

# 8 Phonetic investigation of aggressive to resigned spontaneous speech dialogues

Cornelia Wieck

## 8.1 Introduction

Emotions are not only conveyed by facial expressions. This statement is supported by the case that we are quite able to identify emotional states without any eye contact. Whether in making telephone calls with a familiar person or while listening to a foreign conversation: human beings possess the ability to accurately recognize the emotional condition of their fellow men by reference to the voice.

Several empirical studies have shown how easily listeners imply the emotional state of the speakers exclusively from the sound of the voice (e.g. Fairbanks & Pronovost 1939; Banse & Scherer 1996). This competence suggests that receivers use certain features of speech to perceive and distinguish several emotional qualities.

Since many years researchers have investigated the characteristic elements which enable this distinction. But on the basis of the common acoustic parameters it was not possible to identify a fixed set of features. Therefore, the present study aims at expanding the knowledge about the emotional expression of the voice.

In order to allow an analysis of the acoustic correlates concerning the specific emotions, it is necessary to record emotional speech material. Because of the difficulty in gaining spontaneous material, the majority of studies focuses on simulated and induced emotions.

The purpose of the present study is to collect a data base of spontaneous emotional speech. With the help of this material, the question of whether there are differences between specific emotional expressions vs. neutral speech will be analysed on the basis of several acoustic parameters.

## 8.2 Facets of an emotional construct

In discussing the phenomenon of emotion, most researchers firstly complain about the absence of a generally accepted definition (e.g. Schmidt-Atzert 1996, p. 18; Otto et al. 2000, p. 11). Over and over, scientists are confronted with the question why the term *emotion* is used without any problems in mass media and in interpersonal usage, although an exact description of the phenomenon appears to be quite impossible. A very suitable remark related to this issue was made by Fehr & Russell (1984): "Everyone knows what an emotion is, until asked to give a definition. Then, it seems, no one knows." (1984, p. 464)

### 8.2.1 Adjacent or related: Emotions and other states

The following phenomena are based on the classification of affective states by Siemer (1999). On the one hand, the terms *mood*, *affect* and *feeling* deliver some definitional differentiations among each other as well as with regard to emotions. On the other hand they show commonalities.

#### Mood

Many authors differentiate between emotion and mood. The focus lies on three distinctive features: (1) duration, (2) intensity and (3) diffusion/globality. In the case of duration moods are described as longer than emotions (e.g. Ewert 1983; Davidson 1994; Ekman 1994). Furthermore, Parkinson et al. (2000) claim that moods have a presence in our everyday lives. This leads to the conclusion that these states neither have a fixed starting point nor an ending point. Emotions act as "troublemakers" which induce a phasic interruption (Davidson 1994). Thus, they are characterized by a clear starting point. Parkinson et al. (2000, p. 17) describe emotions as "acute", while moods are "chronic" or "tonic".

Another distinguishing criterion constitutes the intensity. In its level of intensity emotions are classified as more pronounced than moods (e.g. Siemer 1999). Upon closer inspection it can be noticed that moods are also intense in their appearance. Depressions or anxieties, for instance, can involve moods which take a long time. Emotions can also appear at a lower or barely perceptible intensity (e.g. in the form of gratefulness) (Morris 1989).

The most conducive criterion for discriminating mood and emotion is globality or diffusion. Experts agree that emotions are targeted towards a concrete goal (e.g. Parkinson et al. 2000). Moods of course have causes, but they are lacking in directedness on a specific object. Therefore, moods are considered to be non-intentional states.

#### Affect

Juslin & Scherer (2005) claim that "affect" is a universal term which interconnects a plurality of phenomena such as emotions, stress, moods, interpersonal relations, preferences as well as dispositions. The German term "Affekt" does not play an important role, neither in emotion psychology nor in behavioral science. In psychiatric institutions affects describe short-term and intensive emotions which are accompanied by a loss of active control (Otto et al. 2000, p. 13).

## Feeling

Sometimes “feeling” is put on a level with emotion (e.g. James 1884; Cannon 1927; Ewert 1983), sometimes the terms are distinguished. In the latter case feelings were only associated with the subjective component; however, emotions include physiological reactions such as palpitation and shiver as well as the motor expressions (Schmidt-Atzert 1996). That means, feelings only have one aspect of emotion: to feel. Aspects such as motor response and physiological responses were not taken into account. Its sole purpose is to feel an emotion (cf. Merten 2003, p. 10).

### 8.2.2 Structure of emotions

For systematization of emotions the literature offers three different approaches: (1) the categorical or discrete approach, while the representatives postulate a set of basic emotions, whose number fluctuates between two and 18, (2) the dimensional approach that deals with a list of as few basic dimensions as possible and (3) the component process model by Scherer (1984), which considers that physical expression (including the voice) is driven by the nature of cognitive assessment.

#### Concept of basic emotions

The representatives of the discrete emotion approaches attribute a subset of emotions to a separate status, which is usually referred to as elementary, primary or fundamental. In the literature two opinions which support the concept of basic emotions can be found: (1) fundamental emotions are attributable to psychological mechanisms, which have developed during the evolution through adjustment problems; that means they have an adaptive function and/or (2) they provide the basis for complex emotions (similar to the colour concept). However, they are not reducible to other emotions. “Thus, the basic-emotion theorists claim, that a subset of human emotions is psychologically and/or biologically fundamental.” (Reisenzein 2000, p. 205)

One of the most prominent representatives is Plutchik (1993, p. 57), who suggests eight innate basic emotions (acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise). By means of a *two-dimensional* model (Fig. 8.1) he arranges these primary emotions in terms of their qualitative similarity as well as their oppositeness. Qualitatively similar emotions are presented next to each other (anger & disgust); dissimilar ones are oppositional (joy & sadness). As a result of the combination of elementary emotions new and complex ones arise (acceptance + joy = love). By adding a vertical intensity dimension, Plutchik’s *three-dimensional* model is created (Fig. 8.1, upper left corner).

In spite of everything, the theory was confronted with a series of reviews. Ortony & Turner (1990) criticized the poor agreement between the theorists with regard to the *number* and the *designation* of the basic emotions as well as the *admission criteria*, which an emotion has to fulfil to

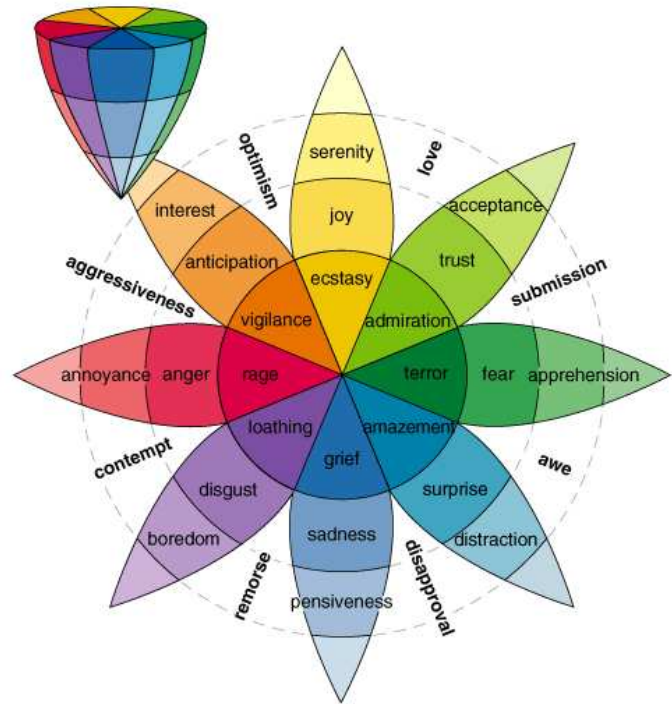


Fig. 8.1: Two-dimensional model as well as three-dimensional cone shaped model of emotions by Plutchik (1993).

rank among the elementary emotions. Furthermore, different researchers assign different terms to the same emotions (e.g. some researchers use the term “anger”, others the item “rage” to describe the same state).

