Title: Experimental study of affect bursts

Author: Marc Schröder

Affiliation: DFKI, Saarbrücken, Germany / Institute of Phonetics, Saarland University

Address for correspondence:

Marc Schröder
DFKI
Stuhlsatzenhausweg 3
D-66123 Saarbrücken
email: schroed@dfki.de

Number of pages: 49

Number of tables: 7

Number of figures: 2

Keywords: Affect Bursts, Interjections, Emotion

Abstract

The study described here investigates the perceived emotional content of “affect bursts” for German. Affect bursts are defined as short emotional non-speech expressions. This study shows that affect bursts, presented without context, can convey a clearly identifiable emotional meaning. The influence of the segmental structure on emotion recognition, as opposed to prosody and voice quality, is investigated. Agreement between transcribers is used as an experimental criterion for distinguishing between reflexive raw affect bursts and conventionalised affect emblems. A detailed account of 28 affect burst classes is given, including perceived emotion and recognition rate in listening and reading perception tests as well as a phonetic transcription of segmental structure, voice quality and intonation.

Zusammenfassung

Die hier vorgestellte Studie untersucht den wahrgenommenen emotionalen Gehalt von “Affect Bursts” für das Deutsche. Affect Bursts werden definiert als kurze, emotionale, nichtsprachliche Ausdrücke. Diese Untersuchung zeigt, dass Affect Bursts, ohne Kontext präsentiert, eine klar identifizierbare emotionale Bedeutung vermitteln können. Der Einfluss der segmentellen Struktur auf die Emotionserkennung, gegenüber Prosodie und Stimmqualität, wird untersucht. Übereinstimmung zwischen Transkribierern wird als ein experimentelles Kriterium zur Unterscheidung zwischen reflexiven Rohen Affect Bursts und konventionalisierten Affekt-Emblemen verwendet. Eine detaillierte Beschreibung von 28 Affect Burst Klassen wird gegeben, die wahrgenommene Emotion

und Erkennungsrate in Hör- und Lese-Perzeptionstests beinhaltet, sowie eine phonetische Transkription von segmenteller Struktur, Stimmqualität und Intonation.

Résumé

L'étude présentée ici s'interroge sur le contenu émotionnel perçu des « Affect Bursts » pour l'Allemand. Les Affect Bursts sont définis comme expressions courtes, émotionnelles et non-verbales. Cette étude démontre que les Affect Bursts, présentées sans contexte, peuvent transmettre un sens émotionnel clairement identifiable. L'influence de la structure ségmentale sur la reconnaissance des émotions, vis-à-vis de la prosodie et du timbre, est étudiée. Le degré d'accord entre les transcripteurs est utilisé comme critère expérimental pour distinguer entre les Affect Bursts Crus, réflexifs, et les Emblèmes Affectives conventionalisées. Un compte rendu détaillé de 28 classes d'Affect Bursts est présenté, comprenant l'émotion perçue et le taux de reconnaissance dans des tests de perception orale et écrite, ainsi qu'une transcription phonétique de la structure ségmentale, du timbre et de l'intonation.

1. INTRODUCTION

Studying emotional expression in speech is inherently difficult. Problematic issues range from the description of emotion itself (Cowie, 2000), via the collection of emotional speech material (Campbell, 2000), to the appropriate evaluation in perception tests (Cauldwell, 2000). Even the delimitation of the domain under study with respect to neighbouring topics such as expression of attitudes (Wichmann, 2000) and even linguistic prosodic structure (Scherer et al., 1984) is difficult and fuzzy.

This study investigates a phenomenon at the heart of many of these difficulties: The so-called “affect bursts”. After introducing the concept, I will summarise some of the difficult aspects of the research domain in general and how these apply to the phenomenon under study here. The particularities of affect bursts compared to the broader field of speech and emotion are highlighted, and research questions are formulated before the experimental study is described.

1.1. The concept of affect bursts

The concept of “affect bursts” has been introduced by Scherer (Scherer, 1994). He defines them as “*very brief, discrete, nonverbal expressions of affect in both face and voice as triggered by clearly identifiable events*” (p. 170). Coined in the context of the psychological literature on emotion expression, the term “affect burst” overlaps strongly with what might be called “affective interjections” in linguistics. However, the limits of the domains of affective interjections and affect bursts differ. On the one hand, a verbal interjection expressing an emotion (“Heaven!”) would not be considered an affect burst

due to its verbal nature. On the other hand, a non-phonemic affect burst like laughter or a rapid intake of breath would probably not be considered an interjection.

