Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Social Sciences 141

Vocal Expression of Emotion

Discrete-emotions and Dimensional Accounts

BY

PETRI LAUKKA

ACTA UNIVERSITATIS UPSALIENSIS
UPPSALA 2004

Dissertation presented at Uppsala University to be publicly examined in Jacobsson Widdingsalen (Room 1022), Trädgårdsgatan 18, Uppsala, Friday, December 10, 2004 at 10:15 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract

Laukka, P. 2004. Vocal Expression of Emotion. Discrete-emotions and Dimensional Accounts. Acta Universitatis Upsaliensis. *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences* 141. 80 pp. Uppsala. ISBN 91-554-6091-7

This thesis investigated whether vocal emotion expressions are conveyed as discrete emotions or as continuous dimensions.

Study I consisted of a meta-analysis of decoding accuracy of discrete emotions (anger, fear, happiness, love-tenderness, sadness) within and across cultures. Also, the literature on acoustic characteristics of expressions was reviewed. Results suggest that vocal expressions are universally recognized and that there exist emotion-specific patterns of voice-cues for discrete emotions.

In Study II, actors vocally portrayed anger, disgust, fear, happiness, and sadness with weak and strong emotion intensity. The portrayals were decoded by listeners and acoustically analyzed with respect to 20 voice-cues (e.g., speech rate, voice intensity, fundamental frequency, spectral energy distribution). Both the intended emotion and intensity of the portrayals were accurately decoded and had an impact on voice-cues. Listeners' ratings of both emotion and intensity could be predicted from a selection of voice-cues.

In Study III, listeners rated the portrayals from Study II on emotion dimensions (activation, valence, potency, emotion intensity). All dimensions were correlated with several voice-cues. Listeners' ratings could be successfully predicted from the voice-cues for all dimensions except valence.

In Study IV, continua of morphed expressions, ranging from one emotion to another in equal steps, were created using speech synthesis. Listeners identified the emotion of each expression and discriminated between pairs of expressions. The continua were perceived as two distinct sections separated by a sudden category boundary. Also, discrimination accuracy was generally higher for pairs of stimuli falling across category boundaries than for pairs belonging to the same category. This suggests that vocal expressions are categorically perceived.

Taken together, the results suggest that a discrete-emotions approach provides the best account of vocal expression. Previous difficulties in finding emotion-specific patterns of voice-cues may be explained in terms of limitations of previous studies and the coding of the communicative process.

Keywords: speech, emotion, vocal expression, emotion dimensions, acoustic cues, categorical perception, nonverbal communication, speech synthesis, cross-cultural communication, decoding accuracy, emotion intensity, meta-analysis, discrete emotions

Petri Laukka, Department of Psychology, Box 1225, Uppsala University, SE-75142 Uppsala, Sweden

ISSN 0282-7492
ISBN 91-554-6091-7
urn:nbn:se:uu:diva-4666 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-4666)

till Erika

List of Papers

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals:

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

II Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion, 1,* 381-412.

III Laukka, P., Juslin, P. N., & Bresin, R. (in press). A dimensional approach to vocal expression of emotion. *Cognition & Emotion.*

IV Laukka, P. (2004). *Categorical perception of vocal emotion expressions.* Manuscript submitted for publication.

Reprints were made with kind permission from the American Psychological Association (Studies I & II) and Psychology Press (Study III).

Contents

1. Introduction....9
   1.1 Theories of emotion....10
      1.1.1 Discrete emotion theories....10
      1.1.2 Dimensional emotion theories....11
      1.1.3 The functions of emotion expressions....12
   1.2 Earlier studies on vocal expression of emotion....13
      1.2.1 Methods of collecting vocal expressions....14
      1.2.2 Decoding of vocal expressions....14
      1.2.3 Encoding studies of vocal expression....15
   1.3 Contribution of this thesis....18
      1.3.1 Limitations of previous studies....19
      1.3.2 Research questions....21

2. Study I....24
   2.1 Background and aims....24
   2.2 Method....24
   2.3 Results....25
      2.3.1 Meta-analysis of decoding accuracy....25
      2.3.2 Acoustic cue patterns used to express discrete emotions....26
   2.4 Conclusions....29

3. Study II....30
   3.1 Background and aims....30
   3.2 Decoding....30
      3.2.1 Method....30
      3.2.2 Results from decoding experiments....32
   3.3 Encoding....33
      3.3.1 Acoustic analysis of the speech stimuli....33
      3.3.2 Acoustic correlates of discrete emotions....34
      3.3.3 Listeners' cue utilization....37
   3.4 Conclusions....39

4. Study III....40
   4.1 Background and aims....40
   4.2 Method....40
   4.3 Results....41

4.3.1 Listening experiment 41
4.3.2 Acoustic correlates of emotion dimensions 42
4.3.3 Listeners' cue utilization..43
4.4 Conclusions 44

5. Study IV 45
5.1 Background and aims 45
5.2 Experiment 1 46
5.2.1 Synthesis of emotional speech 46
5.2.2 Method...47
5.2.3 Results 48
5.3 Experiment 2 50
5.3.1 Method...50
5.3.2 Results 51
5.4 Conclusions 54

6. General discussion 55
6.1 Main findings 55
6.2 Explaining inconsistencies in code usage 57
6.3 Limitations and methodological issues 59
6.4 Implications and future research 61
6.5 Possible applications 63
6.6 Concluding remarks 64

7. Acknowledgments...65

8. References..66

Abbreviations

ANOVA	Analysis of variance
ANS	Autonomic nervous system
CI	Confidence intervals
CP	Categorical perception
F_0	Fundamental frequency
F1	Formant 1
F2	Formant 2
F3	Formant 3
LTAS	Long-term average spectrum
SEC	Stimulus evaluation check
TD-PSOLA	Time-Domain Pitch-Synchronous OverLap-and-Add

1. Introduction

Take some time to think about the last time you came to realize that a person close to you was feeling an emotion. What information did you use to come to that conclusion? Perhaps it was something that the person said. Maybe it was not so much *what* the person said, but rather *how* it was said.

Nonverbal signals, such as facial expressions and tone of voice, are commonly assumed to be particularly suited for communicating emotions (Wallbott & Scherer, 1986). Among the nonverbal cues people use to infer the emotions of others in everyday life, voice cues (e.g., pitch, loudness, and speech rate) are among the most frequently reported (Averill, 1982; Planalp, DeFrancisco, & Rutherford, 1996; see also, Scherer & Ceschi, 2000). This thesis investigated how emotions are expressed nonverbally via the voice; a phenomenon commonly referred to as *vocal expression of emotion*.

The general aim was to investigate if vocal expressions are conveyed as discrete emotions (e.g., anger, fear, happiness, sadness), or as broad emotion dimensions like activation and valence. The following specific questions are addressed: (a) Are vocal emotion expressions of discrete emotions recognized cross-culturally? (b) Are there distinct patterns of voice cues that correspond to discrete emotions? (c) Are vocal expressions perceived as discrete emotions, or as continuous emotion dimensions?

Four studies provide the basis of the thesis. Study I consists of a literature review, and the other studies are empirical investigations (Studies II, III, & IV). A brief introduction to relevant research together with the rationale behind the specific research questions are given in Chapter 1, and then the studies themselves are described (Chapters 2-5). The thesis concludes with a general discussion of the results including a theoretical explanation of the findings (Chapter 6).

1.1 Theories of emotion

> "Everyone knows what an emotion is, until asked to give a definition. Then, it seems, no one knows" (Fehr & Russell, 1984, p. 464).

Emotion is a notoriously hard concept to define, and there are no generally agreed upon criteria for what should count as an emotion and what should not. From an evolutionary perspective (e.g., Buss, 1995), the key to understanding emotions is to look at what *functions* they serve (Keltner & Gross, 1999; Cosmides & Tooby, 2000). From this perspective, emotions have evolved to deal with goal-relevant changes in our environment and may be described as relatively brief and intense reactions to these changes. Most researchers would agree that emotions consist of several components: cognitive appraisal, subjective feeling, physiological arousal, expression, action tendency, and regulation (Oatley & Jenkins, 1996; K. R. Scherer, 2000). However, researchers disagree about how emotions should be conceptualized: as discrete categories (Ekman, 1992), dimensions (Russell, 1980), prototypes (Shaver, Schwartz, Kirson, & O'Connor, 1987), or component processes (Scherer, 2001).

The two research traditions that have most strongly influenced research on vocal expression (and emotion in general) are briefly described below, namely discrete and dimensional emotion theories. Special focus is given to the physiological component of emotion, since it is commonly assumed that physiological variables have an important influence on the acoustic characteristics of vocal expressions. After that, the specific functions of emotion expressions are considered.

1.1.1 Discrete emotion theories

The evolutionary view is closely related to discrete emotion theories according to which each discrete emotion is thought to represent a unique person-environment interaction with its own adaptational significance for the individual. Each discrete emotion is also thought to have its own unique pattern of cognitive appraisal, physiological activity, action tendency, and expression (e.g., Darwin, 1872/1998; Ekman, 1992; Izard, 1992; Tomkins, 1962). According to several discrete emotion theories there exist a limited number of "basic" emotions that have evolved to deal with particularly pertinent life problems such as competition (anger), danger (fear), cooperation (happiness), or loss (sadness; see Power & Dalgleish, 1997, pp. 86-99).

It is often assumed that environmental demands on behavior are reflected in distinct physiological patterns:

> "Behaviors such as withdrawal, expulsion, fighting, fleeing, and nurturing each make different physiological demands. A most important function of emotion is to create the optimal physiological milieu to support the particular behavior that is called forth" (Levenson, 1994, p. 124).

This process of creating the optimal milieu involves the central, somatic, and autonomic nervous systems. Evidence for physiological differentiation of discrete emotions comes from findings of distinct brain substrates associated with discrete emotions (e.g., Murphy, Nimmo-Smith, & Lawrence, 2003; Phan, Wager, Taylor, & Liberzon, 2002). At present, fear is the most well understood emotion in terms of neural mechanisms (LeDoux, 2000; Öhman & Mineka, 2001). The evidence is weaker regarding emotion specific autonomic nervous system (ANS) activity (Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000), but there is substantial empirical support for autonomic discriminability of at least *some* emotions (e.g., Christie & Friedman, 2004; Levenson, 1992; Levenson, Ekman, & Friesen, 1990; Nyklíček, Thayer, & van Doornen, 1997; Stemmler, 1989; see also Herrald & Tomaka, 2002).

The strongest support for discrete emotions has traditionally come from studies of communication of emotions, which suggest that facial expressions are universally expressed and recognized (Ekman, 1992, 1994).

1.1.2 Dimensional emotion theories

The dimensional approach to emotion has largely concentrated on one component of emotion, the subjective feeling state, and focuses on identifying emotions based on their placement on a small number of underlying dimensions. Wundt (1912/1924) suggested that three dimensions (i.e., pleasure-displeasure, strain-relaxation, excitement-calmness) could account for all differences among emotional states. Schlosberg (1941) proposed that the underlying structure of emotional experience can be characterized as an ordering of emotional states on the circumference of a circle. This model, now commonly referred to as the "circumplex" model of emotion, has proved highly influential. Today, most proponents of the circumplex model agree that two orthogonal dimensions underlie the circular ordering. Many authors have postulated an *activation* dimension, and an evaluation, or *valence*, dimension (Larsen & Diener, 1992; Russell, 1980). The valence dimension relates to how well one is doing at the level of subjective experience, and ranges from displeasure to pleasure. The activation dimension, in turn, relates to a subjective sense of mobilization or energy, and ranges from sleep to frenetic excitement (Russell & Feldman Barrett, 1999).[1]

[1] An alternative dimensional model instead posits two orthogonal dimensions, positive affect and negative affect, that are rotated 45° relative to the activation and valence dimensions (e.g., Watson & Tellegen, 1985). However, the circumplex model consisting of activation and valence is more commonly used in studies on vocal expression.

Regarding physiological effects, a two-dimensional model of emotion has received support from studies of ANS activation resulting from various emotion induction procedures (Cacioppo et al., 2000; Lang, Bradley, & Cuthbert, 1998). Functionally, the two dimensions of activation and valence are believed to be relevant for approach and avoidance/withdrawal behavior, respectively. In other words, positive and negative affects are posited to convey information about whether the behavior engaged in is going well or poorly (e.g., Carver, 2001).

The use of only two dimensions has been criticized on the grounds that this does not allow discrimination of certain emotional states (Larsen & Diener, 1992; Lazarus, 1991). Fear and anger, for example, are both unpleasant and highly active. In order to be able to capture qualitative differences among different emotional states, more dimensions are needed. A third dimension that is frequently mentioned in the literature is *potency* (e.g., Russell & Mehrabian, 1977; Osgood, Suci, & Tannenbaum, 1957). Potency may be seen as a dimension that involves cognitive appraisal of an individual's coping potential, or power, in a particular situation; and it has been variably referred to as potency, dominance, power, or control (Lazarus & Smith, 1988). It has been suggested that this is an important dimension for the differentiation of negative emotions (Smith & Ellsworth, 1985).

Another dimension that is generally recognized but poorly understood is *emotion intensity*. Emotions vary not only in quality but also in quantity; a person can be just a little angry or very angry. The relative intensity of emotions is of great importance for the behavioral and physiological responses of an emotion (e.g., Brehm, 1999; Frijda, Ortony, Sonnemans, & Clore, 1992; Sonnemans & Frijda, 1994).

1.1.3 The functions of emotion expressions

The communication of emotions is often viewed as crucial to social relationships and survival (Buck, 1984), and many of the most important adaptive problems faced by our ancestors are assumed to have been social by nature (e.g., Buss & Kenrick, 1998). Emotion expressions can serve as incentives of social behavior through two interrelated mechanisms (Keltner & Kring, 1998). Firstly, by expressing emotions we can communicate important information to others, thereby influencing their behaviors, and the recognition of others' expressions allows us to make quick inferences about their probable behaviors (Darwin, 1872/1998; Plutchik, 1994). Secondly, expressions can regulate social behavior by evoking emotional responses in the decoder (e.g., Russell, Bachorowski, & Fernández-Dols, 2003).

Arguably, the same selective pressures that shaped the development of the emotions in the first place should also favor the development of skills for expressing and recognizing the same emotions. Thus several researchers have proposed that the production and perception of emotion expressions are organized by innate mechanisms (e.g., Buck, 1984; Ekman, 1992; Lazarus, 1991, Tomkins, 1962). Support for this notion comes from, for instance, more or less intact facial and vocal expressions of emotion in children born deaf and blind (Eibl-Eibesfeldt, 1973; Goodenough, 1932), and universality of facial emotion expressions (e.g., Elfenbein & Ambady, 2002; Keltner, Ekman, Gonzaga, & Beer, 2003).

However, as noted by several researchers, expressions are also shaped to a certain degree by salient cultural display rules and contextual factors, such as the immediate social environment (e.g., Ekman, 1972; Izard, 1977). A useful distinction can here be made between so-called *push* and *pull* effects in the determinants of emotion expressions (Scherer, 1989). Push effects involve various internal processes of the organism that are influenced by the emotional response. Pull effects, on the other hand, involve external conditions such as social norms. In any given case of emotion expression, both push and pull effects can be present and affect the resulting expression.

A consequence of the co-existence of push and pull effects is that there is no one-to-one relationship between expression and other components of emotion (e.g., subjective feeling).[2] Individuals are most likely to report an emotion, as well as theorists are most likely to claim that an emotion has occurred, to the extent that many components of emotion co-occur (e.g., cognitive appraisal, subjective feeling, physiological arousal, expression; see Ekman, 1993).

1.2 Earlier studies on vocal expression of emotion

Interest in vocal expression of emotion goes as far back in history as the Ancient Greeks at the least. Early Greek and Roman manuals on rhetoric (e.g., by Aristotle, Cicero) included several examples of how the voice could be used to express different emotions. However, historically vocal expression of emotion has not received as much attention as facial expression (Scherer, 1986). Because of this, our knowledge of how the voice conveys emotions is more limited than our knowledge of facial expressions. Recently there have been increased efforts directed at vocal expression (e.g., Cowie et al., 2001; Scherer, Johnstone, & Klasmeyer, 2003). This probably reflects growing interest in the study of emotions in general coupled with progress in speech science.

Most studies have considered vocal expression as a means to communication. Hence, fundamental issues include (a) the *content* (what is communi-

[2] However, there are several reports in support of a substantive relationship between expression and other emotion components. For instance, it has been found that expressions co-occur with the subjective experience of an emotion (e.g., Ekman, Friesen, & Ancoli, 1980; Fox & Davidson, 1988), and that expressions correspond to emotion-related appraisals (e.g., Bonnano & Keltner, 2004; Smith, 1989).

cated?), (b) the *accuracy* (how accurately is it communicated?), and (c) the *code* (how is it communicated?). The following sections review the methods that have been used to address these questions.

1.2.1 Methods of collecting vocal expressions

Most studies of vocal expression to date have used some variant of the “standard content paradigm“. That is, someone (e.g., an actor) is instructed to read some verbal material aloud, while simultaneously portraying particular emotions chosen by the investigator. The emotion portrayals are first recorded, and then evaluated in listening experiments to see whether listeners are able to decode the intended emotions. The same verbal material is used in portrayals of different emotions, and most typically has consisted of single words or short phrases. The assumption is that because the verbal material remains the same in the different portrayals, whatever effects appear in listeners’ judgments should mainly be the result of the voice cues produced by the speaker. Other common methods include the use of emotional speech from real conversations (e.g., Eldred & Price, 1958; Greasley, Sherrard, & Waterman, 2000; Huttar, 1968), induction of emotions in the speaker using various methods (e.g., Bachorowski & Owren, 1995; Bonner, 1943; Millot & Brand, 2001), and the use of speech synthesis to create emotional speech stimuli (e.g., Burkhardt, 2001; Cahn, 1990; Murray & Arnott, 1995).

Each of these methods has both advantages and drawbacks. Using actor portrayals and the standard content paradigm ensures control of the verbal material and the encoder’s intention, but raises the question about the similarity between posed and naturally occurring expressions. Using real emotional speech, on the other hand, ensures high ecological validity, but renders the control of verbal material and encoder intention more difficult. Induction methods are effective in inducing moods (e.g., Westermann, Spies, Stahl, & Hesse, 1996), but it is harder to induce intense emotional states in controlled laboratory settings. Finally, the use of speech synthesis enables one to conduct controlled experiments where different voice cues can be manipulated separately, in order to investigate diverse hypotheses about cue utilization in vocal expression. This seems a promising route, but the synthesized speech still needs to be modeled on human emotional speech, which in turn must be obtained by one of the above methods.

1.2.2 Decoding of vocal expressions

Listeners’ responses have most often been collected through forced-choice procedures, where the listener is asked to select one among several emotion labels. Another fixed-alternative method is to ask listeners to rate the stimuli on scales representing either emotion labels or emotion dimensions (e.g., Scherer, Banse, Wallbott, & Goldbeck, 1991). Free descriptions have also been used, though more sparsely (e.g., Greasley et al., 2000).

Again, these methods have both advantages and drawbacks. The use of forced-choice methodology produces an ecologically valid task, but the fixed alternatives may produce artifacts (for a hot debate of the pros and cons, see Ekman, 1994; and Russell, 1994). Some of the problems with the forced-choice method are alleviated by the use of rating scales, but the listeners responses are still being influenced by the alternatives present. There have been several suggestions as to how one can improve the validity of the fixed-choice methodology; for instance by correcting for guessing (Wagner, 1993), or including "other emotion" as an response alternative (Frank & Stennet, 2001). The use of free descriptions is the least biasing task, though one still needs to restrict the range of the listener's responses, for instance by asking them what emotion they perceive the stimuli to express. If one does not do this, there is always a risk of getting totally irrelevant answers. Free descriptions are also difficult and time-consuming to classify. It has been reported that free descriptions and the forced-choice task yield similar results, though free descriptions give more detailed information (e.g., Greasley et al., 2000).