#### Dimensional model

In dimensional models, emotions are defined as a position in a continuous multidimensional space where each dimension stands for a fundamental property common to all emotions. At the beginning of the scientific development a three-dimensional concept of Wundt (1896, p. 98) was already used. He formulated three bipolar independent dimensions: pleasantness–unpleasantness, rest–activation and relaxation–attention. Because of reconfirmation the dimensions most accepted were valence and arousal. For the last dimension the findings were ambiguous and contradictory (cf. Schmidt-Atzert 1996, p. 89).

Tischer (1993, p. 31) criticized that the dimensions of valence and arousal have an insufficient power to differentiate emotions. For example, the emotions anger and fear are characterized by a negative valence and a high arousal; but subjectively they are felt very differently.

Moreover, the classification of emotional states can just be implemented in one direction. If a concrete term (e.g. joy) is present, it is possible to classify it with several axes. But in case there is just one point without the respective concept, it is hard to draw a conclusion on the real emotion. Thus, the point could represent an almost unlimited number of emotions.

## 8.3 Theories of emotion

The questions how an emotion develops, which functions it fulfills or which causes it underlies were answered by emotion theorists in various ways. Accordingly, it can be the perception of one's own physiological alteration but also the subjective evaluation of a result which is authoritative for the development of an emotional state.

### 8.3.1 Theory of evolution

Charles Darwin is regarded as a pioneer of the evolutionary psychology approach. In his opus 'The expressions of emotions in man and animals' (1872) he makes a significant contribution to the general nature of emotions as well as to the phylogeny of the facial expressions. His opinion is based on the fact that emotions are "conscious mental states of persons and higher animals" (Meyer et al. 2003, p. 37), which can be triggered by evaluations and appraisals of issues. In accordance with Darwin emotions enhance the chance of individual survival because they supply adequate problem-solving reactions to challenges of the outside world (Niedenthal et al. 2006).

His main field of research focused on the emotional facial expression whereby he used the following three aspects for his hypotheses: (1) the universality of facial expression, (2) its similarity between human and animal, and (3) its actual genesis.

In order to prepare an intercultural comparison of expressive behavior Darwin developed a questionnaire which he sent to different persons in distant countries. A group of observers sent back their impressions which they gained in connection with the facial expressions of several human races.

The majority of these emotional expressions showed constant characteristics across different cultural borders. These and other observations (e.g. of congenitally blind people, children and people with mental diseases) brought Darwin to his universal hypothesis which proposes that the basic forms of expressions are universal and thus have not arisen by imitation but their existence is attributable to inheritance (Darwin 1872, p. 351).

Other evolutionary-biological oriented emotion psychologists (e.g. Plutchik 1984; Ekman 1992) deduced the thesis that only a limited number of primary emotions exist which can be differentiated from each other exactly.

### 8.3.2 Are we sad because we cry?

This simple statement pictures the relevant core of the emotion theory by William James (1884). Contrary to everyday life concerning the emotion genesis the psychologist postulated the following thesis: "the bodily changes follow directly the PERCEPTION of the exciting fact, and that our feeling of the same changes as they occur IS the emotion" (1884, p. 189). Thus, the cause of physiological reactions

is not the emotional experience but the mere perception of a physical change.

Independent from this conception, one year later the physiologist Carl Lange formulated a comparable theory. He assumed that vasomotor reactions are responsible for the development of emotions while James believed that only visceral reactions are important (Lange & James 1922, quoted in Meyer et al. 2001). Due to the huge similarity of both models one speaks of the *James-Lange Theory*.

The essential aspect of James' theory is that the process of object perception and the physical change react by reflex and therefore there is no place for conscious cognitive activities. Furthermore, several emotions were recorded to a different extent because the underlying reason was specific visceral changes.

### 8.3.3 Two-factor theory

Even the postulated theory by Marañon in 1924 (quoted in Schachter & Singer 1962) forms the basis for one of the most popular theory of emotion research. For Marañon physiological components as well as the psychological components play a major role in emotional formation. Building up on this, Schachter & Singer (1962) claim that only the interaction of the perceived activity of the autonomic nervous system and the cognitive interpretation of this activity is responsible for the sensation of a specific emotion.

According to this theory, the perception of one's own unspecific physiological state of excitement (e.g. increased sweat production, trembling hands, increased breath and heart frequency) triggers the need to explore and to categorise the reason of this *arousal*. The excitement is assigned to a 'label' which determines the quality of the emotion. The search for the informative cues from a surrounding area, which are suitable for the clarification of physical arousal, is running at full speed. Which current situation turns out to be responsible for the condition and how is it labeling? From a previously unspecific arousal a qualitatively differentiable emotional state originates (Schachter 1971, pp. 3ff.).

In summary, Schachter & Singer consider — similar to James — the perceived physiological arousal particularly essential for the experience of an emotion. But a second factor is also important: the cognitive evaluation. Thus, they deem, in contrast to the theory of James, the interaction of both factors as sufficient for emotion genesis.

### 8.3.4 Appraisal theories

Why Mr. Smith reacts furiously to the cuts in unemployment benefits while Mr. Potter looks away bashfully during the confrontation with the same event whereas Mr. Cooper collapses sobbingly?

The discovery that different people show varied emotional characteristics even though they are subjected to the same situation leads the researchers to the development of appraisal theories.

The underlying premise of this kind of theory implies that emotions result from a series of evaluations of a significant event or situation. This critical assessment takes place in relation to the personal desires of the individual and influences the form and intensity of the emotional characteristic (Scherer 1986). In this respect, the proponents of the cognitive emotion theory agree. But the number of necessary steps of evaluation for emotional generation as well as the question whether cortical or subcortical regions participate in the stimulus evaluation are highly controversial.

Arnold (1960) presented her thoughts in the age of cognitive revolution by opposing the behaviorism with emphasis on the subjective experiences. In her introspective mentalistically-based theory she postulated three steps of emotional generation: (1) the record or the perception of the circumstances, (2) its critical evaluation (whereby an evaluation is conducted as useful or harmful) and (3) the resulted act, which is experienced as emotion (1960:171). During this procedure three further dimensions are used: the evaluation, presence-absence and the coping skills.<sup>1</sup>

Lazarus (1991) — similar to Arnold — supposed several gradually steps of evaluations. He carried out a rough dichotomy by differentiating between primary and secondary appraisal.

