitz, B. E., Saab, P. G., Llabre, M. M., McCabe, P. M., & Schneiderman, N. (1994). Stimulus dimensions of the cold pressor test and the associated patterns of cardiovascular response. *Psychophysiology, 31*(3), 282-290.

Pieper, S., Brosschot, J. F., van der Leeden, R., & Thayer, J. F. (2007). Cardiac effects of momentary assessed worry episodes and stressful events. *Psychosomatic Medicine, 69*(9), 901-909.

Protopapas, A., & Liebermann, P. (1997). Fundamental frequency of phonation and perceived emotional distress. *Journal of the Acoustical Society of America, 101*(4), 2267-2277.

Quinto, L. R., Thompson, W. F., & Keating, F. L. (2013). Emotional Communication in Speech and Music: The Role of Melodic and Rhythmic Contrasts. *Frontiers in Psychology, 4*.

Renaud, P., & Blondin, J.-P. (1997). The stress of Stroop performance: physiological and emotional responses to color-word interference, task pacing, and pacing speed. *International Journal of Psychophysiology, 27*(2), 87-97

Rothkranz, L. J. M., Wiggers, P., van Wees, J.-W. A., & van Vark, R. J. (2004). Voice stress analysis. In P. K. Sojka, Ivan; Pala, Karel (Ed.), *Text, Speech and Dialogue* 2004 (pp. 449-456). Heidelberg.

Ruff, C. C., Woodward, T. S., Laurens, K. R., & Liddle, P. F. (2001). The Role of the Anterior Cingulate Cortex in Conflict Processing: Evidence from Reverse Stroop Interference. *Neuroimage, 14*(5), 1150-1158.

Russell, J. A. (2009). Emotion, core affect, and psychological construction. *Cognition & Emotion, 23*(7), 1259-1283.

Russell, J. A., Bachorowski, J. A., & Fernandez-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54,* 329-349.

Sabri, M., Melara, R. D., & Algom, D. (2001). A confluence of contexts: Asymmetric versus global failures of selective attention to Stroop dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 27*(3), 515-537.

Scherer, K. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*(2), 143-165.

Scherer, K. (1993). Vocal Expression and Communication of Emotion. In: *Handbook of Emotions*. London.

Scherer, K. (2003). Vocal communication of emotion: a review of research paradigms. *Speech Communication, 40*(1-2), 227-256.

Scherer, K., Banse, R., & Wallbott, H. G. (2001). Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology, 32*(1), 76-92.

Scherer, K., & Johnstone, T. (2000). Vocal communication of Emotion. In: *The Handbook of Emotion*. NewYork: Guilford.

Scherer, K. (2013). Vocal markers of emotion: Comparing induction and acting elicitation. *Computer Speech & Language, 27*(1), 40-58.

Schwabe, L., Haddad, L., & Schächinger, H. (2008). HPA axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology, 33*(6), 890-895.

Siegle, G. J., Steinhauer, S. R., & Thase, M. E. (2004). Pupillary assessment and computational modeling of the Stroop task in depression. *International Journal of Psychophysiology, 52*(1), 63-76.

Streff, A., Kuehl, L. K., Michaux, G., & Anton, F. (2010). Differential physiological effects during tonic painful hand immersion tests using hot and ice water. *European Journal of Pain, 14*(3), 266-272.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Titze, I. R. (1994). *Principles of voice production*. Prentice Hall.

Tomaka, J. B., Jim; Kibler, Jeffrey; Ernst, John M. (1997). Cognitive and physiological antecedents of threat and challenge appraisal. *Journal of Personality and Social Psychology, 73*(1), 63-72.

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication. 49*(1), 2-7.

Victor, R. G., Leimbach, W. N., Jr., Seals, D. R., Wallin, B. G., & Mark, A. L. (1987). Effects of the cold pressor test on muscle sympathetic nerve activity in humans. *Hypertension, 9*(5), 429-436.

Westermann, R., Spies, K., Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: a meta-analysis. *European Journal of Social Psychology, 26*(4), 557-580.

Wittels, P., Johannes, B., Enne, R., Kirsch, K., & Gunga, H.-C. (2002). Voice monitoring to measure emotional load during short-term stress. *European Journal of Applied Physiology, 87*, 278-282.

Yamamoto, K., Iwase, S., & Mano, T. (1992). Responses of Muscle Sympathetic Nerve Activity and Cardiac Output to the cold Pressor Test. *Japanese Journal of Physiology, 42*, 239-252.

Zraick, R. I., Birdwell, K. Y., & Smith-Olinde, L. (2005). The effect of speaking sample duration on determination of habitual pitch. *Journal of Voice, 19*(2), 197-201.

Zraick, R. I., Skaggs, S. D., & Montague, J. C. (2000). The effect of task on determination of habitual pitch. *Journal of Voice, 14*(4), 484-489.

## Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbst verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe. Zudem wurde die Arbeit an keiner anderen Universität zum Erlangen eines wissenschaftlichen Grades eingereicht.

Trier, den 12.08.2014

Debora Elisabeth Plein