The question of the sign status of affect bursts is discussed by Scherer (1994) from the point of view of his push-pull distinction (Scherer, 1988). Push effects are physiological factors (like pain) leading to an expression; pull effects are social rules and expectancies representing culturally shared "targets" for appropriate expressions in a given situation. While both types of effects are always present, one of them may prevail in a given situation. In this line of ideas, Scherer (1994) proposes to make a distinction between 'raw affect bursts' on the "push" end of that continuum, and 'affect emblems' on the "pull" end of the continuum. Consequently, raw affect bursts are raw, reflexive vocalisations that are expected to be barely conventionalised, thus relatively universal, and show strong inter-individual differences. Affect emblems, on the other hand, are conventionalised symbols, i.e. strongly culture-dependent, showing comparatively few and small individual differences. As raw affect bursts and affect emblems are seen as extreme points on a burst-emblem continuum, all sorts of mixtures are expected to exist. The term "affect burst" is used as a general term referring to the entire continuum between raw affect bursts and affect emblems. Figure 1 illustrates the terms used.

---
insert Figure 1 about here

1.2. Problematic issues in studying emotions in speech

The scientific study of emotions in speech is facing major problems, both methodological and conceptual. In a review of the literature, Scherer (1986) comes to the conclusion that due to these problems, "*there has been neither continuity nor*

cumulativeness in the area of the vocal communication of emotion" (p. 143). A number of problems that seem relevant to the present study are summarised below.

1.2.1. Defining emotional states

The first challenge in studying the expression of emotional states in speech is the adequate description of these states themselves. Many studies that simply define the states under study using plain emotion words, such as "anger", "fear", "sadness" etc., come to contradicting results (Scherer, 1986) due to the ambiguity of the terms employed. E.g. for anger, the emotional properties, and consequently the vocal realisations, of "hot anger" and "cold anger" are very different.

In order for research results to be meaningful and interpretable, it seems thus necessary to attempt a more precise description of the emotional states studied (Cowie, 2000). In studies using acted speech, a reasonable way of describing the emotional connotations encoded seems to be the use of frame stories describing the imagined situational context in which an utterance is spoken by the actors (Leinonen et al., 1997; Schröder, 1999). Another approach, capturing some basic properties of perceived speaker emotion in perception studies, uses emotion dimensions (Cowie et al., 2001; Dietz & Lang, 1999; Pereira, 2000, Schröder et al., 2001). Three dimensions are commonly considered most relevant: *arousal* (or *activation*), i.e. the degree of physiological arousal and readiness to take some action; *valence* (or *evaluation*), in terms of positive or negative evaluation of some object or event; and *control* (or *power*), i.e. how dominant or submissive the speaker is. As a perception-oriented tool, emotion dimension ratings provide a quantifiable description of the listener's perception of the speaker's emotional state (see

a discussion of speaker-centred and listener-centred descriptions of emotions in Cowie, 2000).

1.2.2. Delimitating the domain under study

Emotion is not an easy to grasp phenomenon, and as a consequence, a clear delimitation of the research domain against neighbouring fields is difficult. While attempts are sometimes made to distinguish more physiological emotions from more cognitive attitudes (Wichmann, 2000), that distinction seems to be gradual rather than strictly categorical. Similarly, the distinction between emotion expression through prosody on the one hand and the linguistic functions of prosody on the other hand cannot be clearly drawn, because of interaction phenomena: Linguistically defined intonation contours have been shown to convey emotional meaning depending on sentence type (Scherer et al., 1984). This illustrates the difficulty to draw a clear boundary between the linguistic and paralinguistic aspects of vocalisations, a theme on which we will see a variation when discussing the question of the word status of affect emblems.

1.2.3. The use of acted emotional speech material

The impossibility to obtain spontaneous emotional speech in a controlled way is a problem to every single study on emotional speech. Some studies give priority to spontaneity, taking into account the increased difficulty in describing the emotional states expressed (e.g. Campbell, 2000; Douglas-Cowie et al., 2000). Most studies, however, use acted emotion, an approach justified by Banse & Scherer (1996) with the argument that even in real life, people need to enact emotional expression with a certain

amount of voluntary control. The approach is admittedly not without problems: Depending on the methodology and the talent of the actors, acted emotional expression can be well distinguishable from spontaneously occurring emotion (Schröder et al., 1998). Therefore, a number of studies used experts to evaluate the quality of the acted material in pre-selection procedures (Banse & Scherer, 1996; Leinonen et al., 1997, Schröder, 1999), only using in the actual study what was rated "successful displays" by the experts.