It was early agreed that emotions can be communicated accurately through vocal expressions, a finding that is supported by common, everyday experience (e.g., Kramer, 1963). According to some recent estimates, the communication of basic emotions can reach an accuracy about four or five times higher than would be expected by chance alone (e.g., Johnstone & Scherer, 2000). However, it has also been suggested that such estimates might be inflated by methodological artifacts, and that a possibility remains that listeners primarily perceive the activation, or arousal, level of the vocalizations instead of discrete emotions (e.g., Bachorowski & Owren, 2003).

There are also some preliminary findings of cross-cultural recognition of vocal expressions. A couple of studies have shown that people are able to recognize vocal expressions from other cultures with an accuracy above chance (e.g., Albas, McCluskey, & Albas, 1976; Kramer, 1964; Scherer, Banse, & Wallbott, 2001; van Bezooijen, Otto, & Heenan, 1983). This suggests that vocal expressions may be universal. More conclusive evidence of cross-cultural recognition would lend support to discrete emotion theory.

1.2.3 Encoding studies of vocal expression

Speech production

In human speech, linguistic and nonlinguistic information is coded simultaneously in the acoustic signal, and is communicated by the same acoustic voice cues. Also, besides expressive functions, the voice also contains other types of information about the speaker; for instance the identity, age, sex, and body size of the speaker (e.g., Borden, Harris, & Raphael, 1994).

What we hear as speech is produced by the continuous movement of the speech articulators, such as the tongue, lips, and larynx. These articulators modulate airflow in such a way that speech sounds reach our ear. The source-filter model of speech production treats vocal acoustics as a linear combination of an underlying energy source and filtering effects due to resonances of the pharyngeal, oral, and nasal cavities that make up the supralaryngeal vocal tract (Fant, 1960). The harmonics of the glottal sound that correspond to the resonant frequencies of the vocal tract are amplified, and those distant from the resonant frequencies of the vocal tract lose energy. The sound which emerges at the end of the tract (i.e., the lips) has the same harmonics as the sound at the source (i.e., the glottis), but the amplitude of the harmonics has been modified, altering the quality of the sound (e.g., Ladefoged, 2000). Vocal tract resonances are called formant frequencies, or *formants*. The frequencies of the first two formants (*F1* and *F2*) largely determine vowel quality, whereas the higher formants may be speaker-dependent (Laver, 1980). For a more extensive discussion of the principles underlying speech production and associated measurements, see Borden et al. (1994) and Titze (1994).

Voice cues that can be measured from an acoustic speech signal can be broadly divided into those related to (a) *fundamental frequency (F_0)*, (b) *voice intensity*, (c) *voice quality*, and (d) *temporal aspects* of speech. The respiratory process of the lungs builds up subglottal pressure. This pressure, in combination with vocal fold adduction and tension, leads to vibration of the vocal folds. The frequency with which the vocal folds open and close across the glottis during phonation is termed F_0. This is subjectively heard as the pitch of the voice, and mainly reflects the differential innervation of the laryngeal musculature and the extent of subglottal pressure. *Voice intensity* is subjectively heard as the loudness of the voice, and is determined by respiratory and phonatory action. It reflects the effort required to produce the speech. *Voice quality* is subjectively heard as the timbre of the voice, and is largely determined by the settings of the supralaryngeal vocal tract and the phonatory mechanisms of the larynx. *Temporal aspects* of speech, finally, concern the temporal sequence of the production of sounds and silence (e.g., speech rate, pausing).

Acoustic characteristics of vocal expressions

Almost from the beginning of empirical research on vocal expressions, researchers started to acoustically analyze the emotional speech, hoping to find acoustic voice cues that signal various emotional states (e.g., Fairbanks & Hoaglin, 1941; Fairbanks & Provonost, 1939; Isserlin, 1925; Scripture, 1921; Skinner, 1935). Soon enough it was discovered that it was difficult to find specific voice cues that could be used as reliable indicators of vocal expressions. Scherer's (1986) review of the literature revealed an apparent paradox: whereas listeners seem to be accurate in decoding emotions from voice cues, scientists have been unable to identify a set of cues that reliably discriminate among emotions. Inconsistent results regarding the voice cues used to encode emotions abound in the literature (Frick, 1985; Murray & Arnott, 1993). In this thesis several possible reasons for this state of affairs are proposed. Before that, however, let us look at what previous research has found.

A fundamental question is what aspects of the voice signal (i.e., what voice cues) should be measured. The most obvious answer to this question is: "As many as possible". If a larger part of the speech signal is measured, it increases the chances of finding cues that are used to convey emotion. Speech technology has developed rapidly in the last decade, but the analysis of speech signals is still quite time-consuming, makes great demands on the quality of sound recording, and is difficult to make completely automatic because of the dynamic character of speech.

The three "classic" cues are F_0, voice intensity, and speech rate. These three aspects together constitute *prosody* (i.e., the patterns of stress and intonation in speech). The results concerning the prosodic expression of emotions are fairly clear. It is well established that mean, range, and variability of F_0 rises for "active" emotions (e.g., anger, fear, happiness), and decreases for "passive" emotions (e.g., sadness). Also, voice intensity increases for anger and decreases in sadness, and the speech rate is faster for anger, fear, and happiness, than for sadness (e.g., reviews by Cowie et al., 2001; Johnstone & Scherer, 2000).

Voice quality (timbre) is not generally regarded as a component of prosody, but is an important factor in emotion expressions (e.g., Burkhardt, 2001; Gobl & Ní Chasaide, 2003). Voice quality is not easy to measure or to define (e.g., Laver, 1980), and thus fewer studies have analyzed the acoustic correlates of voice quality. One useful index of voice quality is high-frequency energy. As the proportion of high-frequency energy in the acoustic spectrum increases, the voice sounds more "sharp" and less "soft" (von Bismarck, 1974). Studies have shown that the proportion of high-frequency energy increases in anger and decreases in sadness (Banse & Scherer, 1996; Kaiser, 1962; Leinonen, Hiltunen, Linnankoski, & Laakso, 1997; Scherer et al., 1991; van Bezooijen, 1984).

The above minimal review shows that voice cues are systematically affected by different emotion expressions. However, it remains unclear whether the acoustic differentiation reflects discrete emotions or broad emotion dimensions like activation and/or valence. Also, while the results seem relatively consistent when accumulated over several studies, single studies often find conflicting results.

Scherer's (1986) theory of vocal expression

There are few theories of vocal expression of emotion. The one stringent attempt to formulate a theory of vocal expression was made by Scherer

(1986). The general principle that underlies this theory is that physiological variables to a large extent determine the nature of phonation and resonance in vocal expression (see Spencer, 1857, for an early formulation of this principle). For instance, anger yields increased tension in the laryngeal musculature coupled with increased sub-glottal air pressure. This will change the production of sound at the glottis and hence change the timbre of the voice (Johnstone & Scherer, 2000). In other words, depending on the specific physiological state, we may expect to find specific acoustic features in the voice.

Based on his component process theory of emotion, Scherer (1986) made detailed predictions about the patterns of vocal cues associated with different emotions. The predictions were based on the idea that emotions involve sequential cognitive appraisals, or *stimulus evaluation checks* (SEC), of stimulus features like novelty, intrinsic pleasantness, goal significance, coping potential, and norm/self compatibility (for further elaboration of appraisal dimensions, see Scherer, 2001). The outcome of each SEC is assumed to have a specific effect on the somatic nervous system, which, in turn, affects the musculature associated with voice production. In addition each SEC outcome is assumed to affect various aspects of the ANS (e.g., mucous and saliva production) in ways that strongly influence voice production. Scherer (1986) does not favor a discrete-emotions approach, although he offers predictions for acoustic cues associated with anger, disgust, fear, happiness, and sadness; "five major types of emotional states that can be expected to occur frequently in the daily life of many organisms, both animal and human" (Scherer, 1985, p. 227). These predictions are consistent with both discrete emotion theory and component process theory in suggesting that there are distinct patterns of voice cues for different emotions. Predictions that would distinguish between the theories have yet to be proposed and tested. A selection of Scherer's (1986) predictions are shown in Table 1.

1.3 Contribution of this thesis

The difficulties in finding specific patterns of voice cues for discrete emotions have frustrated researchers for years. As a consequence, some authors have concluded that voice cues only reflect the activation dimension of emotions (Pakosz, 1983; Davitz, 1964a), or a combination of activation and valence (Bachorowski, 1999). In this section, a number of limitations of previous studies are pointed out. It is argued that such limitations may account for a large part of the previous difficulties of finding patterns of voice cues that differentiate between discrete emotions. The studies of this thesis were consequently designed to remedy the problems posed by these limitations.

Table 1

Predictions for Emotion Effects on Selected Voice Cues (Adapted from Scherer, 1986)

	Emotion								
Voice cue	Irrita-tion	Rage	Disgust	Anxiety	Terror	Happi-ness	Elation	Sadness	Grief
F_0									
Mean	±		+	+	++	-	+	±	+
Variability	-	++			++	-	+	-	+
Contour	-	=		+	++	-	+	-	+
Voice intensity									
Mean	+	++	+		+	-	+	--	+
Variability		+			+	-	+	-	
Voice quality									
HF energy	++	++	+	+	++	-	±	±	++
F1 (mean)	+	+	+	+	+	-	-	+	+
F1 (bw)	--	--	--	-	--	+	±	±	--
Prec. art.	+	+	+	+	+		+	-	+
Temporal aspects									
Speech rate		+			++	-	+	-	+

Note. + = increase; - = decrease; = represents no change; ± = predictions in opposing directions. Double symbols indicate increased predicted strength of the change. HF energy = high frequency energy; F1 = first formant; bw = bandwidth; prec. art. = precision of articulation.

1.3.1 Limitations of previous studies

Individual differences

Studies have yielded evidence of considerable individual differences in both encoding and decoding accuracy (e.g., Banse & Scherer, 1996; Pakosz, 1983). Particularly, encoders differ widely in their ability to portray specific emotions. Because many researchers have not taken this problem seriously, several studies have investigated only one or two speakers. This makes it hard to control for idiosyncratic effects on voice cue usage.

Few voice cues analyzed

Many studies have analyzed only one, or a few, voice cues. By analogy with facial expressions, such a strategy could be compared to trying to describe facial expressions by only looking at the rising and lowering of the eyebrows. This may of course give valuable information, but all relevant features must be captured before an accurate understanding of how the face conveys emotions can be achieved. Similarly, in vocal expression it is not reasonable to expect differentiation of emotions on the basis of a too small number of voice cues.

There are a number of potentially useful cues that have been measured in only a few studies each. Such cues include, for instance, jitter (small-scale perturbations in F_0; Bachorowski & Owren, 1995; Lieberman, 1961; van Bezooijen, 1984), the mean frequency and bandwidth (i.e., the width of the spectral band containing significant energy) of formants (Kienast & Sendlmeier, 2000; Laukkanen, Vilkman, Alku, & Oksanen, 1997; Williams & Stevens, 1972), pauses (Fairbanks & Hoaglin, 1941; Williams & Stevens, 1972), precision of articulation (e.g., Davitz, 1964b, van Bezooijen, 1984; Kienast & Sendlmeier, 2000), micro-structural regularity (e.g., Davitz, 1964b), and F_0 contours (Fónagy & Magdics, 1963; Mozziconacci, 1998).

A couple of studies have also studied the glottal source (glottal waveform) by methods of inverse filtering (e.g., Laukkanen, Vilkman, Alku, & Oksanen, 1996; Klasmeyer & Sendlmeier, 1997), or by studying the physiology of the source directly (Gendrot, 2003; Svanfeldt, Nordstrand, Granström, & House, 2003; Winkler, 2002). The articulatory settings have also been studied directly by means of radiological methods (Fónagy, 1976). Thus, there exist several potentially important cues, which have not been frequently taken into account in studies of vocal expression. If a larger part of the acoustic voice signal is described by the acoustical measurements, this may increase the possibility of finding cue patterns that differentiate between emotions (e.g., Banse & Scherer, 1996).

Poorly defined emotional states

Many studies have not used well-defined emotion labels, but instead have chosen the labels to be investigated on an *ad hoc* basis. Most studies have used labels that more or less correspond to basic emotions. However, it could be the case that such labels do not capture the full amount of differentiation that may be possible in vocal expression.

For example, it could be that different variants of anger (like irritation and rage) can differ with respect to their acoustic patterns, even though they do belong to the same emotion family (i.e., anger). Banse & Scherer (1996) used such more fine-grained descriptions of emotions, and showed that these differences did have an impact on the patterning of voice cues.

Also, studies have not taken emotion intensity into account (but see Baum & Nowicki, 1998). Since the relative intensity of emotions is believed to have a great impact on the behavioral and physiological responses of an emotion (e.g., Brehm, 1999; Frijda et al., 1992), it would seem important to consider the intensity of emotions in studies of vocal expression. For instance, it could be that weak anger and strong anger have different effects on voice cues. Failure to specify the intensity of the emotion labels to be investigated could thus confound the findings on acoustic emotion differentiation.

Lack of a unifying theory

Much of the work on vocal expression has been atheoretical, searching to empirically determine which changes in the voice will be produced by emotions in the speaker. Such an approach entails the problem of making sense of a multitude of different, often non-replicated results (Banse & Scherer, 1996). Scherer's (1986) theory is a rare exception, but though the article is widely cited, few studies have explicitly tested the predictions made by this theory. Banse and Scherer (1996) found support for some predictions, but more tests are needed. Clearly, future studies would benefit from being based on an explicit set of hypotheses. Such scientific maturity would also facilitate the cumulation of research findings (Scherer, 2003).

Fragmented literature

The literature on vocal expression has been described as fragmented (Murray & Arnott, 1993). The investigation of vocal expressions today is a multidisciplinary enterprise that concerns researchers from areas as diverse as psychology, acoustics, linguistics, phonetics, communication science, speech science and engineering. Most researchers working on vocal expression choose to publish in their own specialized outlets, which are rarely read by researchers in other areas. For instance, engineers seldom read psychology journals and, vice versa, psychologists seldom read engineering journals. This has had the consequence that it is hard to get an overview of what is already known in the field, since even reviews on vocal expression have sampled only a limited subset of the literature.

1.3.2 Research questions

Theoretical background

In his classic treatise, *The Expression of the Emotions in Man and Animals*, Darwin (1872/1998) reviewed different modalities of expression, including the voice, and concluded that:

> "with many kinds of animals, man included, the vocal organs are efficient in the highest degree as a means of expression" (p. 88).

Following Darwin's lead, several researchers have adopted an evolutionary perspective on vocal expression (e.g., Papoušek, Jürgens, & Papoušek, 1992); a perspective that is also adopted in the present thesis (see Study I). Another starting-point is discrete emotion theory, according to which certain emotion states (e.g., anger, fear, happiness, and sadness), have evolved to help the organism deal with pertinent life problems (e.g., Ekman, 1999; Power & Dalgleish, 1997).

Building on the above premises, the following argument is proposed: (a) Emotions may be regarded as adaptive reactions to certain prototypical, goal-relevant, and recurrent life problems (situations) that are common to many living organisms; (b) an important part of what makes emotions adaptive is that they are communicated nonverbally from one organism to another, thereby transmitting important information that helps regulate social behavior; (c) the specific form that vocal emotion expressions take indirectly reflect these situations or, more specifically, the distinct physiological patterns that support the emotional behavior called forth by these situations; and (d) the voice production is influenced in differentiated ways by these physiological reactions.

Further, it has been suggested that *categorical perception* (CP) of emotion expressions is an ability that has evolved in order to enable rapid classification of states in others that can motivate behavior (e.g., Etcoff & Magee, 1992; Laukka, 2003).[3] Several studies have reported evidence of CP of facial expressions (e.g., Calder, Young, Perrett, Etcoff, & Rowland, 1996; Campanella, Quinet, Bruyer, Crommelinck, & Guerit, 2002; de Gelder, Teunisse, & Benson, 1997; Etcoff & Magee, 1992; Young et al., 1997), but no studies have yet investigated CP of vocal expressions. The question of whether vocal expressions are categorically perceived is of great conceptual relevance, since evidence of CP would be hard to reconcile with a dimensional model of perception of vocal expressions, but would fit well within a discrete emotions framework.

Based on the above reasoning, the following hypotheses were stated: If vocal expressions of discrete emotions have evolved because they have had important functions in regulating social behavior, (1) *vocal expressions of discrete emotions should be universally recognized by listeners*. Also, if listeners universally recognize vocal expressions of discrete emotions, (2) *there should exist distinct patterns of acoustic voice cues for discrete emotions*. Finally, in order to facilitate quick recognition of, and response to, vocal expressions, it is hypothesized that (3) *vocal expressions of discrete emotions should be categorically perceived*. The specific research questions considered in this thesis (detailed below) are derived from these three hypotheses, as well as from the preceding review of previous studies on vocal expression.

[3] CP occurs when continuous sensory stimulation is sorted out by the brain into discrete categories. One effect of this is that equal-sized differences between stimuli are perceived as smaller or larger depending on whether the stimuli are perceived as belonging to the same category or to different categories (Harnad, 1987; Repp, 1984).

Specific research questions

The general question studied in this thesis is: *Are vocal expressions conveyed as discrete emotions (e.g., anger, fear, happiness, sadness), or rather as broad emotion dimensions (e.g., activation, valence)?*

The following specific research questions are investigated:

- Q1: Are vocal emotion expressions of discrete emotions recognized cross-culturally?

- Q2: Are there distinct patterns of vocal cues that correspond to discrete emotions?

- Q3: Are vocal emotion expressions perceived as discrete emotions or as continuous emotion dimensions?

Q1 is investigated by means of conducting a meta-analysis of the cross-cultural literature on communication of vocal emotion (Study I).

Q2 is examined (a) by conducting an exhaustive review of the cue-utilization literature (Study I), (b) by using well-defined emotion states (i.e., by controlling for emotion intensity) and analyzing a larger than usual set of acoustic voice cues (Study II), (c) indirectly by investigating acoustic correlates of emotion dimensions (i.e., looking at whether voice cues convey more dimensions than activation and valence; Study III). Also, Scherer's (1986) predictions are compared against the results of both Studies I and II.

Q3, finally, is investigated in Study IV, using the CP paradigm.

2. Study I

2.1 Background and aims

In Study 1, we aimed to assess the current state of knowledge on vocal expression of emotion by conducting a comprehensive review of the vocal expression literature, which included research done in all relevant disciplines. Because studies of vocal expression are disseminated in so many different forums, many previous reviews have only sampled a subset of the existing research. The previous all-encompassing review was published almost 20 years ago (Scherer, 1986), and thus Study I benefits from the additional studies published since.

Firstly, we investigated if vocal expressions are cross-culturally recognized. Secondly, we wanted to see if there are specific patterns of voice cues for discrete emotions. Thirdly, we also tested Scherer's (1986) predictions about code usage against the empirical evidence.[4]

2.2 Method

A search of the literature on vocal expression of emotion was conducted using the internet-based scientific databases PsychINFO, Medline, LLBA, and Ingenta. The goal was to include all English language articles on vocal expression of emotion published in peer-reviewed journals. Additional studies located via informal sources that we were able to locate were also included (including studies reported in conference proceedings, other languages, and in unpublished doctoral dissertations). Two criteria for inclusion were used. First, only studies focusing on nonverbal aspects of speech were included. Second, only studies that investigated the communication of discrete emotions were included. 104 studies of vocal expression that met the inclusion criteria were located and included in the review. The majority of

[4] Study I also investigated the communication of emotions in music performance, and compared communication of emotion in music performance with vocal expression. These aspects of Study I are not presented here, since they are not the focus of this thesis. However, in Study I it is argued that it can prove beneficial to study the communication of emotion in vocal expression and music performance in parallel. The communication of emotions in music performance is further investigated by the author in several other papers (e.g., Juslin & Laukka, 2000, 2003, in press; Laukka, 2004; Laukka & Gabrielsson, 2000).

studies used some sort of the "standard content" paradigm, and the number of emotions included ranged from 1 to 15. Ninety studies used emotion portrayals by actors, 13 studies used manipulations of portrayals (e.g., filtering, masking, reversal), 7 studies used mood induction procedures, 12 studies used natural speech samples, and 21 studies used synthesized speech stimuli.