The ideas that evaluations are sequential and the sensation of an emotion changed with every new appraisal were mentioned first by Scherer (1984); however, they are compatible with other appraisal theories. In connection with this statement he designed the *Component Process Model* of affective states. Emotions are defined as products of different processing steps (Scherer 1984, 1986; Ellsworth & Scherer 2003). According to this model persons consistently perform an appraisal of their actual situation. Scherer (1984) speaks of a successive processing of information on different levels.

In temporally fixed series five (unconscious) examination steps proceed: the so-called ‘Stimulus Evaluation Checks’ (SEC). The starting point of the evaluation process constitutes an internal or external event which shows a change of state which may either comprise the appearance of an unexpected situation or the absence of an expected event. These events are evaluated with regard to (1) novelty, (2) intrinsic pleasantness, (3) goal/need conduciveness, (4) coping potential and (5) norm/self compatibility. By solving this appraisal pattern every emotion can be determined.

Despite of slight deviations between the views of the different appraisal theories, the proponents coincide on one fundamental point: emotions are adaptable and flexible. That means, they have the ability to decouple from the perception of the stimulus, which enable a reappraisal of the respective stimulus at any time (Ellsworth & Scherer 2003). Appraisal theories contrast with categorical theories which suppose a limited number of qualitatively different basic emotions. Appraisal theorists claim an unlimited spread of emotional experiences.

<sup>1</sup> Detailed information is provided by Reisenzein et al. (2003, pp. 61ff.).

## 8.4 Methods of data collection

The experimental investigation and description of emotion is one of the major problems in emotion research. On the one hand, emotional states are very personal. On the other hand, the external emotional utterances are influenced by social rules of conduct and therefore cannot be classified as authentic.

### 8.4.1 Authentic emotions

Previous empirical studies concentrated on analysing recorded material from air traffic, dialogues from psychotherapeutic treatments (e.g. Amir et al. 2010) or highly emotional reports (as the live commentary about the crash of the airship *Hindenburg*, Williams & Stevens 1972). Also, researchers use game or reality shows (Douglas-Cowie et al. 2000).

In addition to the ecological validity<sup>2</sup>, the authenticity of the emotions is an unbeatable argument, provided by the fact that the speaker is not conscious of the recordings.

The disadvantages are considered more closely. Not only short recordings of a small number of speakers complicate an analysis but also the bad quality of sounds (background noise). Furthermore, a control of verbal contents is barely possible. Therefore, an experimental monitoring of important acoustic parameters is very difficult.

Another disadvantage is the uncertainty of which a specific emotion is felt in the respective situation, that means the occurrence of “mixed emotions” cannot be excluded. During recordings taken from TV shows, aspects as e.g. showmanship or social rules of conduct (suppression of real emotions) lead to a negative effect on the material, because the authenticity of emotional utterances is only imperfectly assured (cf. Scherer et al. 2003).

### 8.4.2 Simulated emotions

In this case, professional or amateur actors are requested to imitate emotional states in a natural manner. The induction which the actors receive varies between the individual simulated studies. Either the performer is asked to produce his text (without a framework of content) by giving a precise demand, or the relevant statement is embedded in an emotional scenario in order to enable a better empathy in the respective emotion.

The biggest advantage is to yield emotional utterances without great effort and with a high acoustic quality. Thus, this procedure allows the investigation of any number of emotions. Specification of the produced utterances enables the control of the wording. Therefore, there is an opportunity to ensure a direct comparison of prosodic, phonetic and speech quality realizations (Juslin & Scherer 2005). Tischer

<sup>2</sup> Investigations are considered to be ecologically valid if they can be executed in natural areas of life. Because the results from inspected laboratory experiments could not be transferred to neutral surrounding, these experiments are classified as ecologically invalid.

(1993) points out that firstly the wording must be unspecific and secondly recoverable in every emotional characteristic.

An important point of criticism is the overproduction of stereotyped expressions by the actors. Just as often the presentation of sociocultural standards were criticized (Juslin & Scherer 2005, p. 96). Another factor lies in the incomparableness between acting executions and real emotional expressions. In order to attenuate the argument Banse & Scherer (1996) dealt with the naturalness of emotional everyday utterances. They pointed out that a high percentage of emotional expressions appear in interaction with other persons or rather in social contexts, whereby a control as well as a regulation of emotions is associated with it. Thus, Wallbott & Scherer (1986) doubt the naturalness of daily emotional states and use this weakness to emphasize positive factors of acting portrayals: “although simulated emotional expressions are clearly not *natural* enough, the natural expressions of emotion obtained in most studies to date have not been *emotional* enough.” (1986, p. 690)

Contrary to this, Douglas-Cowie et al. (2000) consider that “acted emotion cannot be a sufficient basis for conclusions about the expression of emotion. Proposals based on it need to be tested against natural material.” (2000, p. 1)

### 8.4.3 Induced emotions

As in all regions of field research the psychology of emotion is endeavored to develop standardized and easy-to-use methods that are inscrutable for the participants. Schmidt-Atzert (1996, p. 31) supports the induction techniques in the laboratory arguing that a showing of connections between cause and effect of emotions is made possible by experiments. A further benefit can be seen in the fact that a control of the surrounding factors, the time as well as the duration of the experiment, is possible.

In order to select out of a multitude of induction techniques, the investigator should know which procedure triggers what emotional state and what intervention level is affected (Sokolowski 1992). Moreover, the choice of induction should be targeted at specific groups, because various methods are suitable for different age and social groups. It must be taken into account that elaborated inductions which require a disproportionate effort, could lead to a negative basic attitude towards the experiment. An additional need is the compliance of ethical standards. Niedenthal et al. (2006) point out that the intensity of artificial emotions must not exceed the intense of daily emotion quantity (e.g. grief or pain). Furthermore, the participants have to be informed about the intention of the experiment. However, as this would lead to a distortion of the results, so-called *cover stories*, which provide a false but plausible explanation purpose, come into play.

In the last decades innumerable processes which are capable to elicit positive, negative and neutral emotions have been established. Westermann et al. (1996) compared and evaluated the effectiveness and validity of eleven different mood induction procedures in a major meta-analysis. In

total, 250 results of euphoric and depressive mood qualities were analyzed. Here, methods like (1) imagination, (2) Velten, (3) films/stories (with/without instruction), (4) music (with/without instruction), (5) feedback, (6) social interactions, (7) gifts or success message, (8) facial expression, and (9) combined methods have been borne in mind (Westermann et al. 1996, p. 567).<sup>3</sup>

## 8.5 Phonetic research of emotional speech

In order to characterize the vocal expression, different empirical studies have proposed a variety of acoustic features. However, it is largely unclear which of these correlates contribute to the specific emotional characteristic. Scientists speculated that the reasons for these difficulties might be the incomparableness of the studies among each other. One of the reasons is the imprecise use of emotion terms. Only few researchers expressed doubt about the fact that e.g. fear is part of the basic emotions (e.g. Frijda 1993), but there is wide disagreement about the demarcation of terror. While several studies subsume all different intensities of one emotion in one single class without any critical comments, other investigations provide an exact differentiation of the concepts. But generally, researchers claim that the usage of the various speech material as well as the different methods of investigation are problematic.