1.2.4. Multiple channels of emotion expression

Emotion is expressed through multiple channels. These include at least facial expression (Ekman, 1982), verbal content (Scherer & Ceschi, 2000), and the voice, comprising gradual (Banse & Scherer, 1996) and categorical (Mozziconacci, 1998) prosodic parameters, voice quality (Gobl & Ní Chasaide, 2000), and articulation precision (Kienast et al., 1999). It is probable that emotion has an effect on other expressive behaviour as well, such as gestures, wording, speaking disfluency, etc.

Many of these channels through which emotion is expressed have been shown to be able to convey emotion on their own, including facial expression, prosody, and voice quality. Other channels may be influenced by the emotion without actually containing sufficient information to allow emotion recognition. This might be the case for articulation precision, gestures and others.

While it is important to establish the contribution of a given channel by studying it in isolation, one question that would merit much more attention is that of the interaction among these channels when they co-occur in a multi-modal and/or situated context. It

has been shown that voice quality, F0 range and intonation contour type do not interact when producing a perceived emotional message (Ladd et al., 1985), but that intonation contour type interacts with verbal content (Scherer et al., 1984); that a given utterance conveys a different emotional message when presented in isolation and with situational context (Cauldwell, 2000); and that a coherent facial and vocal display of a given emotion is perceived as more natural than a facial emotion display accompanied by a neutral voice (Stallo, 2000). However, there seem to be few studies investigating the simultaneous display of conflicting messages as can be observed, e.g., in irony.

1.3. Positioning this study

Affect bursts, although theoretically described in detail, do not seem to have been extensively studied experimentally. Existing descriptions of interjections come from a linguistic background (Ehlich, 1986; Scherer, 1994; Zerling, 1995). However, these studies give definitions and classifications that seem to be based mainly on the authors' intuitions, and do not give any indications of whether and to what extent the corresponding vocalisations are actually perceived as carrying identifiable emotional meaning.

The problems in studying affect bursts are slightly different from those typically encountered in speech and emotion studies (see 1.2). Still, there are useful methods that can be applied in this context. After a discussion of the specificity of the topic under study, the research questions that naturally arise are stated, experimental criteria are formulated, and the methodology adopted is outlined.

1.3.1. The specificity of affect bursts

Typically studied channels for emotion expression (see 1.2.4) are global in the sense that they can co-occur with spoken language, accompanying a spoken utterance. This is different for affect bursts, which by definition are short, delimited events. That different nature of affect bursts is a reason to question whether observations made, e.g., in the domain of speech prosody are generalisable to affect burst prosody, and vice versa.

In affect bursts, the segmental structure itself is expected to carry emotional meaning. In most studies of emotion and speech, some constant verbal content serves as a "carrier" for emotional prosody and voice quality, and the latter are varied. Semantically neutral sentences (e.g., Paeschke & Sendlmeier, 2000) or pseudo-sentences consisting of logatomes (Banse & Scherer, 1996) are often used as carriers, but also single-word utterances such as a name (Leinonen et al., 1997).

Unlike these studies, an investigation of affect bursts will need to consider the segmental structure as an essential part of the affect burst. Along the lines of thought of multi-channel emotion expression (see 1.2.4), it then makes sense to ask for the relative contributions of the segmental structure and of prosody and voice quality on emotion recognition.

A further specificity of studying affect bursts stems from the distinction between raw affect bursts and affect emblems proposed by Scherer (see 1.1). Experimental criteria will need to be formulated for characterising affect bursts as raw bursts or emblems, while taking into account the non-categorical nature of the distinction.

1.3.2. Research questions

From the preceding considerations, the following questions arise:

(I) Can affect bursts, produced by actors, convey the intended emotional meaning when presented in isolation and audio only?

(II) What is the contribution of the segmental structure of affect bursts on emotion recognition?

(III) Can a distinction between raw affect bursts and affect emblems be proposed based on experimental criteria?