All studies that presented forced-choice decoding data relative to some independent criterion of encoding intention were included in a meta-analysis of decoding accuracy. 39 studies featuring a total of 60 decoding experiments met this criterion and were included in the meta-analysis. 12 studies can be characterized as more or less cross-cultural in that they included analyses of encoders or decoders from more than one nation. 77 studies reported acoustic data, and were thus included in the review of acoustic cue patterns used to express emotions. The emotions included in the review were anger, fear, happiness, sadness and love-tenderness.[5]

2.3 Results

2.3.1 Meta-analysis of decoding accuracy

The first issue investigated in Study I was whether vocal expressions can be cross-culturally communicated. The results of the meta-analysis investigating this issue are shown in Table 2, in terms of both within-cultural and cross-cultural decoding accuracy of discrete emotions. One problem in comparing accuracy scores from different studies is that they use different numbers of response alternatives in the decoding task. To allow comparison of data from different studies, the accuracy scores were transformed to Rosenthal and Rubin's (1989) effect size index for one-sample, multiple-choice-type data, *pi* (π). This index transforms accuracy scores involving any number of response alternatives to a standard scale of dichotomous choice, on which .50 is the null value and 1.00 corresponds to 100% correct decoding. Ideally, an index of decoding accuracy should also take into account the response bias in the decoder's judgments (Wagner, 1993). However, this requires that results be presented in terms of confusion matrices, which very few studies have done. Therefore, the data were summarized in terms of the pi index.

The means and confidence intervals presented in Table 2 suggest that the decoding accuracy is typically significantly higher than what would be expected by chance alone for both within-cultural and cross-cultural vocal ex-

[5] Love-tenderness is not included in most lists of basic emotions (e.g., Plutchik, 1994), though it is regarded as a basic emotion by several researchers (e.g., Panksepp, 2000; Scott, 1980; Shaver et al., 1987). The choice of emotion labels in Study I was also dictated by the comparison of studies of speech and music; these emotions were the only ones for which there were enough studies in both modalities to allow for a comparison.

pression. However, the accuracy was significantly higher (*t*-test, $p < .01$) for within-cultural expression ($\pi = .90$) than for cross-cultural expression ($\pi = .84$). Among the individual emotions, anger and sadness were best decoded, followed by fear and happiness. Tenderness received the lowest accuracy although it must be noted that the estimates for this emotion were based on fewer data points.[6]

Table 2

Results From Meta-Analysis of Decoding Accuracy for Discrete Emotions.

	Anger	Fear	Happiness	Sadness	Tenderness	Overall
Within-Cultural Expr.						
Mean accuracy (π)	.93	.88	.87	.93	.82	.90
95% CI	± .021	± .037	± .040	± .020	± .083	± .023
Number of studies	32	26	30	31	6	38
Number of speakers	278	273	253	225	49	473
Cross-Cultural Expr.						
Mean accuracy (π)	.91	.82	.74	.91	.71	.84
95% CI	± .017	± .062	± .040	± .018	-	± .024
Number of studies	6	5	6	7	1	7
Number of speakers	69	66	68	71	3	71

The pattern of results visible in Table 2 is consistent with previous reviews of vocal expression featuring fewer studies (e.g., Johnstone & Scherer, 2000), but differs from the pattern found in studies of facial expression, in which happiness is usually better decoded than other emotions (Elfenbein & Ambady, 2002).

2.3.2 Acoustic cue patterns used to express discrete emotions

The second issue investigated in Study I was whether there are distinct patterns of voice cues that correspond to discrete emotions. The main results from the review of the code usage literature are shown in Table 3. In this table, patterns of voice cues used to express different emotions, as reported in 77 studies of vocal expression, are displayed. Very few studies have re-

[6] Additional data from the meta-analysis were presented in Laukka and Juslin (2002). These data indicate that also disgust and surprise are accurately decoded within-culturally (disgust: $\pi = .83$, *N* studies = 11, *N* speakers = 138; surprise: $\pi = .88$, *N* studies = 12, *N* speakers = 167). Also, there are indications that disgust and surprise may be universally recognized (van Bezooijen et al., 1984).

ported data in such detail that permits inclusion in a meta-analysis. It is also usually difficult to compare quantitative data across different studies because different studies use different baselines (as discussed in Study II). The most prudent approach was thus to summarize findings in terms of broad categories (e.g., high, medium, low) mainly according to how the authors of the various studies themselves had interpreted the data, but whenever possible with support from actual data provided in tables and figures.

The results shown in Table 3 are generally consistent with Scherer's (1986) predictions (see Table 1), which presume a correspondence between emotion-specific physiological changes and voice production. In a direct

Table 3

Patterns of Acoustic Voice Cues Used To Express Discrete Emotions in Studies of Vocal Expression

		Number of studies for each emotion label				
Voice Cue	Category	Anger	Fear	Happiness	Sadness	Tenderness
F_0 mean	High	**33**	**28**	**34**	4	1
	Medium	5	8	2	1	0
	Low	5	3	2	**40**	**4**
F_0 variability	High	**27**	9	**33**	2	0
	Medium	4	6	2	1	0
	Low	4	**17**	1	**31**	**5**
F_0 contour	Up	**6**	**6**	**7**	0	1
	Down	2	0	0	**11**	**3**
Jitter (F_0 perturbation)	High	**6**	4	**5**	1	0
	Low	1	4	3	**5**	0
Voice intensity mean	High	**30**	**11**	**20**	1	0
	Medium	1	3	6	2	0
	Low	1	8	0	**29**	**4**
Voice inten-variability	High	**9**	**7**	**8**	2	0
	Medium	1	4	3	1	0
	Low	2	1	2	**8**	0

(table continues)

Table 3 (*continued*)

		Number of studies for each emotion label				
Voice Cue	Category	Anger	Fear	Happiness	Sadness	Tenderness
Voice onsets (Attack)	Fast	1	1	**2**	1	0
	Slow	1	1	0	1	1
High frequency energy	High	**22**	**8**	**13**	0	0
	Medium	0	2	3	0	0
	Low	0	6	1	**19**	**3**
F1 mean	High	**6**	1	**5**	1	0
	Medium	0	0	1	0	0
	Low	0	**3**	0	**5**	0
F1 bandwidth	Narrow	**4**	0	**2**	0	0
	Wide	0	**2**	1	**3**	0
Precision of articulation	High	**7**	2	**3**	0	0
	Medium	0	2	2	0	0
	Low	0	2	0	**6**	**1**
Glottal waveform	Steep	**6**	2	**2**	0	0
	Rounded	0	**4**	0	**4**	0
Speech rate	Fast	**28**	**24**	**22**	1	0
	Medium	3	3	5	5	1
	Slow	4	2	6	**30**	**3**
Proportion of pauses	Large	0	2	1	**11**	**1**
	Medium	0	3	2	0	0
	Small	**8**	**4**	**3**	1	0
Micro-structural regularity	Regular	0	0	**2**	0	**1**
	Irregular	**3**	**2**	0	**4**	0

Note. Text in bold indicates the most frequent finding for respective voice cue.

test, 82% of 32 comparisons of the predictions matched the results of the review. The test included eight voice cues (Mean F_0, F_0 variability, F_0 contour, Mean voice intensity, Voice intensity variability, High-frequency energy, F1, and Speech rate) and four emotions (anger/rage, fear/terror, happiness/elation, and sadness/sadness), and was conducted on the direction effects only (i.e., if the direction of the prediction and results were the same, it was considered a match). For example, predictions and results did not match in the cases of F_0 variability and F1 for fear, or in the case of F1 for happiness.

2.4 Conclusions

The results, which are based on the most extensive review to date, show that (1) communication of emotions may reach an accuracy well above the accuracy that would be expected by chance alone in vocal expression, at least for broad emotion categories corresponding to basic emotions (i.e., anger, fear, happiness, love, and sadness). (2) The findings further indicate that vocal expressions of emotion are accurately decoded cross-culturally, although the accuracy was somewhat lower (i.e., 7 %) than for within-cultural vocal expression. (3) The results strongly suggest that there are emotion-specific patterns of acoustic voice cues that can be used to communicate discrete emotions in vocal expression. Finally, (4) the patterns of voice cues for discrete emotions are largely consistent with Scherer's (1986) theoretical predictions for anger, fear, happiness, and sadness. [7] These results support an evolutionary perspective on vocal expression, as outlined in the Introduction.

Besides providing evidence of distinct patterns of voice cues that correspond to discrete emotions, the review also revealed many gaps in the literature that must be filled by further research. Many cues have not been systematically investigated, and thus the results for many cues are preliminary at best.

In a meta-analysis that was published after Study I was begun, Elfenbein and Ambady (2002) concluded that emotion expressions (both facial and vocal) are cross-culturally recognized, but that there also exists an "in-group advantage", that is, expressions from one's own culture are slightly better recognized. The results of Study I, which are based on a larger material, support this conclusion for vocal expressions. The in-group advantage probably reflects various *pull* effects resulting from different cultural display norms (see Section 1.1.3).

[7] The general pattern of the results is supported by research that was published after the literature search for Study I was completed (e.g., Airas & Alku, 2004; Bänziger, Grandjean, Bernard, Klasmeyer, & Scherer, 2001; Barrett & Paus, 2002; Gendrot, 2003; Hozjan & Kačič, 2003; Lakshminarayanan et al., 2003; Nakamichi, Jogan, Usami & Erickson, 2003; Oudeyer, 2003; Viscovitch et al., 2003).

3. Study II

3.1 Background and aims

Study I showed evidence of distinct patterns of voice cues for discrete emotions. However, inconsistencies in code usage remain, and need to be explained. Two possible causes for these inconsistencies, namely that studies have not controlled for emotion intensity, and that studies have investigated too few voice cues, were investigated in Study II.

In a study using the "standard content paradigm" we studied both encoding and decoding of vocal expressions of discrete emotions with varying emotion intensity. First, we wanted to see if listeners are able to decode the intensity (quantity), as well as the discrete categories (quality), of emotion from vocal expressions. Second, we investigated the question of whether there exist emotion-specific patterns of voice cues that are used by listeners when decoding emotions from vocal expressions. The findings were further compared with Scherer's (1986) predictions.

3.2 Decoding

3.2.1 Method

Speech stimuli

Eight actors (4 men, 4 women; mean age = 49.5) were asked to vocally portray anger, disgust, fear, happiness, and sadness with weak and strong emotion intensity by using verbal material consisting of two short phrases.[8] In addition to the five emotions, the actors were also instructed to perform the verbal material with no expression. Four of the actors were native British English speakers, and four were native Swedish speakers. For the native English speakers, the phrases were: "It is eleven o'clock" and "Is it eleven o'clock?" For the native Swedish speakers, the phrases were: "Klockan är elva" and "Är klockan elva?" The verbal meaning of the phrases is identical in the two languages.

The portrayals were recorded onto a digital audio tape deck in a laboratory room with dampened acoustics. (The actors also recorded other verbal

[8] These emotions are among the most often proposed basic emotions (e.g., Plutchik, 1994).

material at the same session, but that material is not treated in this thesis). Two sentences, eight actors, five emotions, and two emotion intensities yielded 160 portrayals, plus 16 no expression portrayals (eight actors, two sentences); that is, a total of 176 portrayals.

Participants

A total number of 45 listeners took part in the decoding experiments in Study I; 15 in Decoding experiment 1 (8 women, 7 men; mean age = 24) and 30 in Decoding experiment 2 (15 men, 15 women; mean age = 24). All listeners were Swedish university students, and they were either paid or given course credit for their anonymous and voluntary participation.

Procedure

All listening experiments were conducted individually with specially designed computer software to collect the listener judgments. The listeners were instructed to judge/rate the emotional state expressed by the person who speaks. The order of the stimuli was randomized for each listener, and the participants listened to the stimuli either through loudspeakers (Decoding experiment 1) or headphones (Decoding experiment 2). The sound level was kept the same for all participants. They could listen to each portrayal as many times as they needed to make the judgments. Pre-tests were administered so that the participants could familiarize themselves with the procedure.

In Decoding experiment 1, the listeners were asked to judge the emotional expression of each portrayal by means of forced choice. The alternatives that the listeners could choose from were the same as the intended emotions of the portrayals; that is, *anger, disgust, fear, happiness, sadness,* and *no expression*. The forced-choice method was used because it gives an exact estimate of decoding accuracy that can be compared with results from previous studies.

However, the forced-choice paradigm has been criticized on the grounds that it may artificially inflate decoding accuracy (e.g., Russell, 1994). This problem can be avoided if the participants are allowed to rate the stimuli on several scales simultaneously, and by introducing an additional response alternative (i.e., other emotion) that the participant can choose if none of the provided alternatives seems appropriate (Frank & Stennett, 2001). Therefore, in Decoding experiment 2 the listeners were asked to judge the expression of each portrayal on each of seven scales: *anger*, *disgust*, *fear*, *happiness*, *sadness*, *other emotion*, and *emotion intensity*. All scales were numbered from 0 to 10, where 0 designated minimum, and 10 maximum, of the respective attribute.

3.2.2 Results from decoding experiments

Decoding experiment 1

The results showed that the communication of emotions was generally successful. The overall decoding accuracy (i.e., across emotions, intensities, and languages), in terms of proportion correct, was .56. The decoding accuracy for the individual emotions (across intensities and languages) was .58 (anger), .40 (disgust), .60 (fear), .51 (happiness), .63 (sadness), and .72 (no expression). All emotions were decoded with accuracy better than chance, as indicated by χ^2-tests. (The chance level in a forced choice task with six response alternatives is .167). Portrayals with strong intensity were decoded with significantly higher accuracy (M = .60) than portrayals with weak intensity (M = .49; *t*-test, $p < .01$). Further, all emotions were decoded with accuracy above chance, even when response biases were taken into account (e.g., Wagner, 1993). It can also be noted that there was no pre-selection of effective stimuli in this experiment, in contrast to many previous studies (e.g., Banse & Scherer, 1996).

Decoding experiment 2

The results of Decoding experiment 2 are shown in Table 4. The listeners' mean ratings on each scale were subjected to two-way analyses of variance (ANOVA) with repeated measures of emotion (five levels: anger, disgust, fear, happiness, sadness) and intensity (two levels: weak, strong) as independent variables. Separate ANOVAs were conducted for each rating scale (i.e., anger, disgust, fear, happiness, sadness). Because "no expression" portrayals did not vary with regard to level of intensity, they were excluded from the analysis. The main effect of emotion was highly significant for all scales. Also, the main effect of intensity and the interaction effect were significant for all scales except happiness.

Post hoc multiple comparisons (Tukey's HSD) indicated that for each scale, the emotion portrayals corresponding to this scale were rated higher than all other portrayals. Also, portrayals with strong emotion intensity received higher ratings on the intensity scale for all emotions (M = 6.3) than did portrayals with weak emotion intensity (M = 4.5). Thus, the listeners were successful in discriminating the intended expressions of the emotion portrayals, and were also able to decode the intensity of the portrayals. Further, portrayals with strong intensity received higher ratings on the correct scale than did portrayals with weak intensity, except for the happiness scale. This suggests, as did Decoding experiment 1, that portrayals with strong intensity are easier to decode than portrayals with weak intensity.

Table 4

Listeners' Mean Ratings of Emotion Portrayals as a Function of Intended Emotion and Intended Intensity in Decoding Experiment 2

		Rating Scale					
Emotion portrayed		Anger	Disgust	Fear	Happiness	Sadness	Other emotion
Anger	W	3.16**	1.76	0.60	0.19	0.35	2.22
	S	6.82**	2.01	0.34	0.05	0.11	1.33
Disgust	W	1.38	2.18*	0.21	0.20	0.78	3.16
	S	3.72	3.19**	0.41	0.26	0.34	2.14
Fear	W	0.05	0.23	3.86**	0.18	2.57	1.76
	S	0.35	0.52	5.64**	0.12	2.60	1.05
Happiness	W	0.30	0.18	0.86	2.96**	0.50	2.87
	S	1.01	0.51	1.89	3.11**	0.87	2.50
Sadness	W	0.05	0.27	2.11	0.12	3.78**	1.77
	S	0.21	0.44	3.19	0.07	5.25**	1.21
No Expression		0.44	0.47	0.43	0.18	1.10	2.66

Note. Asterisk indicates that the portrayed emotion received a rating on the corresponding scale which was significantly different (Tukey's *HSD*) from the ratings of other portrayed emotions of the same intensity on the same scale: W = weak emotion intensity; S = strong emotion intensity. * $p < .05$ ** $p < .001$

3.3 Encoding

From the findings of the decoding experiments, it can be expected that the emotion portrayals should show certain characteristics. First, if listeners can decode the emotions, this implies that there are emotion specific patterns of voice cues. Second, if listeners are able to decode the emotion intensity of the portrayals, this implies that the portrayals convey acoustic information that allows listeners to make such inferences.

3.3.1 Acoustic analysis of the speech stimuli

The 176 emotion portrayals were subjected to detailed acoustic analyses regarding 20 acoustic voice cues (see Table 5 for a description). All acoustic measurements were done using the *Praat* (Boersma & Weenink, 1999) and

Soundswell (Ternström, 1996) speech analysis software. For details on how the analyses were made, see Study II, Appendix.

For each encoder, the raw values of each voice cue were transformed into *z*-scores in order to minimize variance caused by individual differences in baseline between encoders (Banse & Scherer, 1996). All within-speaker proportions between cue values for different emotions remain unaltered by this transformation.