The most frequently studied acoustic parameter for measuring emotional expression is the fundamental frequency. Typically, the global  $f_0$ -contour, the mean, the standard deviation as well as the range are taken into consideration (e.g. Williams & Stevens 1972). In addition to the fundamental frequency, prosodic parameters such as (a) energy (or amplitude; perceived as intensity resp. loudness of the voice, e.g. Pfitzinger & Kaernbach 2008), (b) the distribution of the energy in the frequency spectrum (comparison between the relative energy in the high- vs. the low-frequency region, affecting the perception of voice quality or timbre) (c) the location of the formants ( $F_1$ ,  $F_2$ , ...,  $F_n$ , related to the perception of articulation), and (d) a variety of temporal phenomena, including speech rate and pausing, were used to analyze vocal emotional expressions (c.f. Banse & Scherer 1996). Overall, it can be ensured that the focus is increasingly on the investigation of the same parameters. Kienast (2002, p. 48) criticized e.g. the lack of parameters representing the articulatory behaviour of emotional speech. Paeschke (2004) complained about the small emphasis on stress and rhythm.

The synoptical Tab. 8.1 summarizes the changes of acoustic correlates for selected emotions. These results are based on studies that have already taken place and are listed for the emotional states that are of interest for the investigation presented here.

<sup>3</sup> A detailed description of the individual methods can be found in Martin (1990) and Westermann et al. (1996).

Acoustic parameters	Emotion					
	Joy	Sadness	Anger	Rage	Fear	Boredom
<b>F0</b>						
Perturbation	>	>		>	>	
Mean	>	<	<<	>	>	<
Range	>	<	<<	>>	>	
Variability	>	<	<<	>>	>	
Contour	>	<	<			
<b>Formants</b>						
F1 mean	<	>	>	>	>	<
F2 mean		<	<	<	<<	
<b>Intensity</b>						
Mean	>	<<	>	>>	>	<
Range	>	<	>	>		
Variability	>	<		>		
High-frequency energy	>	<	>	>>	>	<
Speech rate	>	<	<<	>	<<	<
Articulation rate	>	<	<<		>	

**Table 8.1:** Changes of acoustic parameters summarized by Banse & Scherer (1996), Murray & Arnott (1993), Scherer (1986) as well as Williams & Stevens (1972). *Note:* > increase, < decrease, >> very sharp increase, << very sharp decrease, <> contrary results.

## 8.6 Method

Spontaneous natural emotions rarely occur in public. On this account the majority of researchers work with simulated emotions with the aid of acting portrayals.

In order to correct the deficiency of spontaneous emotional material the aim of the study is to build a natural emotional database. By means of this data statements will be drawn about the question which changes occur in the acoustic parameters between neutral and emotional utterances. This means, which acoustic correlates distinguish between neutral and anxious emotional expression?

The literature of vocal emotions provides some contradictory findings regarding the acoustic parameters (see Tab. 8.1). However, as the fundamental frequency and the intensity are easy to examine, these acoustic parameters are present in almost all investigations and yield consistent results. Thus, the focus of the current study lies on the fundamental frequency and the intensity parameters.

### 8.6.1 Hypotheses

1. A variety of emotional states exhibits an increase resp. a decrease in the different parameters based on fundamental frequency as opposed to neutral emotional state.
  - a) *Joy* is shown on the basis of a high arousal: an increased mean value as well as an increased standard deviation (sd).
  - b) *Surprise* is marked by an increased mean and an increased dispersion.
  - c) *Sadness* shows a lowering of the parameters mean fundamental frequency and f0-variation.
  - d) *Anger* is characterized by an enhancement in the mean f0 as well as in the standard deviation.

2. Intensity increases resp. decreases for different emotions in comparison to neutral utterances.

- a) *Joy* shows an increase of mean intensity as well as an increase of standard deviation (sd).
- b) *Surprise* is marked by an increased mean and an increased dispersion.
- c) *Sadness* shows a lowering of the parameters mean intensity and variation.
- d) *Anger* is characterized through an enhancement of the mean intensity as well as the standard deviation.

3. The fundamental frequency as well as the intensity increases/decreases along several levels of the emotions joy, anger and sadness in the following way:

- a)  $F0_{Joy1} < F0_{Joy2}$
- b)  $F0_{Anger1} < F0_{Anger2} < F0_{Anger3}$
- c)  $F0_{Sadness1} > F0_{Sadness2} > F0_{Sadness3}$

### 8.6.2 Speech material

In order to obtain speech material containing weak as well as strong emotions, the domain of dialogues between welfare recipients and their advisors were chosen with the main topics of employment and payment cutback. The institution ‘*Grundsicherung für Arbeitsuchende des Landkreises Spree-Neisse*’ (engl. “*Basic security benefits for jobseekers in the Spree-Neisse district*”) in *Spremberg* declared their support for scientific purpose only if the speech data was processed anonymously.

### 8.6.3 Speakers

During one month, 91 speech recordings of 48 female and 45 male volunteers were collected. The participants ranged in age from 16 to 57 years (average age 38 years). With a few exceptions (two females from Russia and two males from Karlsruhe), the place of the applicable school attendance was located in the proximity of the recording region. Hence, a similar dialect could be guaranteed.

### 8.6.4 Recording procedure

A portable digital audio recorder (*Fostex FR-2LE*) with a sampling rate of 48 kHz and 24 Bit linear amplitude quantization was used to record two low-noise large-diaphragm condenser microphones with cardioid directional characteristics (*Røde NT1-A*), which were hidden by specially arranged covers. This hardware guaranteed a dynamic range of 107 dB. Each microphone was oriented to the respective speaker as shown in Fig. 8.2.

Head-worn microphones were intentionally dismissed as they would produce an only slowly fading awareness of the recording situation and consequently bias the naturalness of the dialogue.





**Fig. 8.2:** Experimental situation: Two microphones are directed towards each speaker.

## 8.7 Perception experiment

The aim of the perception experiment was to obtain as reliable emotion labels of speech stimuli as possible.

### 8.7.1 Selection of emotional speech material

With regard to their vocal emotionality, 89 recorded audio samples (average duration: 16 minutes) were analysed with *ProFIS* (= PROMpts For Information Systems, also called *PHD-system*, Tillmann & Pfitzinger 2004). Thereof, 205 segments, which had an emotional tinge, were selected. In addition, one neutral manner of expression of approximately equal length of every speaker was chosen as reference. Overall, 283 (including 78 neutral utterances) segments arose which differ from each other in length and content. In the perception test only the 205 emotional utterances were presented. The whole stimulus set is based on personal judgment of the author.

In order to present to the participants a choice of emotion designations which reflect the content of the speech stimuli, and also to enable an assignment, the first step was to define specific emotion classes. The emotions *joy*, *surprise*, *sadness* and *anger* were selected. Since the emotions *fear*, *disgust* and *apathy* only occurred rarely, they could not be considered in the analysis and were ignored a priori.

joy		anger		sadness		surprise	
f1	content	w1	irritated	t1	disappointed	u	amazed
f2	amused	w2	upset	t2	desperate		
		w3	furious	t3	sad at heart		

**Table 8.2:** Chosen terminology for the nine different emotion categories with the associated label. The numbers correspond to the degree of intensity (1 = less marked, 2 = medium, 3 = much more pronounced).