1.3.3. Working definitions and criteria

Scherer's definition of affect bursts needs to be adapted for the purpose of this experimental study. On the one hand, the facial-vocal interaction and synchronisation that he stresses can be left out, because this study is only concerned with the vocal aspect of affect bursts. On the other hand, the intrinsically fuzzy boundaries of the concept need to be stated as explicitly as possible in order for the definition to be useable as a selection criterion. Therefore, the following working definition was used: Affect bursts are short, emotional non-speech expressions, comprising both clear non-speech sounds (e.g. laughter) and interjections with a phonemic structure (e.g. "Wow!"), but excluding "verbal" interjections that can occur as a different part of speech (like "Heaven!", "No!", etc.).

This definition is meant to delimit the concept of affect bursts as illustrated in Figure 1. Again, it must be clear that the boundaries are fuzzy, and that neither on the

physiological end nor on the verbal end, a clear delimitation is feasible. In this respect, it seems difficult to draw the line around the concept of affect bursts in a more clear-cut way than the boundary between paralinguistic and linguistic phenomena in general (see 1.2.2).

While Scherer's distinction between raw affect bursts and affect emblems, summarised above, is formulated in terms of production, it seems to make sense to propose an additional, perception-based criterion that can be used in this experimental study. As a conventionalised symbol, an emblem should correspond to a reference pattern in a listener's mind. Similar to a lexical entry (a word), that mental pattern comprises an expected form and a meaning. When a given vocalisation of that emblem is matched against the pattern, the expected phonemic form may influence the perception through top-down processing (McQueen & Cutler, 1997), leading to the perception of a more standardised phonemic form. For a raw affect burst, on the other hand, no such reference pattern should exist. Consequently, bottom-up processes would play a more important role in the perception of the phonetic form, leading to more variability in the perceived form, especially for non-expert transcribers.

The following criterion is therefore proposed:

(1) Affect emblems, when transcribed, are expected to show less variability between transcribers than raw affect bursts.

In addition, in the mental representation, a meaning (in this case: an emotion) is associated to the phonemic form of the emblem, leading to a second criterion:

(2) The emotion recognition from a phonemic transcription should be quite accurate for affect emblems.

However, no prediction can be made about the recognition accuracy of raw affect bursts, which may or may not rely on the segmental structure for conveying emotional meaning.

1.3.4. Outline

A list of “German” affect bursts was compiled. On the basis of this list, ten emotion categories were established. Acted realisations of affect bursts intended to express these emotion categories were recorded. The intended connotation of the emotion words was specified through frame stories (see 1.2.1). A pre-selection of the n most successful examples was made based on expert ratings (see 1.2.3).

Question (I), the question of recognisability, was addressed in a forced-choice perception test where affect bursts are presented in isolation and audio only. In addition, the perceived emotional meaning was assessed using scales representing the three emotion dimensions typically studied (see 1.2.1). That information was used to establish the degree of perceived emotional similarity between the different affect bursts. Confusions were interpreted in the light of that information.

Affect bursts were transcribed phonetically and grouped into classes based on segmental phonetic similarity. A detailed account of the perceptual properties of these classes was given.

Question (II), the question of the contribution of segmental structure, was addressed in a written perception test based on orthographic transcriptions.

Question (III), the distinction between raw affect bursts and affect emblems, was addressed by applying the criteria developed under 1.3.3 to the orthographic transcriptions.

2. METHOD

2.1. Collection of a list of affect bursts

While the concept of “raw” affect bursts claims a low degree of conventionality, and thus a relative language-independence, the same is not true for affect emblems that are conventionalised and thus most likely culture- and language-dependent. As any affect burst is considered to be located somewhere on the raw burst-emblem continuum (see Figure 1), any list of affect bursts should, in a first step, be compiled for a given language.

Therefore, a list of “German” affect bursts was assembled from the available literature (Ehlich, 1986; Scherer, 1994; Zerling, 1995) and from personal observation. For the entries in the lists of Italian (Scherer, 1994) and French (Zerling, 1995) interjections, equivalents conveying a similar meaning were sought that could have been produced by a German speaker. In accord with the working definition of affect bursts given above (see 1.3.3), only non-verbal vocalisations expressing emotions were to be included. This requirement excludes purely physiological sounds like sneezing, snoring, or a hiccup, as well as verbal interjections. On the other hand, etymology was not a criterion for exclusion. For example, “Oje!” was included as an affect burst although it has verbal

origins (“O Jesu domine!”, Drosdowski, 1989). Altogether, the resulting list comprises about 80 different affect bursts.