Table 5

Voice Cues Analyzed in Study II

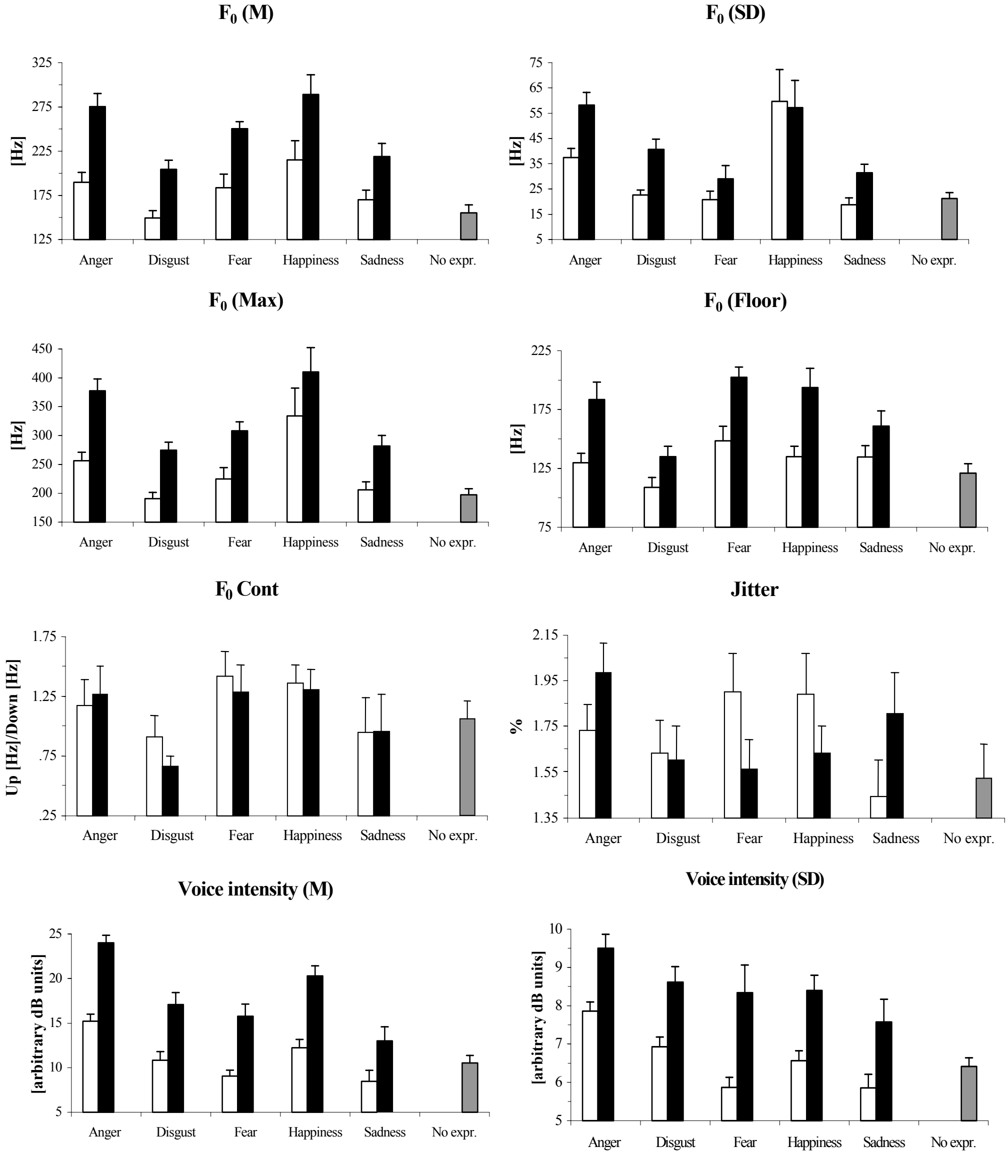
Voice cue	Abbreviation	Perceptual attribute
Fundamental frequency (*M*)	F_0 (*M*)	Mean pitch
Fundamental frequency (*SD*)	F_0 (*SD*)	Pitch variability
Fundamental frequency (max)	F_0 (max)	Pitch maximum
Fundamental frequency (floor)	F_0 (floor)	Pitch base level
Fundamental frequency contour	F_0 contour	Pitch contours (up/down)
Fundamental freq. perturbations	Jitter	Pitch perturbations
Voice intensity (*M*)	VoInt (*M*)	Loudness
Voice intensity (*SD*)	VoInt (*SD*)	Loudness variability
Voice onsets	Attack	Attack of voice onsets
High-freq. energy (cut-off 500 Hz)	HF 500	Voice quality
High-freq. energy (cut-off 1000 Hz)	HF 1000	Voice quality
Formant 1 (*M*)	F1	Voice quality
Formant 1 (bandwidth)	F1 (bw)	Voice quality
Formant 1 (precision)	F1 (prec)	Precision of articulation
Formant 2 (*M*)	F2	Voice quality
Formant 2 (bandwidth)	F2 (bw)	Voice quality
Formant 3 (*M*)	F3	Voice quality
Formant 3 (bandwidth)	F3 (bw)	Voice quality
Speech rate		Velocity of speech
Pause proportion	Pause prop.	Amount of pauses in speech

3.3.2 Acoustic correlates of discrete emotions

A four-way ANOVA, split-plot design, with intended emotion (5 levels) and intended intensity (2 levels) as within-group variables, and encoder language (2 levels) and encoder gender (2 levels) as between-groups variables, was conducted separately for each voice cue.

The ANOVAs confirmed that intended emotion and intended intensity had an effect on most voice cues. No significant main effects of either language or gender were found, which confirmed that the *z*-transformation eliminated baseline differences due to these factors.

The mean values of the voice cues as a function of intended emotion showed characteristic patterns for the different emotions (see Figure 1; only voice cues with significant effects in the ANOVAs are included).



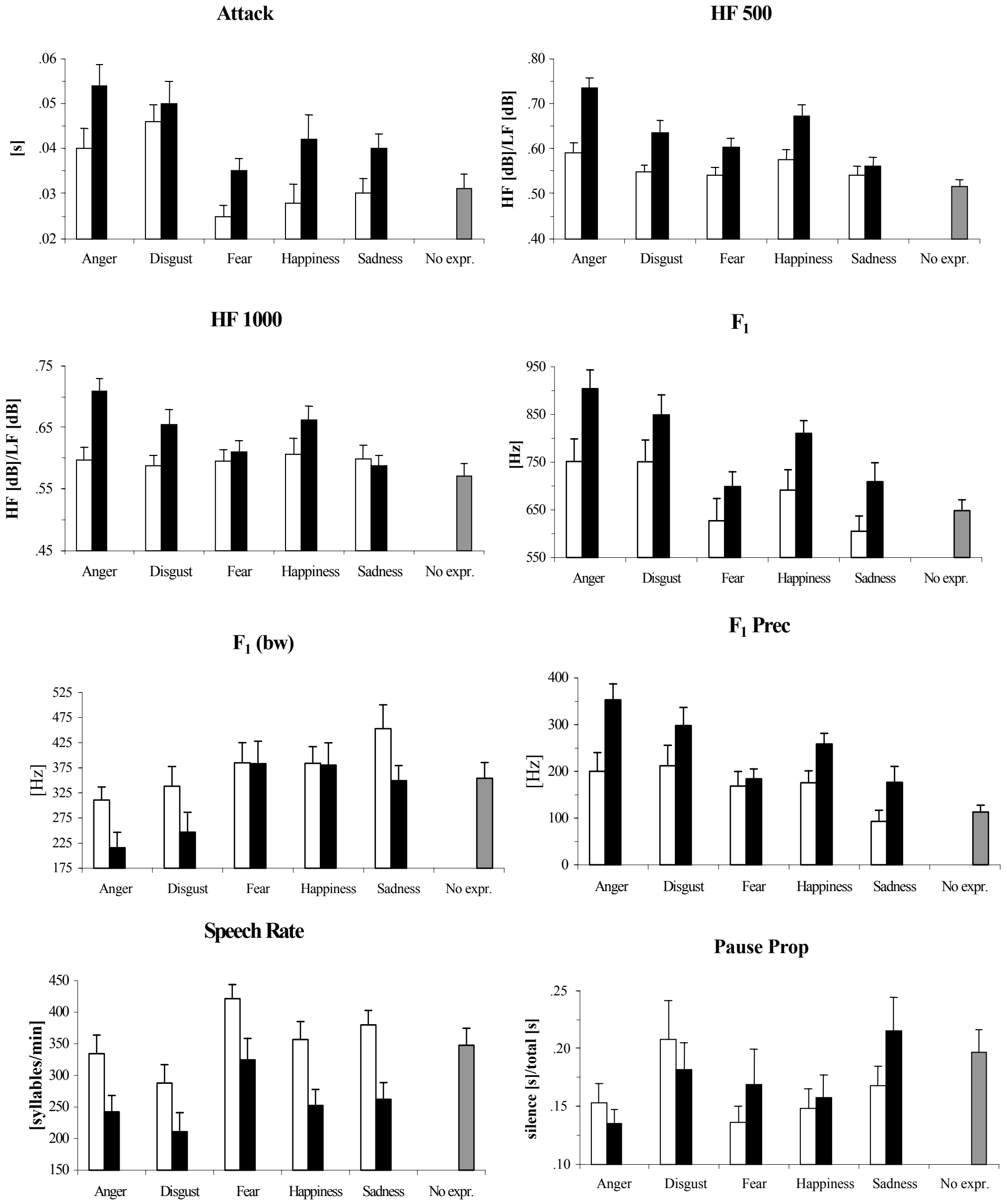


Figure 1 (this and previous page). Mean values (+*SE*) for voice cues as a function of intended emotion and intended emotion intensity. Light bars indicate weak emotion intensity; dark bars indicate strong emotion intensity.

Further, as seen in Figure 1, the results showed that portrayals of the same emotion with different intensity yielded different effects on acoustic cues. For instance, F_0 level (mean, floor, and maximum) varied as a function of intended intensity (the higher the intensity, the higher the F_0). In some cases the differences were actually larger between different intensities of the same emotion (e.g., fear), than between different emotions of the same intensity (e.g., weak happiness vs. weak sadness). The effects of emotion intensity were also strong for voice intensity (higher emotion intensity yielded higher voice intensity) and the measures of spectral energy distribution (higher emotion intensity yielded a larger amount of high-frequency energy in the voice). As for temporal cues, portrayals of strong emotion intensity yielded slower speech rate than portrayals of weak emotion intensity.

The present results regarding code usage were compared with Scherer's (1986) predictions. In a direct test, our results matched the predictions in 57% of 69 comparisons. The test included nine voice cues [F_0 (*M*), F_0 (*SD*), VoInt (*M*), VoInt (*SD*), F1, F1 (bw), F1 (prec), HF 500, and Speech rate] and nine emotion labels (weak anger/irritation, strong anger/rage, strong disgust/disgust, weak fear/anxiety, strong fear/terror, weak happiness/happiness, strong happiness/elation, weak sadness/sadness, strong sadness/grief), and was conducted on the direction effects only (i.e., if the direction of the prediction and results were the same, it was considered a match). Notably, the predictions that fear portrayals with weak emotion intensity are associated with increases in F_0 (*SD*) and high-frequency energy were not supported (see also Banse & Scherer, 1996).

3.3.3 Listeners' cue utilization

To explore which vocal cues are the best predictors of perceived emotions, multiple regression analyses were conducted with the listeners' mean ratings on the emotion scales as dependent variables and the values of the voice cues for each emotion portrayal as independent variables. One simultaneous analysis (see Cohen & Cohen, 1983) was conducted for the listeners' ratings for each emotion scale. Only 9 voice cues were included in the analysis on the basis of two criteria. First, only cues that were significantly influenced by the intended emotion were included. Second, the cues should not be highly intercorrelated. The following cues were included in the analyses: F_0 floor, F_0 (*SD*), VoInt (*M*), F_0 contour, F1, HF 500, speech rate, pause proportion, and attack.

The results are displayed in Table 6, and show that all of the multiple correlations (*R*) were significant ($p < .000001$). The multiple correlation was largest for anger (.76), followed by disgust (.75), sadness (.70), and fear and happiness (.67). Approximately 51% of the beta weights were significant. No less than three cues yielded significant beta weights for each emotion, which suggests that the listeners utilized many cues. However, the actual set of cues was unique for each emotion, thus showing the importance of taking a large set of voice cues into consideration when predicting listeners' decoding of emotions from vocal expression. Furthermore, the generally low beta weights imply that no single cue is sufficient for predicting listeners' emotion judgments. These results provide preliminary evidence that listeners use emotion-specific patterns of cues to decode emotion portrayals.

Table 6

Summary of Results from Multiple Regression Analyses of Relationships between Voice Cues and Listeners' Emotion Ratings in Terms of Standardized Regression Coefficients (β) and Multiple Correlations (R)

Voice cue	Anger	Disgust	Fear	Happiness	Sadness
β					
F_0 (floor)	-.22*	-.37*	.55*	.15*	.45*
F_0 (*SD*)	-.02	-.12	-.24*	.75*	-.17*
F_0 Cont	.02	-.13*	.08	.02	.02
VoInt (*M*)	.25*	-.13	.06	-.18	-.44*
Attack	.25*	.31*	-.07	-.05	.02
HF 500	.44*	.52*	-.29*	-.18	-.15
F1	.20*	.32*	-.12	-.03	-.16*
Speech rate	.03	-.08	.29*	-.17	.05
Pause proportion	-.11	-.08	.28*	-.15*	.20*
R (R^2)	.76 (.58)	.75 (.57)	.67 (.45)	.67 (.45)	.70 (.48)

Note. For explanation of abbreviations, see Table 5. *N* = 160 (The regression analyses included all emotion portrayals except the 16 ”no expression” portrayals). * *p* < .05

We also conducted a multiple regression analysis specifically for the ratings of emotion intensity. We selected cues that were strongly correlated with the ratings of emotion intensity, with additional constraints that the cues (a) should have yielded significant effects of emotion and intensity in the ANOVAs, and (b) should not be highly intercorrelated. The following six cues were selected: F_0 floor, F_0 (*SD*), VoInt (*M*), F1, HF 500, and attack. A simultaneous multiple regression with mean emotion intensity rating as the dependent variable and the cues as independent variables yielded a highly significant multiple correlation (*R* = .84, *p* < .000001), with significant beta weights for F_0 floor (β = .40), F_0 (*SD*) (β = .20), F1 (β = .15), HF 500 (β = .12), and attack (β = .19) albeit not for VoInt (*M*). In fact, removal of VoInt (*M*) from the regression equation did not appreciably influence either the multiple correlation (*R* = .84, *p* < .000001) or the beta weights for any of the remaining cues. Thus, five cues alone can account for the majority of variance in listeners' judgments of emotion intensity in the portrayals.

3.4 Conclusions

The results showed that (a) portrayals with strong emotion intensity yielded higher decoding accuracy than portrayals of weak intensity, (b) listeners were able to decode the intensity of portrayals,[9] (c) there were specific patterns of voice cues for discrete emotions, (d) emotion intensity had a large impact on voice cues, and (e) it was possible to predict the listeners' ratings of emotion categories, as well as of emotion intensity, from a selection of voice cues. Also, the patterns of voice cues for discrete emotions of varying intensity found in this study supported at least some of Scherer's (1986) predictions.

These results show that it is very important to take the emotion intensity into account when studying vocal expression. Failure to consider emotion intensity may have caused some of the inconsistencies in the code usage literature. For instance, Study 1 revealed that fear had a bimodal distribution of some voice cues (e.g., voice intensity, high-frequency energy; see Table 3). It is possible that this result is a consequence of different studies having looked at fear of different intensities.

[9] The finding that emotion intensity can be decoded from vocal expressions has recently been replicated by Rothman and Nowicki (2004), and Song, Chen, Bu and You (2004). See also additional data from Study II reported in Juslin & Laukka (2004).

4. Study III

4.1 Background and aims

Studies I and II strongly suggested that there are distinct patterns of voice cues that correspond to discrete emotions, though inconsistency in code usage was also evident. Hence it has also been suggested that vocal expressions do *not* express discrete emotions directly, but merely the activation dimension of emotions (Pakosz, 1983; Davitz, 1964a), or a combination of activation and valence (Bachorowski, 1999). In Study III, we thus consider the possibility that a discrete-emotions framework may not provide the best possible account of vocal expression by exploring a dimensional approach to vocal expression.

It has been suggested that the affective states most often expressed vocally in everyday life may not be prototypical emotion episodes corresponding to discrete emotions (Cowie & Cornelius, 2003). Instead, weaker affective states like moods, attitudes or stress may be more prevalent.[10] Such affective states may be better described by adopting a dimensional approach to emotions (e.g., Russell & Feldman Barrett, 1999). Thus another reason for exploring a dimensional approach to vocal expressions is that it may be useful in investigating the subtleties of everyday vocal expression.

Consequently, in Study III we were concerned with the task of finding acoustic correlates of emotion dimensions. Questions raised included: Which emotion dimensions have acoustic correlates? Is it only activation and valence, as suggested above, or can other important emotion dimensions like potency also be conveyed by vocal expressions?

4.2 Method

In Study III, listeners judged the speech stimuli from Study II on scales reflecting emotion dimensions. Thirty students (15 men, 15 women; mean age = 24) and 6 expert judges (speech researchers from the Speech, Music, and

[10] Moods are usually distinguished from emotions by their longer duration, weaker intensity, and lack of an object (e.g., Frijda, 1993).

Hearing Department, Royal Institute of Technology, Stockholm: 3 women, 3 men; mean age = 44) participated in the study.

The listeners were asked to rate the portrayals on each of 4 scales reflecting emotion dimensions: *activation*, *valence*, *potency*, and *emotion intensity*.[11] The ratings were made on scales ranging from 0 (low activation, negative valence, low potency, and low intensity) to 10 (high activation, positive valence, high potency, and high intensity). In other respects, the procedure was the same as in Study II (Decoding experiment 2). The acoustic analyses of the speech stimuli presented in Study II were used to investigate the acoustic correlates of emotion dimensions (see Table 5).

4.3 Results

4.3.1 Listening experiment

The ratings of the students and the expert judges were highly correlated for all scales, wherefore data from the two groups were collapsed. A two-way ANOVA, repeated measures, with emotion (5 levels) and intensity (2 levels) as independent variables was conducted separately for the listeners' ratings on each scale (i.e., activation, valence, potency, intensity). The listeners' ratings were aggregated across portrayals within each listener for each condition (i.e., emotion and intensity). Because no expression portrayals did not vary with regard to level of intensity, they were excluded from the ANOVAs.

Significant main effects of emotion and intensity were obtained for all dimension scales. There were also significant interactions between emotion and intensity for the activation, valence, and potency scales. *Post hoc* multiple comparisons (Tukey's *HSD*) revealed that all differences between emotions in the activation scale were significant; especially anger portrayals were rated higher, and sadness portrayals lower, on activation than the other portrayals. For the valence scale, happiness portrayals received higher ratings than the other portrayals, and for the potency scale, fear and sadness were rated lower than the other portrayals. Anger portrayals were further rated higher, and sadness lower, than the other portrayals on the intensity scale.

As in Study II, portrayals with strong intensity were rated higher on the intensity scale than portrayals with weak intensity for each emotion. Differences between the ratings of portrayals of strong and weak intensity were largest for anger and disgust concerning activation and potency, and largest for anger and happiness concerning valence. Further, potency involved different effects of intensity depending on the emotion. Specifically, potency

[11] These four dimensions have been obtained in many studies of subjective feeling states (Smith & Ellsworth, 1985).

increased in strong anger and strong disgust, whereas it decreased in strong happiness, strong fear, and strong sadness. One possible explanation could be that when the intensity of fear and sadness increases, the appraised coping potential decreases (e.g., Scherer, 2001).

The differences among emotions were smallest for the intensity scale. This makes sense because each emotion may be strong or weak in a particular portrayal, but we would not expect strong overall differences between the emotions. As regards the other dimension scales, different emotions seemed to involve partly different patterns of ratings: *happiness* (high activation, positive valence, high potency); *fear* (moderate activation, negative valence, low potency); *sadness* (low activation, negative valence, low potency). These three patterns were different from the patterns for *anger* and *disgust* that, however, were quite similar to each other (high activation, negative valence, high potency).

4.3.2 Acoustic correlates of emotion dimensions

A crucial aim of Study III was to explore whether specific voice cues are associated with specific emotion dimensions. First, the listeners' dimensional ratings were aggregated across all listeners. Then, the correlations (Pearson *r*) between the listeners' mean ratings on each dimension scale and the values of the voice cues for each portrayal were calculated. All four dimensions were significantly correlated with a number of voice cues. The overall effect size was largest for intensity (*r* = .40) and activation (*r* = .39), followed by potency (*r* = .23) and valence (*r* = .17). The patterns of voice cues associated with each emotion dimension are shown in Table 7.

Table 7

*Correlations (*r*) between Listeners' Mean Ratings on Each Emotion Dimension and Voice Cues*

Voice cue	Activation	Valence	Potency	Intensity
F_0 (*M*)	.62***	-.21**	-.02	.72***
F_0 (*SD*)	.62***	.08	.34***	.54***
F_0 (max)	.68***	-.09	.12	.72***
F_0 (floor)	.36***	-.36***	-.25***	.53***
F_0 contour	.07	.08	-.07	.06
Jitter	.07	-.03	-.01	.17*
VoInt (*M*)	.80***	-.26***	.44***	.74***
VoInt (*SD*)	.66***	-.34***	.41***	.67***

(table continues)

Table 7 (*continued*)

Voice cue	Activation	Valence	Potency	Intensity
Attack	.39***	-.21**	.34***	.39***
HF 500	.74***	-.33***	.40***	.74***
HF 1000	.54***	-.31***	.30***	.55***
F1	.57***	-.19*	.39***	.50***
F1 (bw)	-.31***	.12	-.37***	-.23**
F1 (prec)	.40***	-.23**	.30***	.41***
F2	.11	-.14	.00	.17*
F2 (bw)	.05	-.17*	-.03	.18*
F3	-.02	-.09	-.21**	.09
F3 (bw)	-.18*	.09	-.20**	-.09
Speech rate	-.41***	.20**	-.28***	-.47***
Pause proportion	-.25***	-.04	-.14	-.15*

Note. For explanation of abbreviations, see Table 5. $N = 176$. * $p < .05$. ** $p < .01$. *** $p < .001$

4.3.3 Listeners' cue utilization

To explore which voice cues are the best predictors of perceived emotion dimensions, multiple regression analyses with the listeners' mean ratings of the emotion dimensions as dependent variables and the values of the voice cues for each emotion portrayal as independent variables were conducted in a similar way as in Study II. F_0 (*SD*), F_0 floor, mean voice intensity, F1, HF 500, and speech rate were the cues chosen for inclusion in these analyses, based on their correlations with the ratings on respective dimension.