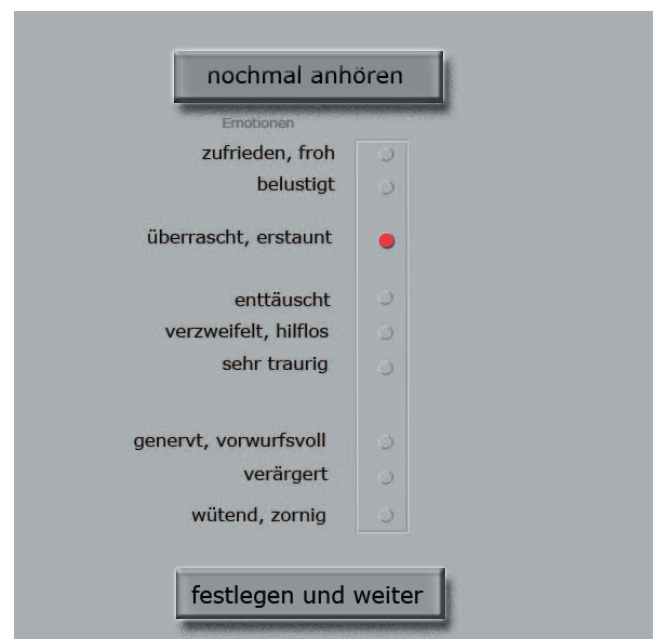
During the manual selection process of the speech segments, it became obvious that different degrees of emotion intensity were found within a class. Therefore, an additional subclassification was made in the respective emotion categories. In this way a total of nine distinct categories were created, excluded the neutral mode of speaking (Tab. 8.2). For simple processing each of them achieved a short label (e.g. f1 = joy with low intensity). Furthermore, commonly used emotion words were determined which clearly and unambiguously denoted the emotions. Tab. 8.2 shows a summary of the selected emotion classes with their associated representatives and the assigned short labels.

### 8.7.2 Participants

Students of phonetics and other fields of the Christian-Albrechts-University of Kiel were recruited. 14 of the 23 participants were female and nine were male. Their mean age was 28.6 years with a range from 22 to 31 years. All were native speakers of German.

### 8.7.3 Instructions

Before the actual start of the experiment, each participant got the same instruction: their task would be to sort 205 short utterances into nine different emotion categories. There was no fixed time limit. However, if the participants reacted too slow, they were instructed to reply more spontaneously. Furthermore, they got the possibility to listen to the utterances several times.



**Fig. 8.3:** Computer-controlled graphical user interface of the forced choice perception test.

Emotion	Label	Total number	Female speaker	Male speaker
joy	f1	5	4	1
	f2	30	26	4
surprise	u	17	14	3
sadness	t1	12	7	5
	t2	23	22	1
	t3	22	19	3
anger	w1	52	39	13
	w2	36	25	11
	w3	8	2	6
<b>Total</b>		205	158	47

**Table 8.3:** Number of stimuli in an emotion category as judged by the listeners.

### 8.7.4 Perception test procedure

A graphical user interface (GUI) was developed in *Matlab* which enabled the listener a rapid emotion assignment by using one mouse click. The test was designed as a forced choice experiment. Fig. 8.3 shows the emotion words which were arranged into four groups on the GUI. The stimuli were presented in a randomized order through headphones.

Depending on the listener, the test time varied from 25 to 40 minutes. A following survey of the participants showed that in certain cases a definite distinction of the stimuli regarding the emotional characteristic was difficult to conduct. The participants declared that “pure” emotions could be assigned only rarely, rather a *multi-emotional* form (mixed form) appeared. Moreover, the majority explained that some of the stimuli were too short to provide unambiguous indications of the emotions. Some participants claimed that the speakers’ dialectal background aggravated the decisions.

### 8.7.5 Results and revision of the short labels

In total 4715 listener judgments (205 stimuli x 23 subjects) were submitted. For each of the 205 stimuli, the emotional category chosen by the majority of the listeners was determined and overwrote the author’s short label.

Some stimuli received two emotion categories with the same number of ratings. In these cases the category which was closest to the author’s short label was chosen to be the revised short name.

Take, for example, the case that the author’s short label was *t1*, though the listeners judged *t2* and *t3* equally frequently. Then, the short label was changed to *t2*. Tab. 8.3 illustrates the results which arise from the assessment of the listener group.

## 8.8 Acoustic analysis

### 8.8.1 Fundamental frequency measurement

By comparing the different *f0* detection algorithms implemented in *ProFIS*, it turned out that *get\_f0* (originally introduced by Talkin 1995) worked best. In some cases a manual

adjustment of the algorithm parameters had to be made (e.g. octave errors or incorrect voiced/voiceless decision).

After calculating *f0*, zero values were removed to avoid a distortion of the mean value estimation. Although some utterances showed strong *f0*-variations, only 20% of the lower as well as of the upper values were processed further. This procedure guaranteed to ignore extreme outliers without significantly influencing mean values. Subsequently, the mean scores of the lower and the upper quintile were calculated. Accordingly, the lower mean represented the *f0*-minimum (*F0min*) and the upper mean the *f0*-maximum (*F0max*). Incidentally, the percentage value can be chosen within a wide range of e.g. 10% to 30% as it has only little impact.

Furthermore, all frequency values were converted into semitones. The conversion of the measured *f0* in Hertz (Hz) into semitones (*st*) ensued through the following formula<sup>4</sup>:

$$F0_{st} = \frac{\log(F0/F0_{bez})}{\log(2^{\frac{1}{12}})} \quad (8.1)$$

### 8.8.2 Amplitude measurement

In order to estimate the speech amplitude (or intensity) of an utterance, the root mean square amplitude was chosen. A Kaiser window with 50 ms duration was used to extract amplitude contours. Since some parts of the chosen material contain short speech pauses or other low amplitude values at the beginning and end of each utterance, only 50% of the loudest frames were subjected to averaging. This procedure guaranteed that amplitude values of e.g. speech pauses were removed from mean value estimation.

## 8.9 Results

The extracted *f0* and amplitude values were submitted to a statistical analysis. To test for statistical significance, a one-way analysis of variance separated into male and female speakers was applied.

### 8.9.1 Fundamental frequency

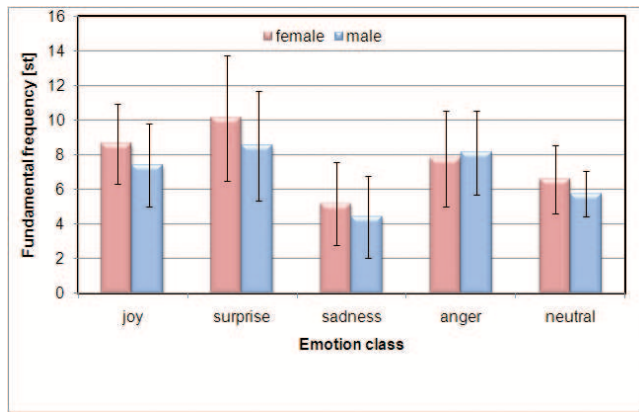
#### Mean and standard deviation

Fig. 8.4 shows the results separated into male and female speakers. On the x-axis the emotion classes are located, the y-axis shows fundamental frequency in semitones (*st*).