The results from the multiple regression analyses are shown in Table 10. The multiple correlations (*R*) were high for activation (.86), potency (.75), and intensity (.83), but considerably lower for valence (.50; all *p*'s < .001). Whereas a linear model including 6 cues could explain 56 - 74% of the variance in the listeners' ratings of activation, potency, and intensity, only 25% of the variance in valence ratings could be explained. Different cues were important for predicting different dimensions, and all included cues received significant beta weights for at least some dimension. The most important predictors, in terms of variance accounted for, were: mean voice intensity for activation; F_0 (*SD*) and HF 500 for valence; mean voice intensity and F_0 floor for potency; and HF 500 and F_0 floor for intensity. Further, F1 was an important predictor of activation but not of intensity, and F_0 floor was an important predictor of intensity but not of activation. F_0 floor was also an important predictor of potency, though *low* F_0 floor was predictive of high potency, whereas *high* F_0 floor was predictive of high intensity.

Table 10

Summary of Results from Multiple Regression Analyses of Relationships between Voice Cues and Listeners' Dimensional Ratings in Terms of Standardized Regression Coefficients (β) and Multiple Correlations (R)

Voice cue	Activation	Valence	Potency	Intensity
β				
F_0 (floor)	-.06*	-.21*	-.69*	.22*
F_0 (SD)	.20*	.34*	-.07	.17*
VoInt (M)	.50*	.01	.61*	.19*
HF 500	.20*	-.33*	.20*	.31*
F1	.15*	-.04	.14*	.07
Speech rate	.01	.15*	-.05	-.14*
R (R^2)	.86 (.74)	.50 (.25)	.75 (.56)	.83 (.69)

Note. For explanation of abbreviations, see Table 5. *N* = 176. * *p* < .05

4.4 Conclusions

The results from Study III indicate that all investigated emotion dimensions (activation, valence, potency, and emotion intensity) were correlated with many voice cues, and that the listeners' ratings could be successfully predicted from the voice cues for all dimensions except valence. Also, the finding from Study II, suggesting that listeners can decode emotion intensity, was replicated using a different group of listeners.

These results clearly show that activation is not the only dimension that is reflected in the acoustics of emotional speech. Nor is a combination of activation and valence the only thing that is conveyed by vocal expressions, contrary to some previous claims (Russell et al., 2003). This conclusion is also supported by the findings from Studies I and II, where evidence for emotion-specific patterns of voice cues were obtained.

Valence was not as well conveyed by the voice cues as were the other dimensions. However, because a larger than usual set of voice cues was included, we did find several acoustic correlates of valence, as opposed to some earlier studies that have investigated fewer cues (e.g., Davitz, 1964a; Pereira, 2000; see also Schröder, Cowie, Douglas-Cowie, Westerdijk & Gielen, 2001).

5. Study IV

5.1 Background and aims

The previous studies strongly suggest that vocal expressions of discrete emotions are acoustically differentiated. However, these studies cannot answer the question of how a listener in fact perceives vocal expressions. In principle it could be possible for the acoustical structure of vocal expressions to be structured mainly according to discrete emotion categories, while the same expressions are *perceived* as continuous dimensions. The possibility thus remains that listeners may perceive primarily the activation, or arousal, level of the vocalizations instead of discrete emotions (e.g., Bachorowski & Owren, 2003).

The aim of Study IV was thus to investigate how vocal expressions are perceived; as continuous dimensions or as discrete categories. One way of doing this is by investigating if emotion expressions are categorically perceived. Evidence of CP of vocal expressions would fit well within a discrete emotions approach to emotion perception, but would be harder to reconcile with a dimensional approach to emotion perception. Study IV attempted a first investigation of CP of vocal expressions using the standard methodology for assessing CP (described below).

The hallmark of CP is greater sensitivity to a physical change when it crosses the boundary between two perceptual categories, than to the same change occurring within a particular category (Harnad, 1987; Repp, 1984). In a classic study, Liberman, Harris, Hoffman and Griffith (1957) used synthetic speech to generate a series of consonant-vowel syllables (e.g., /be/ and /de/) going from one syllable to another. Firstly, they found that subjects identified stimuli on one side of an abrupt boundary as /be/ and stimuli on the other side as /de/, even though the stimuli were randomly sampled from smoothly varying continua. Secondly, subjects showed better discrimination of stimuli that crossed the category boundary defined by their identification behavior than of stimuli belonging to the same category, even though the physical differences between the stimuli were identical. Thirdly, they found that discrimination could be predicted from the identification function. This method of operationally assessing CP in terms of behavioral measures has since become known as the standard methodology, and the above pattern of results are often cited as standard requirements of CP.

5.2 Experiment 1

In Experiment 1, evidence of CP of emotion in vocal expression was tested using the standard methodology for assessing CP (e.g., Liberman et al., 1957; Young et al., 1997). Continua of vocal expressions were created using speech synthesis. Each continuum consisted of a series of vocal expressions, differing by constant physical amounts, interpolated (morphed) from one expression to another. Subjects were asked to identify the emotion of each expression and to discriminate between pairs of expressions.

It was hypothesized that if vocal expressions are perceived in a categorical fashion, (1) subjects should identify expressions as belonging to two distinct sections separated by a sharp category boundary, even though the expression information is linearly manipulated; (2) it should be easier for subjects to discriminate between two stimuli that are perceived as expressing different emotions than between two stimuli that are perceived as expressing the same emotion, even though the physical differences are identical; and (3) it should be possible to predict the form of the discrimination function from the subjects' identification performance.

5.2.1 Synthesis of emotional speech

Recent development in speech synthesis has made it possible to synthesize vocal expressions that are recognized by listeners with an accuracy above chance (for a review, see Schröder, 2001). Six continua (anger-fear, anger-sadness, fear-happiness, fear-sadness, happiness-anger, and happiness-sadness) of vocal expressions were created using concatenative speech synthesis. Each continuum consisted of a series of vocal expressions differing by constant physical amounts, morphed from one prototype emotional expression to another. The prototype expressions were spoken by a female actor and were taken from Study II.

Synthesized expressions were created based on the acoustic measurements of the prototype portrayals (see Study II). The syllables of the neutral portrayal were used as building blocks ("Klo-ckan-är-el-va"), and these were manipulated to resemble the prototype portrayals as closely as possible. The speech synthesis was based on the neutral expression so that properties that were not manipulated (e.g., formant structure) would be emotionally neutral. Vocal cues related to F_0 and temporal aspects were manipulated using the *TD-PSOLA* (Time-Domain Pitch-Synchronous OverLap-and-Add) technique (Moulines & Charpentier, 1990), as implemented in the Praat software (Boersma & Weenink, 1999). Voice intensity was manipulated by up- or downscaling the intensity of the neutral syllables so that they received the same intensity as the syllables of the prototype expressions. Voice quality, finally, was manipulated by amplifying the relative amount of high-frequency energy for the anger expression, and then manipulating the other expressions relative to the anger expression.

Morphed expressions were created by blending between two prototype expressions. The values of the manipulated vocal cues were linearly interpolated between the values of the prototype portrayals. Continua of all possible combinations between the emotions anger, fear, happiness, and sadness were created (i.e., anger-fear, anger-sadness, fear-happiness, fear-sadness, happiness-anger, and happiness-sadness). Morphs were created in proportions 90:10 (e.g., for the happiness-sadness continuum, 90% happiness and 10% sadness), 70:30 (70% happiness and 30% sadness), 50:50 (50% happiness and 50% sadness), 30:70 (30% happiness and 70% sadness), and 10:90 (10 % happiness and 90% sadness). These will be referred to as 90%, 70%, 50%, 30%, and 10% morphs along the appropriate continuum. A total of 34 synthesized vocal expressions (4 prototype expressions and 30 morphed expressions) were made. For more details about the stimulus creation, the reader is referred to Study IV.

5.2.2 Method

Thirty-four Swedish students (14 men, 20 women; mean age = 25) participated in the experiment. They were either paid or given course credit for their confidential and voluntary participation. Listening experiments were conducted individually with specially designed software to collect the individual judgments. No feedback was given as to the appropriateness of the response. The participants listened to the stimuli through headphones. Pretests were conducted prior to the real experiment, so that the participants could get acquainted with the procedure.

The subjects were first asked to discriminate between pairs of expressions in a sequential discrimination (ABX) task, in which stimuli A, B, and X were presented sequentially and the subjects had to decide whether X was the same as A or B. For all stimulus pairs, X was either identical to A or identical to B. The pairs to be compared consisted of all combinations for each emotion continuum that differed by 20%; e.g., the 90% morph vs. the 70% morph, the 70% morph vs. the 50% morph, the 50% morph vs. the 30% morph, and the 30% morph vs. the 10% morph. The presentation order of the stimulus pairs was randomized within and across all emotion continua, and the subjects received four trials of each stimulus pair representing all possible orders of presentation (ABA, ABB, BAA, and BAB). Each subject was thus asked to discriminate between 96 stimulus pairs (4 emotions x 6 emotion continua x 4 presentation orders).

The listeners were then asked to judge the emotion of each synthesized expression by means of forced-choice. The alternatives they could choose from were the same as the two end-emotions of the continuum to which the respective expression belonged. The order in which the different continua were presented was randomized, and the order of presentation of the stimuli was randomized within each emotion continuum. Each expression was presented three times.

5.2.3 Results

Results from the identification experiment showed that each emotion continuum was perceived as two distinct sections separated by a sudden category boundary (see Figure 2).

The results from the discrimination task were assessed using two different approaches. Firstly, using a subject-by-subject method (e.g., de Gelder et al., 1997), two measures of discrimination performance were calculated for each subject. (1) a "peak" value – the mean discrimination accuracy for the expression pair that crosses the identification category boundary (i.e., contains the 50% identification point), and (2) a "nonpeak" value – the mean discrimination accuracy for the remaining pairs. A repeated measures ANOVA, with type of discrimination (2 levels: peak and nonpeak) and continua (6 levels) as within-subject factors, was conducted on the discrimination performance values. There were significant main effects of type of discrimination and continua, but the interaction effect was not significant. Across all continua, peak pairs were easier to discriminate (mean accuracy = .82) than nonpeak pairs (mean accuracy = .76). *Post hoc* multiple comparisons (Fisher's *LSD*) showed significant differences for the anger-fear, fear-happiness, and happiness-anger continua (p's < .05), but not for the continua that included sadness.

Secondly, subjects' discrimination performance was predicted from their identification rates, and these predictions were compared with the observed discrimination results using a procedure proposed by Liberman et al. (1957).[12] The fit between predicted and observed performance was assessed by correlating predicted and observed scores, and converting this to a t-value (Young et al., 1997). This showed a significant correlation of predicted and observed performance ($r = .42$, $t = 2.17$, $df = 22$, $p < .05$). The magnitude of the correlation is similar to that of some previous studies on categorical perception of emotion expressions (e.g., Young et al., 1997).

Taken together, the results from Experiment 1, show categorical effects in the perception of emotion from vocal expressions, at least for some emotions.

[12] An underlying assumption of this procedure is that the subjects covertly classify each of the stimuli presented in the discrimination task. For details on the calculation of the predicted discrimination performance, the reader is referred to Study IV, Appendix (see also Liberman et al., 1957, and Macmillan, Kaplan, and Creelman, 1977, p. 454).

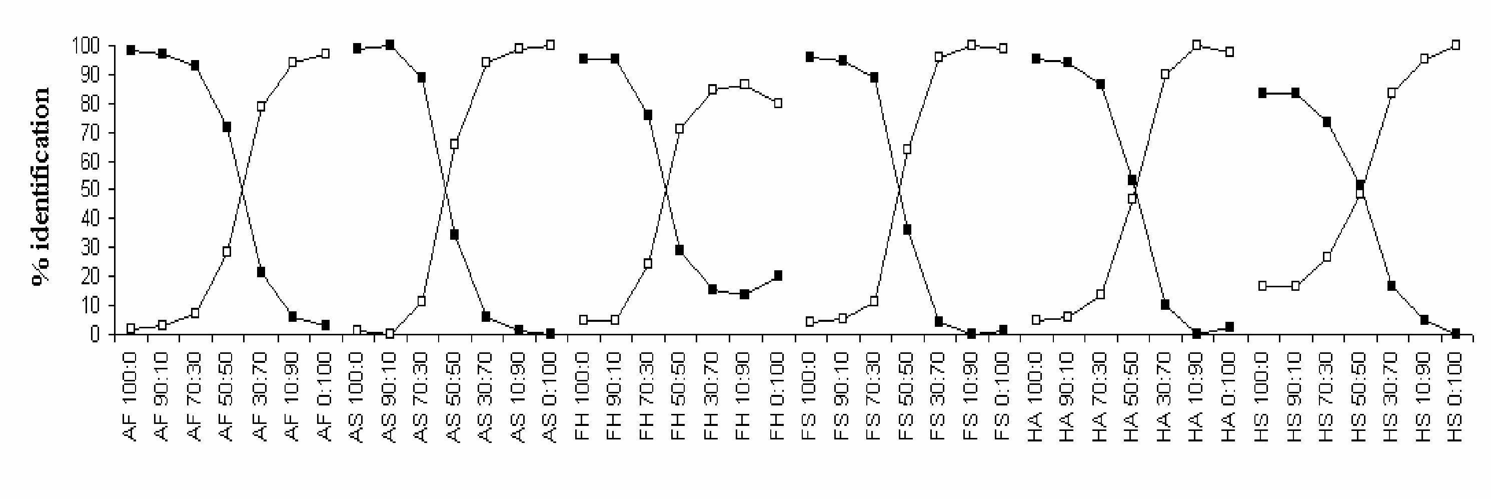


Figure 2. Decoding accuracies with which the vocal expressions were identified for each continuum [from left to right, anger-fear (AF), anger-sadness (AS), fear-happiness (FH), fear-sadness (FS), happiness-anger (HA), and happiness-sadness (HS)], in Experiment 1. For each continuum the morphs are shown in the following order; from left to right, 100%, 90%, 70%, 50%, 30%, 10% and 0%. [Dark squares = percentage of identification of the first emotion of each continuum (e.g., anger for the anger-fear continuum), light squares = percentage of identification of the second emotion of each continuum (e.g., fear for the anger-fear continuum)].

5.3 Experiment 2

In Experiment 2, a replication of the findings from Experiment 1 was attempted, using a slightly modified procedure. Firstly, a different sort of discrimination task was used; a dual-pair (4IAX) task. The 4IAX task makes less demands on short-term memory than the ABX discrimination test, since a correct decision can be made by a pairwise comparison. Thus it is assumed to be more sensitive to purely auditory information (e.g., Pisoni, 1975). It has also been noted that the 4IAX procedure is less biased (i.e., less dominated by a subjective internal criterion) than the ABX procedure (e.g., Schouten, Gerrits, & van Hessen, 2003). Secondly, a single continuum ranging anger – fear – happiness – sadness – anger was used. Because such a continuum has no fixed end-points, all morphs could be used equally often in the discrimination task. Thirdly, in the identification test, the listeners were asked to rate the expression of the stimuli on rating scales, instead of the two-alternative forced-choice task used in Experiment 1, to get a more reliable estimate of the decoding accuracy.

5.3.1 Method

A single continua ranging anger – fear – happiness – sadness – anger was created. This is equal to the four continua anger – fear, fear – happiness, happiness – sadness, and sadness – anger. For each of these continua, seven synthesized expressions (five morphs and two prototype expressions) were created (proportions 100:0, 90:10, 70:30, 50:50, 30:70, 10:90, and 0:100). Five morphs in each of four continua, together with four prototype expressions, leads to a total of 24 synthesized vocal expressions. The synthesized vocal expressions were the same as in Experiment 1, with the exception that a different prototype portrayal of sadness was used to create the continua that included sadness. A different synthesized portrayal of sadness was used to see if the failure to find a statistically significant enhancement of between-category differences for continua that contained sadness was due to possible artifacts of the stimuli used.

25 Swedish students (4 men, 21 women; mean age = 22) participated in the experiment. The subjects were first asked to discriminate between pairs of expressions in a dual-pair discrimination (4IAX) task. In the 4IAX test, two pairs of stimuli are presented on every trial; one pair is always the same and one pair is different. The subjects' task was to determine which pair contains the same stimuli, the first pair or the second pair. The stimuli pairs to be compared consisted of all combinations in the anger – fear – happiness – sadness – anger continua that differed by 20% (i.e., the 90% morph vs. the 70% morph, the 70% morph vs. the 50% morph, the 50% morph vs. the 30% morph, the 30% morph vs. the 10% morph, and the 10% morph vs. the 90% morph of the next emotion). The order of presentation of the stimulus pairs

was randomized across all emotion continua, and the stimuli were arranged in the following 4IAX sequences: AA–AB, AA–BA, AB–AA, and BA–AA. Each subject was thus asked to respond to 80 trials (5 stimuli pairs x 4 emotion continua x 4 presentation orders). Note that all morphs occurred in equally many trials. In other respects, the procedure was the same as in Experiment 1.

The participants were next asked to judge the expression of all stimuli on each of four scales: *anger*, *fear*, *happiness*, and *sadness*. All scales were numbered from 0 to 10, where 0 designated minimum and 10 designated maximum of the respective attribute. The order in which the stimuli were presented was randomized across all emotion continua. The participants could listen to each stimulus as many times as they needed to make the judgment.

5.3.2 Results

To obtain a measure of decoding accuracy, the adjective ratings were coded in the following manner: For each expression, the highest rated emotion-scale was coded as 1, and the other emotion-scales were coded as 0. If an expression received equally high ratings on two emotion-scales, both emotion-scales were coded as 0.5, and the other emotion-scales were coded as 0. In the present case, this coding procedure is roughly equivalent to a forced-choice format with four response options (e.g., Scherer et al., 1991).

The decoding accuracies (in percent correct) for all synthesized expressions in each continuum are shown in Figure 3. As can be seen, all prototype expressions were decoded with an accuracy well above chance (chance level = .25). Also, each continuum fell in two regions separated by a category boundary, and each region was primarily identified as the emotion category corresponding to the prototype at that end. As in Experiment 1, the shift from one category to the next was abrupt and at or near the center of the continuum. This shows that the listeners identified the portrayals of a continuum as belonging to either of the respective end-emotions, even though they had four alternatives to choose from and were not forced to choose one particular response alternative.

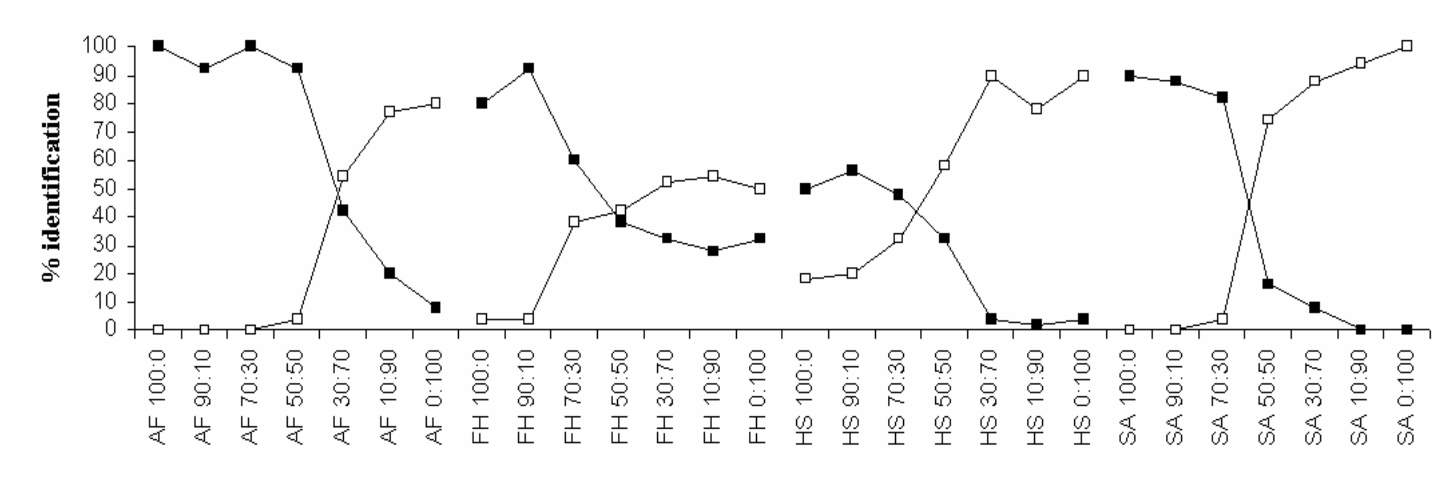


Figure 3. Decoding accuracies with which the vocal expressions were identified for each continuum [from left to right, anger-fear (AF), fear-happiness (FH), happiness-sadness (HS), and sadness-anger (SA)], in Experiment 2. For each continuum the morphs are shown in the following order; from left to right, 100%, 90%, 70%, 50%, 30%, 10% and 0%. [Dark squares = percentage of identification of the first emotion of each continuum (e.g., anger for the anger-fear continuum), light squares = percentage of identification of the second emotion of each continuum (e.g., fear for the anger-fear continuum)].

The results from the 4IAX discrimination task were assessed using the same subject-by-subject method as in Experiment 1 (i.e., peak and non-peak values of discrimination performance were calculated for each subject). A repeated measures ANOVA, with type of discrimination (two levels: peak and non-peak) and continua (4 levels) as within-subjects factors, was conducted on the above measures of discrimination performance. The effects of interest were the main effect of type of discrimination and the interaction effect.

A significant main effect was found for type of discrimination, but not for continua. Also, the interaction was not significant. Across all continua, peak pairs were easier to discriminate (mean accuracy = .91) than non-peak pairs (mean accuracy = .81). Multiple comparisons (Fisher's *LSD*) revealed that the difference between peak and nonpeak pairs was significant for all continua (*p*'s < .01). This suggests that the failure to find enhanced between-category discrimination for continua including sadness in Experiment 1 may have been due to artifacts of the particular stimuli used in that experiment. The discrimination accuracy was generally higher in Experiment 2 than in Experiment 1, which is in accord with earlier findings that the 4IAX procedure yields better discrimination performance than the ABX procedure (e.g., Pisoni, 1975).

Further, the predicted discrimination performance was calculated, based on the decoding accuracy data from the identification experiment, using the same procedure as in Experiment 1 (see Study IV, Appendix, for details). The predicted discrimination performance was then compared with the observed results from the 4IAX experiment. The fit between predicted and observed performance was assessed by correlating predicted and observed scores, and converting this to a *t*-value. This yielded a positive correlation of predicted and observed performance ($r = .39$, $t = 1.80$, $df = 18$, $p < .10$). The correlation is of a similar magnitude as in Experiment 1, but did not reach the conventional criteria for significance. This result can be interpreted in two ways, either as showing that the 4IAX method results in "less" CP than the ABX method (e.g., Pisoni & Lazarus, 1974; Schouten et al., 2003), or as merely reflecting the weak statistical power of the test, which was based on only 20 cases. The latter interpretation is favored by the fact that the correlation between predicted and observed discrimination performance across Experiments 1 and 2 is stronger than for the two experiments separately ($r = .46$, $t = 3.36$, $df = 42$, $p < .005$). This shows that, across the two experiments, discrimination and identification of vocal expressions are not independent. The magnitude of the correlation is comparable to that of previous studies on CP of emotion expressions that have used facial expressions (e.g., Young et al., 1997).

5.4 Conclusions

Evidence of categorical effects in the perception of vocal emotion expressions was found in two experiments. Firstly, subjects identified expressions as belonging to two distinct sections separated by an abrupt category boundary, even though the expression information was linearly manipulated. Secondly, subjects were better at discriminating between two stimuli that were perceived as expressing different emotions, than between two stimuli that were perceived as expressing the same emotion, even though the physical differences were identical. Thirdly, it was possible to predict the form of the discrimination function from the subjects' identification performance.

These findings have a bearing on the question of whether vocal expressions are perceived as varying continuously along underlying emotion dimensions such as activation or valence, or as belonging to qualitatively discrete emotion categories. In general the findings of Study IV are inconsistent with two-dimensional accounts of emotion perception in vocal expression. On such accounts, continua that are close to or run along the hypothesized dimensions would be expected to move through a region with no identifiable expression in the middle, and in that region the subject should respond at random. Also, on an account based on two underlying dimensions, the continua that cross these dimensions should have a discrete region in the middle which belongs to neither prototype (Young et al., 1997). Neither pattern is discernible from the present results. On the contrary, the morphed expressions were identified mainly in terms of the prototypes at either end of the relevant continuum, with high identification rates at each end, and sudden shifts in the middle. Moreover, as evident from Experiment 2, in the regions where identification changed from one emotion to another, there were few intrusions from other categories that did not belong to that continuum. The results from the present study are thus more in line with the idea that vocal expressions are perceptually coded in terms of their conformity to prototype expressions that correspond to discrete emotions. Though emotion dimensions are relevant to the perception of vocal expression, as for instance seen in Study III, it may be that they correspond to intellectual, rather than perceptual, constructs.

6. General discussion

6.1 Main findings

So, are vocal expressions conveyed as discrete emotions or as broad emotion dimensions like activation and valence? Judging from the results of this thesis, the answer is that a discrete-emotions framework provides the best account of vocal emotion expression. A dimensional model consisting of activation and valence cannot do full justice to vocal expression neither with regard to the amount of emotion differentiation that can occur, nor the way vocal expressions are commonly perceived by listeners.

The specific research questions stated in the Introduction can now be answered in the following way:

- A1: The results of Study I show that vocal expressions of discrete emotions are cross-culturally recognized.

- A2: The results from Studies I, II, and III strongly suggest that there exist distinct patterns of voice cues that correspond to discrete emotions.

- A3: The results from Study IV suggest that vocal expressions are perceived as discrete emotions, and not as continuous dimensions.

Table 9 summarizes the results on code usage and presents emotion-specific patterns of voice cues, as well as patterns corresponding to emotion dimensions. The patterns for discrete emotions are generally in concordance with Scherer's (1986) predictions, which presumed a correspondence between emotion-specific physiological changes and voice production. However some predictions do not receive support, and thus may have to be revised. For instance, the prediction that F_0 variability increases in fear likely needs to be revised on the basis of the present results. It should be noted that the findings of emotion-specific patterns of voice cues are consistent with both discrete emotion theory and component process theory. However, this thesis did not test the most important assumption of component process theory, namely that there are highly differentiated, sequential patterns of cues that reflect the cumulative result of the adaptive changes produced by a specific appraisal profile (Scherer, 2001).

Table 9

Summary of Patterns of Acoustic Cues for Discrete Emotions and Emotion Dimensions Obtained in the Present Thesis

Anger:	high mean F_0, much F_0 variability, high maximum F_0, rising F_0 contours, much jitter, high voice intensity, much voice intensity variability, much high-frequency energy, high mean F1, narrow bandwidth of F1, precise articulation, steep glottal waveform, fast speech rate, little pauses, and micro-structural irregularity
Fear:	high mean F_0, little F_0 variability, rising F_0 contours, low voice intensity (except in panic fear), much voice intensity variability, little high-frequency energy, low mean F1, wide bandwidth of F1, rounded glottal waveform, fast speech rate, and micro-structural irregularity
Disgust:	low mean F_0, medium F_0 variability, falling F_0 contours, medium voice intensity, medium voice intensity variability, slow voice onsets, medium high-frequency energy, high mean F1, narrow bandwidth of F1, precise articulation, slow speech rate, and much pauses
Happiness:	high mean F_0, much F_0 variability, high maximum F_0, rising F_0 contours, medium-high voice intensity, much voice intensity variability, fast voice onsets, medium high-frequency energy, high mean F1, steep glottal waveform, fast speech rate, and micro-structural regularity
Sadness:	low mean F_0, little F_0 variability, low maximum F_0, falling F_0 contours, little jitter, low voice intensity, little voice intensity variability, little high-frequency energy, low mean F1, wide bandwidth of F1, slackened articulation, rounded glottal waveform, slow speech rate, much pauses, and micro-structural irregularity
Tenderness:	low mean F_0, little F_0 variability, falling F_0 contours, low voice intensity, slow voice onsets, little high-frequency energy, slow speech rate, and micro-structural regularity
Activation: (*high*)	high mean F_0, much F_0 variability, high maximum F_0, high voice intensity, much voice intensity variability, slow voice onsets, much high-frequency energy, high mean F1, narrow bandwidth of F1, precise articulation, and few pauses

(table continues)

Table 9 (*continued*)

Valence: (*positive*)	low mean F_0, low F_0 base level, low voice intensity, little voice intensity variability, fast voice onsets, low mean F1, slackened articulation, little high-frequency energy, and fast speech rate
Potency: (*high*)	large F_0 variability, low F_0 base level, high voice intensity, large voice intensity variability, slow voice onsets, much high-frequency energy, high mean F1, narrow bandwidth of F1, precise articulation, and slow speech rate
Intensity: (*high*)	high mean F_0, large F_0 variability, high maximum F_0, high F_0 base level, jitter, high voice intensity, large voice intensity variability, slow voice onsets, much high-frequency energy, high mean F1, narrow bandwidth of F1, precise articulation, slow speech rate, and few pauses

Note. Results are based on both the review in Study I and the empirical studies (Studies II and III).

Further, the results of this thesis support a discrete-emotions approach to emotion in general. It is true that the findings apply only to one component of emotion (i.e., expression), but like similar findings on facial expressions, evidence of universal vocal expressions of discrete emotions, which are communicated by emotion-specific patterns, can be viewed as supporting discrete-emotions theory (Ekman, 1992, 1994). Likewise, the present findings of CP of vocal expressions of discrete emotions, like similar previous findings regarding facial expressions, lend support to a discrete-emotions approach (Ekman, 1994).

Finally, it is proposed that a dimensional approach to vocal expression may be a viable alternative if one wishes to study milder affective states (e.g., moods, attitudes, or stress) that do not correspond to full-blown emotions. It has been suggested that such affective states can be usefully described in terms of broad emotion dimensions (Russell & Feldman Barrett, 1999). Such affective states have also been suggested to play an important role in everyday vocal expression (Cowie & Cornelius, 2003). Thus, a dimensional approach can make a contribution to the understanding of how the voice offers subtle cues to affective states in everyday life.

6.2 Explaining inconsistencies in code usage

In this thesis, several limitations of previous studies, which may have contributed to the noted difficulties in finding distinct patterns of voice cues that differentiate between discrete emotions, were pointed out. The results showed that if these limitations are attended to, the problem of inconsistency in cue usage can be alleviated. Especially the results of Study II imply that it is very important to take the intensity of emotion into account, since it greatly affects the acoustic characteristics of vocal expressions.

However, some inconsistency still remains and needs to be explained. This explanation can be sought in terms of the coding of the communicative process.[13] As can be seen from the results of Studies I and II, the same voice cues are often used in the same way in order to convey more than one emotion. For instance, F_0 rises and speech rate increases for both anger and fear. Therefore F_0 and speech rate are not perfect indicators of respective emotion. Any voice cue taken alone is not sufficient for communicating emotion expressions, but what is needed are combinations of several cues. The acoustic voice cues that are involved in expression of emotions thus seem to be partly redundant, and the relevant cues are coded probabilistically (i.e., the cues are not perfectly reliable indicators of the expressed emotion; see Juslin & Scherer, in press).[14]

The redundancy between the cues largely reflects the sound production mechanisms of the voice. For instance, an increase in subglottal pressure increases not only the intensity of the voice, but also F_0 to some degree (e.g., Borden et al., 1994). The probabilistical nature of the cues reflects individual differences among encoders and structural constraints of the verbal material, as well as the fact that the same cue can be used in the same way in more than one expression. The redundancy and probabilistical nature of the voice cues entail that decoders have to combine many cues for successful communication to occur. This is not simply a matter of pattern matching, however, because the cues contribute in an additive fashion to listeners' judgments (e.g., Ladd, Silverman, Tolkmitt, Bergmann, & Scherer, 1985; Scherer & Oshinsky, 1977). Each cue is neither necessary nor sufficient, but the larger the number of cues used, the more reliable the communication becomes (Juslin, 2000). Obviously this emphasizes the importance of considering a large number of voice cues in studies of vocal expression.

The nature of the coding described above further has important implications. Because the voice cues are intercorrelated to some degree, more than one way of using the cues might lead to a similarly high level of decoding accuracy (e.g., Dawes & Corrigan, 1974; Juslin, 2000). This might explain why accurate communication is regularly found in studies of vocal expres-

[13] It has been suggested that an adapted version of Brunswik's (1956) lens model could provide a useful way of conceptualizing the relations between the encoder, the message, and the decoder (Juslin, 2000; Scherer, 1982).

[14] This is also evident from studies that have manipulated the voice stimuli so that some voice cues are altered while others are left unaltered. Such studies have found that listener's can still recognize the expressions from the manipulated stimuli (e.g., Alpert, Kurtzberg, & Friedhoff, 1963; Bergmann, Goldbeck, & Scherer, 1988; Friend & Farrar, 1994; Starkweather, 1956).

sion, despite considerable inconsistency in code usage (see Study I). Multiple cues that are partly redundant yield a robust communicative system that is forgiving of deviations from optimal code usage. However, this robustness comes with a price. The redundancy of the cues means that the same information is conveyed by many cues, which limits the complexity of the information that can be conveyed (Shannon & Weaver, 1949).

It has indeed been reported that actors are able to communicate broad emotion categories (e.g., basic emotions), but not finer nuances within these categories (e.g., Greasley et al., 2000; Kaiser, 1962). This could reflect a limit on how fine nuances can be reliably communicated. As argued in Study I, seen from an evolutionary point of view it is ultimately more important to avoid making serious mistakes, like mistaking anger for sadness, than to be able to make subtle discriminations like detecting different kinds of anger. The communication of basic emotions would have been supported by evolutionary mechanisms since their function is to deal with issues of life and death. This approach does not, however, deny that also some more subtle affective states could be reliably conveyed by the voice. It only states that it is more likely that basic emotions, rather than more subtle states, will have specific acoustic patterns. It is possible that *some* more subtle states can also be reliably communicated. If so, it is also likely that pull effects are more prominent in such cases.

Of course, evolution did not associate vocal expressions with emotional states in a random fashion, but the mapping of expressions to emotions has likely evolved in tandem with the need to communicate emotional states efficiently concerning both production and perception of vocal expressions (e.g., Dailey, Cottrell, Padgett, & Adolphs, 2002). Seen from this perspective, the redundancy of the coding has further beneficial consequences. For instance, it helps to counteract the degradation of acoustic signals during transmission that occur in natural environments due to factors such as attenuation and reverberation (Wiley & Richards, 1978).

6.3 Limitations and methodological issues

There are several limitations in the present material that deserve to be mentioned. First, the empirical results of the thesis are based on a relatively small database of vocal expressions, which may limit the generalizability of the obtained results.

With regard to this point, it should be noted that though the database in Studies II and III consisted of only 8 speakers, 2 languages, and 2 sentences, it is still a larger than average sample. It can also be noted that the results of Study I are based on an exhaustive review of the vocal expression literature, and thus are based on a quite large database. The issue of idiosyncratic effects of individual speakers may be most pressing considering the results of

Study IV, since these were based on only one speaker. In Study IV, this practice was necessitated by the time-consuming creation of the synthesized speech stimuli, and the pioneering nature of the study. However, the acoustic characteristics of the voice stimuli in question were in concordance with the average code usage reported in the literature (e.g., as detailed in Table 8), so there is little reason to suspect that the results were due to idiosyncrasies of the particular speaker. However, given the frequent appearance of individual differences in vocal expression, this brings attention to the need for replication using different speakers, languages, and verbal material.

Concerning Study IV, it should further be noted that speech synthesis has only recently reached a level where it is possible to create morphed continua of vocal expressions. Also, the relationships between the psychological dimensions of pitch and loudness, and the acoustic parameters of frequency and intensity are complicated (e.g., Zwicker & Fastl, 1999). Especially the psychological dimension of voice quality is not fully understood acoustically (e.g., Gerratt & Kreiman, 2001; Laver, 1980). This renders the morphing of multi-dimensional acoustic stimuli, like speech sounds, a challenging task. The perceptual organization of vocal expressions has barely begun to be addressed, and studies on this subject are an important concern for future research.

Another limitation is that the empirical studies, and a large part of the reviewed studies in Study I, were conducted on posed vocal expressions. The degree to which emotion portrayals are similar to naturally occurring expressions is an important concern (e.g., Bachorowski, 1999). It is generally assumed that emotion portrayals need to be similar to naturally occurring expressions in order to be effective (e.g., Davitz, 1964c; Owren & Bachorowski, 2001; Banse & Scherer, 1996). Nevertheless, posed expressions may be influenced by conventionalized stereotypes of vocal expression and may also yield expressions that are more intense and prototypical than naturally occurring expressions (Scherer, 1986).

Unfortunately, the number of studies that have used natural expressions is too small to allow for a comparison, and the data of the existing studies is often presented in ways that make a direct comparison difficult. Thus, the actual extent to which posed expression is similar to natural expression is an empirical question that requires further research. Besides, control of the verbal material and the intention of the encoder can often be problematic in studies using naturally occurring emotion expressions. To conduct more ecologically valid studies of vocal expression without sacrificing internal validity clearly represents a challenge for future research.

It can be argued that the choice of methodology should be informed by the research questions that one wishes to address. For instance, if one wishes to study effects of weaker affective states on voice production, mood-induction methods, or recordings of real conversations, combined with dimensional listening tests could be a successful combination. If, on the other hand, one wishes to study discrete emotions of a fairly strong intensity, one may be forced to rely on emotion portrayals to a large extent. However, to conduct vocal expression studies where relatively intense discrete emotions are induced in controlled laboratory settings is an important undertaking for future research. For some recent attempts, see Laukka, Åhs, Furmark, Michelgård, and Fredrikson (2004), and T. M. Scherer (2000).

Finally, it should be noted that in everyday life, vocal expressions usually involve a combination of both “natural” and posed expressions (i.e., consist of both push and pull effects; Scherer, 1989). Also, intentionally portraying vocal expressions appears to produce emotional effects in the speaker through the process of emotional contagion (e.g., Hatfield, Hsee, Costello, Weisman, & Denney, 1995; Siegman, Anderson, & Berger, 1990). Thus the distinction between natural and posed expression is not always as clear-cut as it is sometimes made out to be.