In either instance a highly significant effect was revealed for the factor emotion class (female:  $F = 16,373$ ;  $df = 4$ ;  $p < .001$ , male:  $F = 8,317$ ;  $df = 4$ ;  $p < .001$ ). The following post-hoc-comparisons show that sadness as the only emotion exhibits significant differences for both sexes to all other emotion classes. The distinction from neutral speech is marginally significant for women, results for men are not significant. While the female speakers show marked differences between the neutral speaking mode and joy as well as surprise, the men expose a significant difference between

<sup>4</sup> The reference level amounted 50 Hz.





**Fig. 8.4:** Average fundamental frequency and standard deviation of the emotion classes and the neutral speaking mode, separated according to gender.

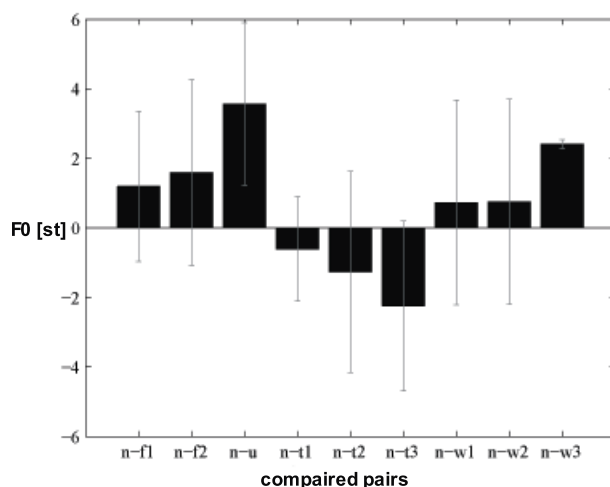
anger and neutral speech. Moreover, the female speakers exhibit a significant contrast between surprise and anger. Among the men a significant difference between joy and anger was found.

### Differences

Besides the calculation of arithmetic mean and standard deviation in a further step the differences between neutral and emotional speech were formed. For each speaker neutral utterances were paired with emotionally produced speech. A total of 112 utterances were distributed among nine pairs. Fig. 8.5 shows the differences ( $\Delta$ ) of the neutral speech and the emotional mode of expression of all speakers.

The statistical evaluation showed a highly significant effect for the factor emotion ( $p < .001$ ). A post-hoc test demonstrated the following results for the individual comparison:

- Joy in the highest level of intensity ( $f_2$ ) differs significantly from the emotion  $t_3$  (sad at heart).



**Fig. 8.5:** Difference of mean and standard deviation between neutral and emotional speech of all speakers.

- The contrast between neutral speech and surprise differs significantly from the changes of neutral to all sad utterances. Further significant values occurred for the lowest level of anger.
- $t_3$  shows significant differences from the highest characteristic of anger.
- $f_1$ ,  $w_2$  and  $w_3$  show no significant differences with regard to all other emotions.

Summarizing the results in terms of the differences between neutral and emotional speech it can be said that sadness in each of its characteristics shows significant differences to joy, surprise and anger. Between the intensity levels of an emotion class no significant effects were observed.

By means of a subsequent multiple analysis of variance the influence of the sex was examined. The results demonstrated that emotion also has a significant influence to the fundamental frequency in the multiple design ( $p = .003$ ).

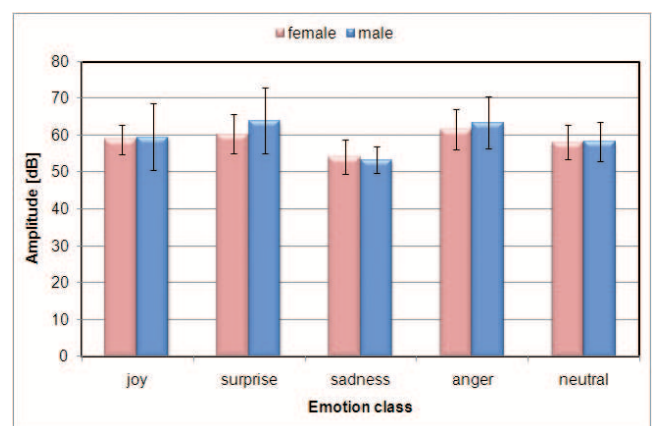
However, for sex no significant main effect was found ( $p = .969$ ), which indicates that investigating the differences instead of absolute values perfectly normalizes the gender effect. This is additionally evidenced by the fact that the interaction between the two factors (emotion\*sex) shows no effect ( $p = .786$ ).

## 8.9.2 Amplitude

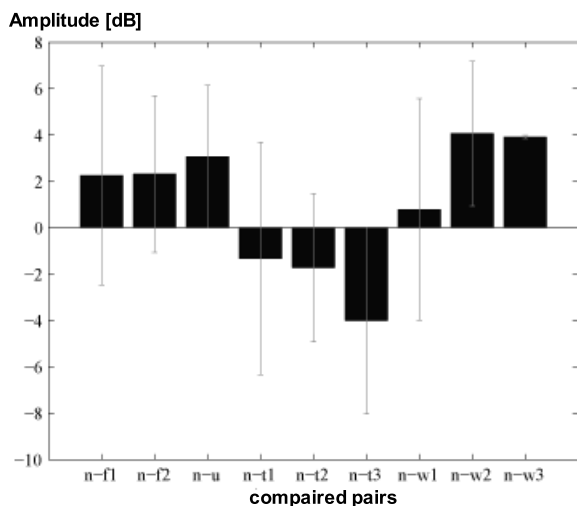
### Mean and standard deviation

As it was the case with the fundamental frequency, the mean amplitude values as well as the standard deviations were calculated for the four emotion classes. In Fig. 8.6 the results of the amplitude are illustrated. The x-axis shows the emotion classes, the y-axis the intensity values in dB.

For the factor emotion the statistical evaluation revealed a highly significant result for the male and female speakers (male  $p = .001$ , female  $p < .001$ ). Afterwards a paired comparison showed significant results for the emotion sadness. Besides the highly significant differences between all



**Fig. 8.6:** Average speech signal amplitude and standard deviation of the emotion classes and the neutral speaking mode, separated according to gender.



**Fig. 8.7:** Mean-difference and the variation between neutral and emotional speech.

three remaining emotion classes, the female speakers also demonstrated clear differences between sadness and neutral speech.

The male speakers exhibited significant differences only between sadness and anger. Except the significant differences of the emotion anger from neutral speech (for both genders) there were no other clear differences.

### Differences

The amplitude differences between neutral utterances and the respective emotional sensation are illustrated in Fig. 8.7.

The statistical analysis revealed a significant result for the factor emotion ( $p < .001$ ). In the post-hoc-comparisons the findings were the following:

- The distinctions between neutral speech and both intensity levels of the emotion joy ( $f1$ ,  $f2$ ) differ significantly from the change of neutral speech to the strongest form of sadness ( $t3$ ).
- Surprise also shows significant differences to  $t3$ .
- All intensity levels of sadness ( $t1-t3$ ) exhibit significant deviations to anger  $w2$ ;  $t3$  also shows differences to  $w1$ .
- The strongest form of anger demonstrates, as the only emotion, no significant effects in contrast to the other intensity levels.

The present results show that amplitude differences of sadness in its highest intensity level, significantly differ from all other emotions. Between the intensity levels within any class there are no significant differences.