6.4 Implications and future research

The results of this thesis have several implications for future research on vocal expressions. Firstly, it is suggested that it will be of paramount importance to take emotion intensity into account in future studies, since the intensity of emotion has a great impact on emotion-relevant voice cues. Secondly, the nature of the coding of vocal expressions (i.e., that voice cues are probabilistic and partly redundant) renders it necessary to measure many aspects of the speech signal (i.e., many voice cues) if one wishes to find acoustic differentiation of emotional states. Thirdly, Study I revealed many gaps regarding the use of specific voice cues in the vocal expression literature (see Table 3). Since these gaps point out where knowledge about code usage is missing, they can be used to guide future research efforts. Fourthly, the results regarding specific cue patterns for discrete emotions and emotion dimensions (as summarized in Table 8) could be subjected to direct tests in listening experiments using synthesized and systematically varied speech stimuli.

However, the results of this study are only one step towards understanding the nature of vocal expression. Below, various further issues that need to be considered to achieve a fuller understanding of the complexities of vocal expression are outlined.

One key issue is what sort of emotional states (e.g., discrete emotions, moods, attitudes) are most commonly expressed in everyday vocal expression. It has been suggested that milder affective states are particularly prominent in everyday speech (Cowie & Cornelius, 2003), but evidence supporting this claim is at the moment largely lacking. To achieve a more thorough understanding of vocal expression in everyday speech, more studies need to be conducted in naturalistic settings.

Another pressing issue is what acoustic measures should be utilized. To date, acoustic measures that are averaged over time are most commonly utilized. However, in order to be able to capture the dynamic nature of speech, more detailed analyses of local features must be conducted. This could be done, for instance, by using continuous measurements of listener responses (e.g., Cowie et al., 2000), which can then be coupled with time-linked measurements of voice cues. Similarly, it will also become increasingly important to try to predict more specific aspects of vocal changes, for instance on the basis of appraisal results, as suggested by Scherer (2003). In addition, more research needs to be directed to the spectral parameters of speech to clarify the relationships between voice quality, voice production, and acoustic measurements. All hypotheses derived from empirical studies should also be verified experimentally using speech synthesis techniques where cues are manipulated in a systematic fashion.

It has been reported that expression of emotions can be mediated by personality characteristics (Feldman Barrett & Niedenthal, 2004; Gross, John, & Richards, 2000). This entails that some people are more expressive than others, and may be more emotional than others; something that has been often reported in studies of vocal expression (see Feldman Barrett & Gross, 2001). Future studies should take the issue of individual differences seriously and, besides using large numbers of encoders and decoders, also systematically study the reasons for these differences.

In vocal expression, there is a constant interaction between the nonverbal and verbal (linguistic) aspects of communication (e.g., Buck & VanLear, 2002). There is preliminary evidence that the context in which vocal expressions are heard affects the interpretation of their content (e.g., Cauldwell, 2000). Future research should pay attention to the interaction between verbal and nonverbal aspects, and also take the context into account.

There is also a need for more cross-cultural research, especially on the possible universality of code usage. It would be worthwhile to make more fine-grained comparisons of perception of vocal expressions, taking into account the relations between the particular cultures (e.g., Elfenbein & Ambady, 2003).

The face and the voice together constitute the most effective means of communication of emotions (e.g., de Gelder, 2000), but more research is needed on the interaction of these two modalities. Interesting future topics thus include the bimodal expression of emotion in the face and voice (de Gelder & Vroomen, 2000; Massaro & Egan, 1996). This issue has implications for the question of whether there exists a general processor for the perception of emotional content across different modalities (e.g., Borod et al., 2000). A related topic in need of more research is the influence of facial expressions on the acoustics of vocal expression (e.g., Aubergé & Cathiard, 2003; Tartter, 1980; Tartter & Braun, 1994).

Finally, studies on the neural bases of production and perception of emotional speech is an important way forward in this area (e.g., Adolphs, Damasio, & Tranel, 2002; Buchanan et al., 2000; Gandour et al., 2003; George et al., 1996; Wildgruber et al., 2002). Especially work on the production of vocal expressions is today largely missing. Knowledge about the underlying brain mechanisms responsible for perception and production of vocal expressions will be crucial for definitively answering the question whether vocal expressions are conveyed as discrete emotions or emotion dimensions (or perhaps a little of both, depending on the situation and affective state).

6.5 Possible applications

The focus of this thesis has been on the basic research implications of research on vocal expression, and its relations to psychological theory. However, knowledge of vocal expression of emotions is important for many questions of a more applied nature.

Studies of vocal expressions have implications for human-computer interaction. For instance adding expressiveness (including emotional expression) to synthesized speech is an important concern for researchers who wish to make synthetic speech more natural and acceptable to users in various applications (e.g., Campbell, 2004; Tatham & Morton, 2004). Also, in automatic recognition of speech it is important to be aware of the impact of vocal expressions on the acoustic signal (Schröder, 2001; for an interesting application, see Slaney & McRoberts, 2003).

Further, the ability to decode vocal expressions can affect the quality of close relationships (Koerner & Fitzpatrick, 2002; see also Ekman, 2003). Proficiency of emotional communication, including vocal expressions, is an aspect of "emotional intelligence" (Salovey & Meyer, 1990). Knowledge of how the voice conveys emotion could for instance be used in empathy training, or in the development of training programs for emotion regulation. Vocal expression also has further implications for social relationships since it is an important component of early infant-caregiver interaction. It has indeed been suggested that what foremost marks out infant-directed speech, is its emotional expressiveness (Trainor, Austin, & Desjardins, 2000). Also, how we express our emotions vocally has an impact on health-related physiology and emotion regulation (Siegman et al., 1990; Siegman & Boyle, 1993). Thus research on vocal expression has wide implications on issues of importance to our health and well-being (see also Booth & Pennebaker, 2000; Giese-Davis & Spiegel, 2003).

Finally, continua of facial expressions have found several interesting applications in applied psychological research, for instance in studies on mood bias in emotion recognition (Richards et al., 2002), and in studies of pathological (e.g., Teunisse & de Gelder, 2001; Calder, Keane, Lawrence, &

Manes, 2004) and developmental (e.g., Pollak & Kistler, 2002) aspects of emotion perception. It is believed that continua of vocal expressions (like the ones developed in Study IV) likewise can be a useful research tool, and find many applications for research on emotion perception, as well as for psychological research on emotions in general.

6.6 Concluding remarks

The results of this thesis suggest that a discrete-emotions framework provides the best account of vocal expression, as implicated by the combined weight of the following findings: (a) Vocal expressions of discrete emotions are universally recognized, (b) distinct patterns of voice cues correspond to discrete emotions, and (c) vocal expressions are perceived as discrete emotion categories, and not as broad emotion dimensions.

In the light of these results, it may now be time to start looking beyond the simple, though admittedly important, question about whether discrete emotions can or cannot be conveyed by vocal expressions. To find evidence of emotion-specific patterns of voice cues is only a first step toward the understanding of vocal emotion expression. In order for further progress to be made, research on vocal expression now needs to start asking more subtle questions. What particular changes in each emotion-component (e.g., physiology, appraisal) produce what particular effects on the resulting vocal expression? What are the relations between push and pull effects on vocal expression, or between posed and naturally occurring expressions?

7. Acknowledgments

Many people deserve credit for making this thesis possible. Two people without whom this work most certainly would not have seen the light of day are Alf Gabrielsson – for getting me started with psychological research in the first place – and, most importantly, my supervisor and friend Patrik Juslin – for guiding me through many pitfalls of research. Patrik is a man with many talents, not the least significant being his ability to provide well-deserved criticisms in a constructive manner. Throughout my work, he has always encouraged and supported me beyond the call of duty.

I have had the good fortune to be employed by two research projects during my time at the department. Many thanks go to the *Expressive performance* and *Feel-ME* groups for great collaboration: Roberto Bresin, Marie Djerf, Gertrud Ericson, Anders Friberg, Jessika Karlsson, Ingrid Lagerlöf, Erik Lindström, Guy Madison, Maria Sandgren, and Erwin Schoonderwaldt.

I would also like to thank the following people in no particular order: – Ulf Dimberg and Gunilla Bohlin for providing valuable comments on preliminary versions of this thesis. – All colleagues, staff and students that I have had the opportunity to work and socialize with during the years. Ingen nämnd, ingen glömd. – The librarians Siv Vedung and Hans Åhlén for great assistance when I was searching for those obscure articles from decades ago. – Many reviewers and editors who have shared their knowledge and commented on the studies of this thesis; Beatrice de Gelder, Stevan Harnad, Bruno Repp, Craig A. Smith, Klaus Scherer (and of course all the anonymous ones). – All the people who have participated in my experiments. It really would not have been possible without you!

Warm regards to my friends and musical collaborators during the years; especially Rundfunk, SSOP, and Jonas A. Without you I would have had far too much time to spend on research … I'm sure that would not have been good for me!

Haluan kiittää vanhempiani Terttua ja Maunoa, jotka ovat aina kannustaneet sen pariin, joka tuntuu itsestä kiinnostavalta. Many thanks to my brother Raul for being a kindred spirit.

Last, but at the top of the list, I would like to thank my loving wife Erika (you're smashing!). I really owe everything to you.

Uppsala, October 2004
Petri Laukka

8. References

Adolphs, R., Damasio, H., & Tranel, D. (2002). Neural systems for recognition of emotional prosody: A 3-D lesion study. *Emotion, 2,* 23-51.

Airas, M., & Alku, P. (2004). Emotions in short vowel segments: Effects of the glottal flow as reflected by the normalized amplitude quotient. *Lecture Notes in Artificial Intelligence, 3068,* 13-24.

Albas, D. C., McCluskey, K. W., & Albas, C. A. (1976). Perception of the emotional content of speech: A comparison of two Canadian groups. *Journal of Cross-Cultural Psychology, 7,* 481-490.

Alpert, M., Kurtzberg, R. L., & Friedhoff, A. J. (1963). Transient voice changes associated with emotional stimuli. *Archives of General Psychiatry, 8,* 362-365.

Aubergé, V., & Cathiard, M. (2003). Can we hear the prosody of smile? *Speech Communication, 40,* 87-97.

Averill, J. R. (1982). *Anger and aggression: An essay on emotion.* New York: Springer.

Bachorowski, J.-A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science, 8,* 53-57.

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

Bachorowski, J.-A., & Owren, M. J. (2003). Sounds of emotion: Production and perception of affect-related vocal acoustics. *Annals of the New York Academy of Sciences, 1000,* 244-265.

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

Bänziger, T., Grandjean, D., Bernard, P.-J., Klasmeyer, G., & Scherer, K. R. (2001). Prosodie de l'émotion: Etude de l'encodage et du décodage [Prosody of emotion: Study of encoding and decoding]. *Cahiers de Linguistique Française, 23,* 11-37.

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

Baum, K. M., & Nowicki, S., Jr. (1998). Perception of emotion: Measuring decoding accuracy of adult prosodic cues varying in intensity. *Journal of Nonverbal Behavior, 22,* 89-107.

Bergmann, G., Goldbeck, T., & Scherer, K. R. (1988). Emotionale Eindruckswirkung von prosodischen Sprechmerkmalen [Emotional inference from prosodic features of speech]. *Zeitschrift für Experimentelle und Angewandte Psychologie, 35,* 167-200.

Boersma, P., & Weenink, D. (1999). Praat (Computer software). Amsterdam, The Netherlands: Institute of Phonetic Sciences, University of Amsterdam.

Bonanno, G. A., & Keltner, D. (2004). The coherence of emotion systems: Comparing "on-line" measures of appraisal and facial expressions, and self-report. *Cognition & Emotion, 18,* 431-444.

Bonner, M. R. (1943). Changes in the speech pattern under emotional tension. *American Journal of Psychology, 56,* 262-273.

Booth, R. J., & Pennebaker, J. W. (2000). Emotions and immunity. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 558-570). New York: Guilford Press.

Borden, G. J., Harris, K. S., & Raphael, L. J. (1994). *Speech science primer: Physiology, acoustics and perception of speech* (3rd ed.). Baltimore: Williams & Wilkins.

Borod, J. C., Pick, L. H., Hall, S., Sliwinski, M., Madigan, N., Obler, L. K., Welkowitz, J., Canino, E., Erhan, H. M., Goral, M., Morrison, C., & Tabert, M. (2000). Relationships among facial, prosodic, and lexical channels of emotional perceptual processing. *Cognition & Emotion, 14,* 193-211.

Brehm, J. W. (1999). The intensity of emotion. *Personality and Social Psychology Review, 3,* 2-22.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments.* Berkeley: University of California Press.

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., & Jäncke, L. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Cognitive Brain Research, 9,* 227-238.

Buck, R. (1984). *The communication of emotion.* New York: Guilford Press.

Buck, R., & VanLear, C. A. (2002). Verbal and nonverbal communication: Distinguishing symbolic, spontaneous, and pseudo-spontaneous nonverbal behavior. *Journal of Communication, 52,* 522-541.

Burkhardt, F. (2001). *Simulation emotionaler Sprechweise mit Sprachsyntheseverfahren* [Simulation of emotional speech by means of speech synthesis]. Doctoral dissertation, Technische Universität Berlin, Berlin, Germany.

Buss, D. M. (1995). Evolutionary psychology: A new paradigm for psychological science. *Psychological Inquiry, 6,* 1-30.

Buss, D. M., & Kenrick, D. T. (1998). Evolutionary social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, pp. 982-1026). New York: McGraw-Hill.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotions. In M. Lewis & J. M.

Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 173-191). New York: Guilford Press.

Cahn, J. E. (1990). The generation of affect in synthesized speech. *Journal of the American Voice I/O Society, 8,* 1-19.

Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition, 3,* 81-117.

Calder, A. J., Keane, J., Lawrence, A. D., & Manes, F. (2004). Impaired recognition of anger following damage to the ventral striatum. *Brain, 127,* 1958-1969.

Campanella, S., Quinet, P., Bruyer, R., Crommelinck, M., & Guerit, J.-M. (2002). Categorical perception of happiness and fear facial expressions: An ERP study. *Journal of Cognitive Neuroscience, 14,* 210-227.

Campbell, N. (2004). Specifying affect and emotion for expressive speech synthesis. *Lecture Notes in Computer Sciences, 2945,* 395-406.

Carver, C. S. (2001). Affect and the functional bases of behavior: On the dimensional structure of affective experience. *Personality and Social Psychology Review, 5,* 345-356.

Cauldwell, R. T. (2000). Where did the anger go? The role of context in interpreting emotion in speech. In R. Cowie, E. Douglas-Cowie, & M. Schröder (Eds.), *Proceedings of the ISCA workshop on speech and emotion* [CD-ROM]. Belfast, Northern Ireland: International Speech Communication Association.

Christie, I. C., & Friedman, B. H. (2004). Autonomic specificity of discrete emotion and dimensions of affective space: A multivariate approach. *International Journal of Psychophysiology, 51,* 143-153.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 91-115). New York: Guilford Press.

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

Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., & Schröder, M. (2000). Feeltrace: An instrument for recording perceived emotion in real time. In R. Cowie, E. Douglas-Cowie, & M. Schröder (Eds.), *Proceedings of the ISCA workshop on speech and emotion* [CD-ROM]. Belfast, Northern Ireland: International Speech Communication Association.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine, 18(1),* 32-80.

Dailey, M., N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). EMPATH: A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience, 14,* 1158-1173.

Darwin, C. (1998). *The expression of the emotions in man and animals* (with introduction, afterword, and commentaries by P. Ekman). New York: Oxford University Press. (Original work published 1872).

Davitz, J. R. (1964a). Auditory correlates of vocal expressions of emotional meanings. In J. R. Davitz (Ed.), *The communication of emotional meaning* (pp. 101-112). New York: McGraw-Hill.

Davitz, J. R. (1964b). Personality, perceptual, and cognitive correlates of emotional sensitivity. In J. R. Davitz (Ed.), *The communication of emotional meaning* (pp. 57-68). New York: McGraw-Hill.

Davitz, J. R. (1964c). A review of research concerned with facial and vocal expressions of emotion. In J. R. Davitz (Ed.), *The communication of emotional meaning* (pp. 13-29). New York: McGraw-Hill.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81,* 95-106.

de Gelder, B. (2000). Recognizing emotions by ear and by eye. In R. D. Lane & L. Nadel (Eds.), *Cognitive neuroscience of emotion* (pp. 84-105). New York: Oxford University Press.

de Gelder, B., Teunisse, J. P., & Benson, P. J. (1997). Categorical perception of facial expressions and their internal structure. *Cognition & Emotion, 11,* 1-23.

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion, 14,* 289-311.

Eibl-Eibesfeldt, I. (1973). The expressive behavior of the deaf-and-blind born. In M. von Cranach & I. Vine (Eds.), *Social communication and movement: Studies of interaction and expression in man and chimpanzee* (pp. 163-193). New York: Academic Press.

Ekman, P. (1972). Universals and cultural differences in facial expression of emotion. In J. Cole (Ed.), *Nebraska symposium on motivation* (Vol. 19, pp. 207-283). Lincoln, NE: University of Nebraska Press.

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

Ekman, P. (1993). Facial expressions and emotion. *American Psychologist, 48,* 384-392.

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin, 115,* 268-287.

Ekman, P. (2003). *Emotions revealed. Recognizing faces and feelings to improve communication and emotional life.* New York: Henry Holt.

Ekman, P. Friesen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology, 39,* 1125-1134.

Eldred, S. H., & Price, D. B. (1958). A linguistic evaluation of feeling states in psychotherapy. *Psychiatry, 21,* 115-121.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128,* 203-235.

Elfenbein, H. A., & Ambady, N. (2003). Cultural similarity's consequences. A distance perspective on cross-cultural differences in emotion recognition. *Journal of Cross-Cultural Psychology, 34,* 92-110.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44,* 227-240.

Fairbanks, G., & Hoaglin, L. W. (1941). An experimental study of the durational characteristics of the voice during the expression of emotion. *Speech Monographs, 8,* 85-90.

Fairbanks, G., & Provonost, W. (1939). An experimental study of the pitch characteristics of the voice during the expression of emotion. *Speech Monographs, 6,* 87-104.

Fant, G. (1960). *Acoustic theory of speech production.* ‘s-Gravenhage, The Netherlands: Mouton.

Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General, 113,* 464-486.

Feldman Barrett, L., & Gross, J. J. (2001). Emotional intelligence: A process model of emotion representation and regulation. In T. J. Mayne & G. A. Bonnano (Eds.), *Emotions: Current issues and future directions* (pp. 286-310). New York: Guilford Press.

Feldman Barrett, L., & Niedenthal, P. M. (2004). Valence focus and the perception of facial affect. *Emotion, 4,* 266-274.

Fónagy, I. (1976). La mimique buccale. Aspect radiologique de la vive voix [Radiological aspects of emotive speech]. *Phonetica, 33,* 31-44.

Fónagy, I., & Magdics, K. (1963). Emotional patterns in intonation and music. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, 16,* 293-326.

Fox, N. A., & Davidson, R. J. (1988). Patterns of brain electrical activity during facial signs of emotion in 10-month old infants. *Developmental Psychology, 24,* 230-236.

Frank, M. G., & Stennet, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology, 80,* 75-85.

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin, 97,* 412-429.

Friend, M., & Farrar, M. J. (1994). A comparison of content-masking procedures for obtaining judgments of discrete affective states. *Journal of the Acoustical Society of America, 96,* 1283-1290.

Frijda, N. H. (1993). Moods, emotion episodes, and emotions. In M. Lewis, & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 381-403). New York: Guilford Press.

Frijda, N. H., Ortony, A., Sonnemans, J., & Clore, G. L. (1992). The complexity of intensity. Issues concerning the structure of emotion intensity. In M. S. Clark (Ed.), *Review of personality and social psychology* (Vol. 13, pp. 60-89). Newbury Park, CA: Sage.

Gandour, J., Wong, D., Dzemidzic, M., Lowe, M., Tong, Y., & Li, X. (2003). A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. *Human Brain Mapping, 18,* 149-157.