In order to analyze the effect of sex, a multiple analysis of variance was conducted. A significant influence of emotion to amplitude differences was obvious ( $p = .002$ ). The factor sex does not show any significant effect ( $p = .463$ ). The interaction between the two factors (emotion\*sex) is also not

significant ( $p = .214$ ). Both indicate that amplitude differences, similar to  $f0$  differences, rather than absolute values also normalize the gender effect.

## 8.10 Discussion

The aim of this investigation was to determine whether and how acoustic parameters of spontaneous emotional utterances differ from neutral speech. The presented results of the acoustic analysis suggest that vocal indicators which distinguish emotional from neutral speech signals absolutely exist. Similar findings are obtained for emotion categories that can be compared with those from earlier studies (Banse & Scherer 1996; Williams & Stevens 1972). For anger contradictory statements exist with regard to  $f0$ : For instance Murray & Arnott (1995) determined a highly increased  $f0$ , whereas Scherer (1986) found both upper and lower  $f0$  in angry speech. The results of the current study agree with those of Murray & Arnott (1995).

However, it must be remarked that the acoustic alterations did not turn out in the extent as it has been postulated in other empirical studies (e.g. Scherer et al. 1991), thus mainly statements in form of only a trend can be formulated. This is due to the spontaneous nature of the speech stimuli. In addition to the imperfect recording quality, the small number of emotional utterances was the main problem. Despite the huge quantity of spontaneous material, only a slight number of emotional utterances could be extracted. This can be attributed to the following factors:

- In public areas neutral conversations occur more frequently because of learned rules of conduct. The interactive situations are restricted to resolve any given case respectively the submission of required documents. Citizens rarely give in to their feelings. And even when emotional expressions are uttered, the intensity is in the lower range.
- Acoustic problems such as background noises are typical for recording outside of a recording studio environment. Due to the disturbing conversations outside the room or the continuous operation of the computer keyboard as well as telephone bells, some emotional segments could not be selected.

The study was unilateral because of the greater emphasis on females. Although an equal number of male and female speakers was recorded, only 47 of the 205 emotional utterances were produced by males (which is remarkable). Thus, the statistical analysis of the female speakers more often revealed significant effects than that of the men. Moreover, the emotions joy and sadness were sparsely distributed.

During inspection of the spontaneous speech stimuli the need to split up the emotion qualities within a class became obvious. Because the vast majority of vocal emotion research works with simulated performances, they typically refer to coarse emotion classes (Wallbott & Scherer 1986) and seldomly to fine gradation (Banse & Scherer 1996) formed several pairs, which have the same quality, yet dif-

ferent intensity, e.g. *hot anger vs. cold anger*). In the light of the authentic emotions this generalizing approach seems to be incorrect. Not only do we have the ability to produce different intensities of one specific emotion, but also do we perceive a wide range from subtle nuances to strong emotional expressions. Although there were no significant differences between the intensity levels within the emotion classes in the present study, the results demonstrate the existence of various acoustic profiles for different levels. Thus, e.g. f0 rises when producing anger with an increasing level of arousal for female speakers (from 7.5 st to 8.0 st and on to 10.5 st) and for male speakers (from 7.1 st to 8.7 st and on to 9.5 st).

## 8.11 Outlook

A further evaluation of the present study would be feasible in several aspects. On the one hand it would be reasonable to expand the spontaneous data base in order to yield a larger number of authentic emotional expressions. In addition to the extension of the material a higher amount of various emotional states could be examined. Consequently, the research of secondary emotions might provide valuable informations about similarities and differences between related emotions.

Furthermore, the acoustic analysis and perception test were impeded by background noise and reverberation. As records of authentic emotions in soundproof areas are barely possible, the first step should be to take precautions in order to reduce the reverb. The options are, for example, wearing a headset, whereby the distance between sound source and microphone can be kept constant. A better, but expensive procedure is the redesign of the recording rooms (e.g. by the use of partitions or wall coverings).

In connection with the perception test it became clear that the dialect has effects on the classification of emotional expressions. It is quite evident from the judgment that the selected dialect (*Niederlausitzer Mundart*) is experienced as “harsh” or “impudent” (even in situations in which the speakers produced joy).<sup>5</sup> The question that arises here is: are the emotional expressions influenced by the place of origin of the speakers? Do the inhabitants in the eastern part of Germany, for instance, use other patterns in order to produce a specific emotional mode of expression than people in northern Germany? In other words, do speakers from different places produce joy for example in another manner? Or is it just the selection of the recording area that produces predominantly negative emotional expressions?

To find out if a dialectal impact on the recognition of emotions exists or if solely the selected situation determines the appearance of negative emotion, it would be (1) necessary to record emotional speech material from a different situation at the same place. The data necessarily has to be from a situation in which mainly positive emotions appear (e.g. a birthday party). (2) Similar emotional utterances might be collected from another place in Germany to verify the dialectal influence.

<sup>5</sup> None of the judges was from the recording location.

## Bibliography

- Amir, N., Mixdorff, H., Amir, O., Rochman, D., Diamond, G. M., Pfitzinger, H. R., Levi-Isserlish, T. & Abramson, S. (2010). Unresolved anger: Prosodic analysis and classification of speech from a therapeutic setting. In: *Proc. of the 5th Int. Conf. on Speech Prosody*. Chicago.
- Arnold, M. (1960). *Emotion and Personality*, vol. 1. New York: Columbia University Express.
- Banse, R. & Scherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology* 70(3), 614–636.
- Bösel, J. & Pfitzinger, H. R. (2010). Kulturübergreifende Unterschiede bei Akustik und Wahrnehmung authentischer Emotionen. *Arbeitsberichte (AIPUK)* 38, Inst. für Phonetik und digitale Sprachverarbeitung, Univ. Kiel, 63–76.
- Cannon, W. (1927). The James-Lange theory of emotion: A critical examination and an alternative theory. *American Journal of Psychology* 39(1), 106–124.
- Darwin, C. (1965). *The expression of emotions in man and animals*. Chicago: University of Chicago Press. (Original work published in 1872).
- Davidson, R. (1994). On emotion, mood and related affective constructs. In: Ekman, P. & Davidson, R. (Eds.), *The nature of emotion: fundamental questions*. New York: Oxford University Express, 51–55.
- Douglas-Cowie, E., Cowie, R. & Schröder, M. (2000). A new emotion database: considerations, sources and scope. In: *Proceedings of the ISCA Workshop on Speech and Emotion*. Newcastle, Northern Ireland.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion* 6(3/4), 169–200.
- Ekman, P. (1994). All emotions are basic. In: Ekman, P. & Davidson, R. J. (Eds.), *The nature of emotion: fundamental questions*. New York: Oxford University Express, 15–19.
- Ellsworth, P. & Scherer, K. (2003). Appraisal processes in emotion. In: Davidson, R., Scherer, K. R. & Goldsmith, H. (Eds.), *Handbook of affective sciences*. Oxford, New York: University Press, Ch. 29, 572–595.
- Ewert, O. (1983). Ergebnisse und Probleme der Emotionsforschung. In: Thomae, H. (Ed.), *Enzyklopädie der Psychologie. Motivation und Emotion: Theorien und Formen der Motivation*. Göttingen: Hogrefe, C IV 1, Ch. 7, 389–452.
- Fairbanks, G. & Pronovost, W. (1939). An experimental study of the pitch characteristics of the voice during the expression of emotion. *Speech Monographs* 6, 87–104.
- Fehr, B. & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General* 113(3), 464–486.
- Frijda, N. (1993). Moods, emotion episodes and emotions. In: Lewis, M. & Haviland, J. (Eds.), *Handbook of Emotions*. New York: Guildford Press, 381–403.
- James, W. (1884). What is an emotion? *Mind* 9, 188–205.
- Juslin, P. & Scherer, K. (2005). Vocal expression of affect. In: Harrigan, J., Rosenthal, R. & Scherer, K. (Eds.), *The new handbook of methods in nonverbal behavior research*. New York: Oxford University Express, Ch. 3, 65–135.