Gendrot, C. (2003). Rôle de la qualité de la voix dans la simulation des émotions: Une étude perceptive et physiologique [The role of voice quality in the simulation of emotions: A study of perception and physiology]. *Revue PArole, 27,* 137-158.

George, M. S., Parekh, P. I., Rosinsky, N., Ketter, T. A., Kimbrell, T. A., Heilman, K. M., Herscovitch, P., & Post, R. M. (1996). Understanding emotional prosody activates right hemisphere regions. *Archives of Neurology, 53,* 665-670.

Gerratt, B. R., & Kreiman, J. (2001). Measuring vocal quality with speech synthesis. *Journal of the Acoustical Society of America, 110,* 2560-2566.

Giese-Davis, J., & Spiegel, D. (2003). Emotional expression and cancer progression. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 1053-1082). New York: Oxford University Press.

Gobl, C., & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication, 40,* 189-212.

Goodenough, F. L. (1932). Expressions of emotions in a blind-deaf child. *Journal of Abnormal and Social Psychology, 27,* 328-333.

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

Gross, J. J., John, O. P., & Richards, J. M. (2000). The dissociation of emotion expression from emotion experience: A personality perspective. *Personality and Social Psychology Bulletin, 26,* 712-726.

Harnad, S. (Ed.). (1987). *Categorical perception. The groundwork of cognition.* New York: Cambridge University Press.

Hatfield, E., Hsee, C. K., Costello, J., Weisman, M. S., & Denney, C. (1995). The impact of vocal feedback on emotional experience and expression. *Journal of Social Behavior and Personality, 10,* 293-312.

Herrald, M. M., & Tomaka, J. (2002). Patterns of emotion-specific appraisal, coping, and cardiovascular reactivity during an ongoing emotional episode. *Journal of Personality and Social Psychology, 83,* 434-450.

Hozjan, V., & Kačič, Z. (2003). Context-independent multilingual emotion recognition from speech signals. *International Journal of Speech Technology, 6,* 311-320.

Huttar, G. L. (1968). Relations between prosodic variables and emotions in normal American English utterances. *Journal of Speech and Hearing Research, 11,* 481-487.

Isserlin, M. (1925). Psychologisch-phonetische Untersuchungen. II. Mitteilung [Psychological-phonetic studies. Second communication]. *Zeitschrift für die Gesamte Neurologie und Psychiatrie, 94,* 437-448.

Izard, C. E. (1977). *Human emotions.* New York: Plenum.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review, 99,* 561-565.

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 220-235). New York: Guilford Press.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1797-1813.

Juslin, P. N., & Laukka, P. (2000). Improving emotional communication in music performance through cognitive feedback. *Musicae Scientiae, 4,* 151-183.

Juslin, P. N., & Laukka, P. (2003). Emotional expression in speech and music: Evidence of cross-modal similarities. In P. Ekman, J. J. Campos, R. J. Davidson, & F. B. M. de Waal (Eds.), *Emotions inside out: 130 years after Darwin's The Expression of the Emotions in Man and Animals* (pp. 279-282). New York: New York Academy of Sciences.

Juslin, P. N., & Laukka, P. (2004). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. In A. Kappas (Ed.), *Proceedings of the XIth Conference of the International Society for Research on Emotions* (pp. 278-281). Amsterdam, The Netherlands: ISRE Publications.

Juslin, P. N., & Laukka, P. (in press). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research.*

Juslin, P. N., & Scherer, K. R. (in press). Vocal expression of affect. In J. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *Handbook of nonverbal behavior research methods in the affective sciences.* New York: Oxford University Press.

Kaiser, L. (1962). Communication of affects by single vowels. *Synthese, 14,* 300-319.

Keltner, D., Ekman, P., Gonzaga, G., & Beer, J. (2003). Facial expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 415 – 429). New York: Oxford University Press.

Keltner, D., & Gross, J. J. (1999). Functional accounts of emotions. *Cognition & Emotion, 13,* 465-466.

Keltner, D., & Kring, A. M. (1998). Emotion, social function, and psychopathology. *Review of General Psychology, 2,* 320-342.

Kienast, M., & Sendlmeier, W. F. (2000). Acoustical analysis of spectral and temporal changes in emotional speech. In R. Cowie, E. Douglas-Cowie, & M. Schröder (Eds.), *Proceedings of the ISCA workshop on speech and emotion* [CD-ROM]. Belfast, Northern Ireland: International Speech Communication Association.

Klasmeyer, G., & Sendlmeier, W. F. (1997). The classification of different phonation types in emotional and neutral speech. *Forensic Linguistics, 1,* 104-124.

Koerner, A.. F., & Fitzpatrick, M. A. (2002). Nonverbal communication and marital adjustment and satisfaction: The role of decoding relationship relevant and relationship irrelevant affect. *Communication Monographs, 69,* 33-51.

Kramer, E. (1963). Judgment of personality characteristics and emotions from nonverbal properties of speech. *Psychological Bulletin, 60,* 408-420.

Kramer, E. (1964). Elimination of verbal cues in judgments of emotion from voice. *Journal of Abnormal and Social Psychology, 68,* 390-396.

Ladd, D. R., Silverman, K. E. A., Tolkmitt, F., Bergmann, G., & Scherer, K. R. (1985). Evidence of independent function of intonation contour type, voice quality, and F_0 range in signaling speaker affect. *Journal of the Acoustical Society of America, 78,* 435-444.

Ladefoged, P. (2000). *A course in phonetics* (4th ed.). New York: Harcourt Brace.

Lakshminarayanan, K., Shalom, D. B., van Wassenhove, V., Orbelo, D., Houde, J., & Poeppel, D. (2003). The effect of spectral manipulations on the identification of affective and linguistic prosody. *Brain and Language, 84,* 250-263.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology. *Biological Psychiatry, 44,* 1248-1263.

Larsen, R. J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. In M. S. Clark (Ed.), *Review of personality and social psychology* (Vol. 13, pp. 25-59). Newbury Park, CA: Sage.

Laukka, P. (2003). Categorical perception of emotion in vocal expression. *Annals of the New York Academy of Sciences, 1000,* 283-287.

Laukka, P. (2004). Instrumental music teachers' views on expressivity: A report from music conservatoires. *Music Education Research, 6,* 45-56.

Laukka, P., Åhs, F., Furmark, T., Michelgård, Å., & Fredrikson, M. (2004). *The voice of anxiety: Acoustical correlates of perceived anxiety and distress in patients with social phobia before and after treatment.* Manuscript in preparation.

Laukka, P., & Gabrielsson, A. (2000). Emotional expression in drumming performance. *Psychology of Music, 28*, 181-189.

Laukka, P., & Juslin, P. N. (2002). *Accuracy of communication of emotions in speech and music performance: A quantitative review.* Paper presented at the 7th International Conference on Music Perception and Cognition, Sydney, Australia.

Laukkanen, A.-M., Vilkman, E., Alku, P., & Oksanen, H. (1996). Physical variations related to stress and emotional state: A preliminary study. *Journal of Phonetics, 24,* 313-335.

Laukkanen, A.-M., Vilkman, E., Alku, P., & Oksanen, H. (1997). On the perception of emotions in speech: The role of voice quality. *Logopedics Phoniatrics Vocology, 22,* 157-168.

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

Lazarus, R. S. (1991). *Emotion and adaptation.* New York: Oxford University Press.

Lazarus, R. S., & Smith, C. A. (1988). Knowledge and appraisal in the cognition-emotion relationship. *Cognition & Emotion, 2,* 281-300.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23,* 155-184.

Leinonen, L., Hiltunen, T., Linnankoski, I., & Laakso, M.-L. (1997). Expression of emotional-motivational connotations with a one-word utterance. *Journal of the Acoustical Society of America, 102,* 1853-1863.

Levenson, R. W. (1992). Autonomic nervous system differences among emotions. *Psychological Science, 3,* 23-27.

Levenson, R. W. (1994). Human emotion: A functional view. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 123-126). New York: Oxford University Press.

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology, 27,* 363-384.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54,* 358-368.

Lieberman, P. (1961). Perturbations in vocal pitch. *Journal of the Acoustical Society of America, 33,* 597-603.

Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. (1977). The psychophysics of categorical perception. *Psychological Review, 84,* 452-471.

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review, 3,* 215-221.

Millot, J.-L., & Brand, G. (2001). Effects of pleasant and unpleasant ambient odors on human voice pitch. *Neuroscience Letters, 297,* 61-63.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9,* 453-467.

Mozziconacci, S. J. L. (1998). *Speech variability and emotion: Production and perception.* Eindhoven, the Netherlands: Technische Universiteit Eindhoven.

Murphy, F. C., Nimmo-Smith, I., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: A meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience, 3,* 207-233.

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 Acoustical Society of America, 93,* 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.

Nakamichi, A., Jogan, A., Usami, M., & Erickson, D. (2003). Perception by native and non-native listeners of vocal emotion in a bilingual movie. *Bulletin of Gifu City Women's College, 52,* 87-91.

Nyklíček, I., Thayer, J. F., & van Doornen, L. J. P. (1997). Cardiorespiratory differentiation of musically-induced emotions. *Journal of Psychophysiology, 11,* 304-321.

Oatley, K., & Jenkins, J. M. (1996). *Understanding emotions.* Oxford, England: Blackwell.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108,* 483-522.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana, IL: University of Illinois Press.

Oudeyer, P.-Y. (2003). The production and recognition of emotions in speech: Features and algorithms. *International Journal of Human-Computer Studies, 59,* 157-183.

Owren, M. J., & Bachorowski, J.-A. (2001). The evolution of emotional expression: A "selfish-gene" account of smiling and laughter in early hominids and humans. In T. J. Mayne & G. A. Bonanno (Eds.), *Emotions: Current issues and future directions* (pp. 152-191). New York: Guilford Press.

Pakosz, M. (1983). Attitudinal judgments in intonation: Some evidence for a theory. *Journal of Psycholinguistic Research, 12,* 311-326.

Panksepp, J. (2000). Emotions as natural kinds within the mammalian brain. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 137-156). New York: Guilford Press.

Papoušek, H., Jürgens, U., & Papoušek, M. (Eds.). (1992). *Nonverbal vocal communication: Comparative and developmental approaches.* Cambridge, England: Cambridge University Press.

Pereira, C. (2000). Dimensions of emotional meaning in speech. In R. Cowie, E. Douglas-Cowie, & M. Schröder (Eds.), *Proceedings of the ISCA workshop on speech and emotion* [CD-ROM]. Belfast, Northern Ireland: International Speech Communication Association.

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage, 16,* 331-348.

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory & Cognition, 3,* 7-18.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America, 55,* 328-333.

Planalp, S., DeFrancisco, V. L., & Rutherford, D. (1996). Varieties of cues to emotion in naturally occurring situations. *Cognition & Emotion, 10,* 137-153.

Plutchik, R. (1994). *The psychology and biology of emotion.* New York: Harper-Collins.

Pollak, S. D., & Kistler, D. J. (2002). Early experience is associated with the development of categorical representations for facial expression of emotion. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 9072-9076.

Power, M., & Dalgleish, T. (1997). *Cognition and emotion: From order to disorder.* Hove, England: Psychology Press.

Repp, B. H. (1984). Categorical perception: Issues, methods and findings. In N. Lass (Ed.), *Speech and language. Advances in basic research and practice* (Vol. 10, pp. 244-335). New York: Academic Press.

Richards, A., French, C. C., Calder, A. J., Webb, B., Fox, R., & Young, A. W. (2002). Anxiety-related bias in the classification of emotionally ambiguous facial expressions. *Emotion, 2,* 273-287.

Rosenthal, R., & Rubin, D. B. (1989). Effect size estimation for one-sample multiple-choice-type data: Design, analysis, and meta-analysis. *Psychological Bulletin, 106,* 332-337.

Rothman, A. D., & Nowicki, S., Jr. (2004). A measure of the ability to identify emotion in children's tone of voice. *Journal of Nonverbal Behavior, 28,* 67-92.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161-1178.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin, 115,* 102-141.

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

Russell, J. A., & Feldman Barrett, L. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology, 76,* 805-819.

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality, 11,* 273-294.

Salovey, P., & Meyer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality, 9,* 185-211.

Scherer, K. R. (1982). Methods of research on vocal communication: Paradigms and parameters. In K. R. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal research* (pp. 136-198). Cambridge, England: Cambridge University Press.

Scherer, K. R. (1985). Vocal affect signalling: A comparative approach. In J. Rosenblatt, C. Beer, M.-C. Busnel, & P. J. B. Slater (Eds.), *Advances in the study of behavior* (Vol. 15, pp. 189-244). New York: Academic Press.

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

Scherer, K. R. (1989). Vocal correlates of emotional arousal and affective disturbance. In H. Wagner, & A. Manstead (Eds.), *Handbook of social psychophysiology* (pp. 165-197). New York: Wiley.

Scherer, K. R. (2000). Psychological models of emotion. In J. Borod (Ed.), *The neuropsychology of emotion* (pp. 137-162). New York: Oxford University Press.

Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92-120). New York: Oxford University Press.

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

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32,* 76-92.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15,* 123-148.

Scherer, K. R., & Ceschi, G. (2000). Criteria for emotion recognition from verbal and nonverbal expression: Studying baggage loss in the airport. *Personality and Social Psychology Bulletin, 26,* 327-339.

Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 433-456). New York: Oxford University Press.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilisation in emotion attribution from auditory stimuli. *Motivation and Emotion, 1,* 331-346.

Scherer, K. R., Schorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal processes in emotion: Theory, methods, research.* New York: Oxford University Press.

Scherer, T. M. (2000). *Stimme, Emotion und Psyche. Untersuchungen zur emotionalen Qualität der menschlichen Stimme* [Voice, emotion, and mind. Studies on the emotional quality of the human voice]. Doctoral dissertation, University of Marburg, Germany.

Schlosberg, H. (1941). A scale for the judgment of facial expressions. *Journal of Experimental Psychology, 29,* 497-510.

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication, 41,* 71-80.

Schröder, M. (2001). Emotional speech synthesis: A review. In *Proceedings of the 7th European Conference on Speech Communication and Technology* (Vol. 1, pp. 561-564). Aalborg, Denmark: International Speech Communication Association.

Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., & Gielen, S. (2001). Acoustic correlates of emotion dimensions in view of speech synthesis. In *Proceedings of the 7th European Conference on Speech Communication and Technology* (Vol. 1, pp. 87-90). Aalborg, Denmark: International Speech Communication Association.

Scott, J. P. (1980). The function of emotions in behavioral systems: A systems theory analysis. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion* (pp. 35-56). New York: Academic Press.

Scripture, E. W. (1921). A study of emotions by speech transcription. *Vox, 31,* 179-183.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication.* Urbana, IL: University of Illinois Press.

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology, 52,* 1061-1086.

Siegman, A. W., Anderson, R. A., & Berger, T. (1990). The angry voice: Its effects on the experience of anger and cardiovascular reactivity. *Psychosomatic Medicine, 52,* 631-643.

Siegman, A. W., & Boyle, S. (1993). Voices of fear and anxiety and sadness and depression: The effects of speech rate and loudness on fear and anxiety and sadness and depression. *Journal of Abnormal Psychology, 102,* 430-437.

Skinner, E. R. (1935). A calibrated recording and analysis of the pitch, force and quality of vocal tones expressing happiness and sadness. *Speech Monographs, 2,* 81-137.

Slaney, M., & McRoberts, G. (2003). BabyEars: A recognition system for affective vocalizations. *Speech Communication, 39,* 367-384.

Smith, C. A. (1989). Dimensions of appraisal and physiological response in emotion. *Journal of Personality and Social Psychology, 56,* 339-353.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48,* 813-838.

Song, M., Chen, C., Bu, J., & You, M. (2004). Speech emotion recognition and intensity estimation. *Lecture Notes in Computer Sciences, 3046,* 406-413.

Sonnemans, J., & Frijda, N. H. (1994). The structure of subjective emotional intensity. *Cognition & Emotion, 8,* 329-350.

Spencer, H. (1857). The origin and function of music. *Fraser's Magazine, 56,* 396-408.

Starkweather, J. A. (1956). Content-free speech as a source of information about the speaker. *Journal of Abnormal and Social Psychology, 52,* 394-402.

Stemmler, G. (1989). The autonomic differentiation of emotions revisited: Convergent and discriminant validation. *Psychophysiology, 26,* 617-632.

Svanfeldt, G., Nordstrand, M., Granström, B., & House, D. (2003). Measurements of articulatory variation in expressive speech. In M. Heldner (Ed.), *Phonum: Reports in phonetics* (No. 9, 53-56). Umeå, Sweden: Department of Philosophy and Linguistics, Umeå University.

Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics, 27,* 24-27.

Tartter, V. C., & Braun, D. (1994). Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustical Society of America, 96,* 2101-2107.

Tatham, M., & Morton, K. (2004). *Expression in speech. Analysis and synthesis.* New York: Oxford University Press.

Ternström, S. (1996). Soundswell Signal Workstation 3.4 (Computer software). Stockholm, Sweden: Soundswell Music Acoustics HB.

Teunisse, J.-P., & de Gelder, B. (2001). Impaired categorical perception of facial expressions in high-functioning adolescents with autism. *Child Neuropsychology, 7,* 1-14.

Titze, I. R. (1994). *Principles of voice production.* Englewood Cliffs, NJ: Prentice-Hall.

Tomkins, S. (1962). *Affect, imagery, and consciousness: Vol 1. The positive affects.* New York: Springer.

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science, 11,* 188-195.

van Bezooijen, R. (1984). *Characteristics and recognizability of vocal expressions of emotion.* Dordrecht, the Netherlands: Foris.

van Bezooijen, R., Otto, S. A., & Heenan, T. A. (1983). Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics. *Journal of Cross-Cultural Psychology, 14,* 387-406.

von Bismarck, G. (1974). Sharpness as an attribute of the timbre of steady state sounds. *Acustica, 30,* 146-159.

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior, 17,* 3-28.

Wallbott, H. G., & Scherer, K. R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology, 51,* 690-699.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98,* 219-235.

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,* 557-580.

Wildgruber, D., Pihan, H., Ackermann, H., Erb, M., & Grodd, W. (2002). Dynamic brain activation during processing of emotional intonation: Influence of acoustic parameters, emotional valence and sex. *NeuroImage, 15,* 856-869.

Wiley, R. H., & Richards, D. G. (1978). Physical constraints on acoustic communication in the atmosphere: Implications for the evolution of animal vocalizations. *Behavioral Ecology and Sociobiology, 3,* 69-94.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America, 52,* 1238-1250.

Winkler, R. (2002). *Laryngographische Untersuchungen zum stimmlichen Ausdruck von affektiven Sprecherzuständen* [Laryngographic investigations of vocal expressions of speaker’s affective states]. Unpublished master’s thesis, Technische Universität Berlin.

Wundt, W. (1924). *An introduction to psychology* (R. Pintner, Trans.). London: Allen & Unwin. (Original work published in 1912)

Viscovich, N., Borod, J., Pihan, H., Peery, S., Brickman, A. M., Tabert, M., Schmidt, M., Spielman, J. (2003). Acoustical analysis of posed prosodic expressions: Effects of emotion and sex. *Perceptual and Motor Skills, 96,* 759-771.

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition, 63,* 271-313.

Zwicker, E., & Fastl, H. (1999). *Psychoacoustics: Facts and models* (2nd ed.). Berlin: Springer.

Acta Universitatis Upsaliensis

Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences
Editor: The Dean of the Faculty of Social Sciences

A doctoral dissertation from the Faculty of Social Sciences, Uppsala University, is either a monograph or, as in this case, a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences*. (Prior to July 1985, the series was published under the title "Abstracts of Uppsala Dissertations from the Faculty of Social Sciences".)

Distributor:
Uppsala University Library,
Box 510, SE-751 20 Uppsala, Sweden
www.uu.se, acta@ub.uu.se

ISSN 0282-7492
ISBN 91-554-6091-7