- Kienast, M. (2002). *Phonetische Veränderungen in emotionaler Sprechweise*. Ph. D. thesis, TU Berlin.
- Lazarus, R. S. (1991). *Emotion and adaption*. New York, Oxford: University Express.
- Martin, M. (1990). On the induction of mood. *Clinical Psychology Review* 10, 669–697.
- Merten, J. (2003). *Einführung in die Emotionspsychologie*. Stuttgart: W. Kohlhammer.
- Meyer, W., Reizenzein, R. & Schützwohl, A. (2001). *Einführung in die Emotionspsychologie. Die Emotionstheorien von Watson, James und Schachter* (2. ed.). Band I. Bern: Hans Huber.
- Meyer, W., Schützwohl, A. & Reizenzein, R. (2003). *Einführung in die Emotionspsychologie. Evolutionspsychologische Emotionstheorien* (3. ed.). Band II. Bern: Hans Huber.
- Morris, W. (1989). *Mood: The frame of mind*. New York: Springer.
- Murray, I. R. & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *Journal of the Acoustic Society of America* 93(2), 1097–1108.
- Murray, I. R. & Arnott, J. L. (1995). Implementation and testing of a system for producing emotion-by-rule in synthetic speech. *Speech Communication* 16, 369–390.
- Niedenthal, P. M., Krauth-Gruber, S. & Ric, F. (2006). *Psychology of emotion. Interpersonal, experiential, and cognitive approaches*. New York: Psychology Press.
- Ortony, A. & Turner, T. T. (1990). What's basic about basic emotions? *Psychological Review* 97(3), 315–331.
- Otto, J., Euler, H. A. & Mandl, H. (2000). Begriffsbestimmungen. In: Otto, H. A., Euler, H. & Mandl, H. (Eds.), *Emotionspsychologie: Ein Handbuch*. Weinheim: Beltz, 11–18.
- Paeschke, A. (2004). Global trend of fundamental frequency in emotional speech. In: *Speech Prosody*. Nara, Japan, 671–674.
- Parkinson, B., Totterdell, P., Briner, R. & Reynolds, S. (2000). *Stimmungen. Struktur, Dynamik und Beeinflussungsmöglichkeiten eines psychologischen Phänomens*. Stuttgart: Klett-Cotta.
- Pfützinger, H. R. (2002). Intrinsic phone durations are speaker-specific. In: *Proc. of the 7th ICSLP*, vol. 2. Denver, 1113–1116.
- Pfützinger, H. R. & Kaernbach, C. (2008). Amplitude and amplitude variation of emotional speech. In: *Proc. of Interspeech 2008*. Brisbane, 1036–1039.
- Plutchik, R. (1984). Emotions: A general psychoevolutionary theory. In: Scherer, K. & Ekman, P. (Eds.), *Approaches to emotion*. Hillsdale, New Jersey: Erlbaum, Ch. 8, 197–219.
- Plutchik, R. (1993). Emotions and their vicissitudes: emotions and psychopathology. In: Lewis, M. & Haviland, J. M. (Eds.), *Handbook of emotions*. New York: Guilford Press, Ch. 5, 53–66.
- Reizenzein, R. (2000). Worum geht es in der Debatte um die Basisemotionen? In: Försterling, J., Stiensmeier-Pelster, T. & Silny, L. (Eds.), *Kognitive und motivationale Aspekte der Motivation*. Göttingen: Hogrefe, 205–237.
- Reizenzein, R., Meyer, W.-U. & Schützwohl, A. (2003). *Einführung in die Emotionspsychologie. Kognitive Emotionstheorien*. Band III. Bern: Hans Huber.
- Schachter, S. (1971). *Emotion, obesity, and crime*. New York: Academic Press.
- Schachter, S. & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional states. *Psychological Review* 69, 379–399.
- Scherer, K. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin* 99(2), 143–165.
- Scherer, K., Banse, R., Wallbott, H. & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion* 15(2), 123–148.
- Scherer, K. R. (1984). On the nature and function of emotion: a component process approach. In: Scherer, K. R. & Ekman, P. (Eds.), *Approaches to emotion*. Hillsdale, NJ: Erlbaum, Ch. 14, 293–317.
- Scherer, K. R., Johnstone, T. & Klasmeyer, G. (2003). Vocal expression of emotion. In: Davidson, R., Scherer, K. & Goldsmith, H. (Eds.), *Handbook of the Affective Sciences*. New York: Oxford University Press, Ch. 23, 433–456.
- Schmidt-Atzert, L. (1996). *Lehrbuch der Emotionspsychologie*. Stuttgart: Kohlhammer.
- Siemer, M. (1999). *Stimmungen, Emotionen und soziale Urteile*. Frankfurt am Main: Peter Lang.
- Sokolowski, K. (1992). Emotionsinduktion ohne Bewusstheit und Verhaltensänderung ohne Absicht. *Zeitschrift für experimentelle und angewandte Psychologie* 39, 329–344.
- Talkin, D. (1995). A robust algorithm for pitch tracking (RAPT). In: Kleijn, W. & Paliwal, K. (Eds.), *Speech coding and synthesis*. Amsterdam: Elsevier Science, 495–518.
- Tillmann, H. G. & Pfützinger, H. R. (2004). Applying the Munich Parametric High Definition (PHD) speech synthesis system to the problem of teaching Chinese tones to L1-speakers of German. In: *Proc. of the Int. Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages (TAL)*. Beijing, 185–188.
- Tischer, B. (1993). *Die vokale Kommunikation von Gefühlen*. Weinheim: Psychologie Verlags Union Beltz.
- Wallbott, H. & Scherer, K. (1986). Cues and channels in emotion recognition. *Journal of personality and social psychology* 51(4), 690–699.
- Watson, D. & Clark, L. A. (1994). Emotions, moods, traits, and temperaments: conceptual distinctions and empirical findings. In: Ekman, P. & Davidson, R. (Eds.), *The nature of emotion: fundamental questions*. New York: Oxford University Express, 89–93.
- 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.
- Wierzbicka, A. (1999). *Emotions across languages and cultures: diversity and universals*. Cambridge University Press.
- Williams, C. & Stevens, K. (1972). Emotions and speech: some acoustical correlates. *Journal of the Acoustical Society of America* 52(4(2)), 1238–1250.
- Wundt, W. M. (1896). *Grundriss der Psychologie*. Leipzig: Engelmann.