Due to the way the list was created, no claim can be made that the list contains all or even the most frequent affect bursts typically employed by German speakers. It is merely considered a starting point from which to extract recording material.

2.2. Definition of emotion categories

Based on informal inspection of this list of affect bursts, ten emotion categories were established that seemed to be typically expressed by affect bursts. The emotion categories are: “Bewunderung” (admiration), “Drohung” (threat), “Ekel” (disgust), “Große Freude” (elation), “Langeweile” (boredom), “Erleichterung” (relief), “Schreck” (startle), “Sorge” (worry), “Verachtung” (contempt), and “Wut” (hot anger). For each of these ten emotion categories, the author selected two affect bursts from the list to be used in recordings.

Most of the emotion categories considered “basic” or “primary” (Murray & Arnott, 1993) by some authors (but see Ortony & Turner, 1990) are represented in this list: anger, joy (through elation), disgust, and maybe fear (through startle). Sadness, on the other hand, is absent. This only reflects the observation that few affect bursts that seemed to express sadness were found in the established list of affect bursts. As that list does not pretend to cover all affect bursts typically used, any conclusion stating that sadness would not typically be expressed through affect bursts is likely to be premature.

In order to define the intended emotions for the recordings, a frame story was constructed for each of the ten emotions[1].

2.3. Recordings

During recordings, speakers silently read the frame story for a given emotion (see 2.2), and then produced an affect burst *of their choice* that they spontaneously associated with the situation described by the frame story. Only after that did they see the two affect bursts chosen from the list (see 2.2) for the same emotion, and produce them in the same spirit evoked by the frame story. Thus, each speaker produced 30 vocalisations, three per emotion.

Six speakers (three male, three female) between the age of 25 and 32 years took part in the recordings. Four of them (two male, two female) were amateur actors. Recordings were conducted in a sound-treated room, with a Sennheiser MKH 20 P48 microphone, and recorded onto a DAT tape. The speech material was re-digitised at 16 kHz, 16 bit during the transfer to a PC. A higher sampling rate was purposefully avoided in order to guarantee a sound quality similar to the one typically used in state-of-the-art concatenative speech synthesis (e.g., Dutoit et al., 1996; see also section 6).

2.4. Pre-selection

A pre-selection procedure was carried out in order to reduce the data to a subset of good quality (see 1.2.3). For this purpose, each speaker rated the other five speakers' vocalisations. The 180 vocalisations (6 speakers * 30 vocalisations/speaker) were presented ordered by emotion and by speaker, along with the intended emotion label.

Judges were to rate how well the vocalisation expressed the intended emotion, using German school marks (i.e., a 6-point scale where 1 is best and 6 worst; only values ≤ 4 are considered acceptable).

The selection of stimuli for the listening test was a two-step procedure. First, if two vocalisations of a given emotion by a given speaker seemed auditorily very close, the one with the worse quality rating was discarded. Then, the eight vocalisations with the best quality ratings (= lowest numbers) were selected as stimuli. The ratings still considered acceptable ranged from 2.6 for relief and admiration to 3.8 for startle and contempt, an indication that the former might have been easier to express convincingly than the latter.

The contributions of the different speakers were very unequal (see Table 1). Amateur actors contributed significantly more stimuli than non-actors (repeated measures ANOVA, $F(1,4)=14.6$, $p=0.019$). This is an indication that actors seem to be more capable of producing affect bursts “on command” than non-actors.

---
insert Table 1 about here

2.5. Listening test

In the listening test, two types of rating were collected. On the one hand, subjects had to identify each stimulus as one of the ten emotion categories (see 2.2.). On the other hand, they had to position each stimulus on the following three seven-point scales: “Aufgeregt-ruhig” (excited-calm, the *arousal* scale), “positiv-negativ” (positive-negative, the *valence* scale), and “dominant-untergeordnet” (dominant-subordinate, the

control scale). These scales represent the three emotion dimensions introduced earlier (see 1.2.1) and provide information about basic emotion properties of the stimuli.
20 naïve subjects (ten male, ten female, between 23 and 49 years old) participated in the listening test. Stimuli were individually randomised and presented over headphones. Subjects could listen to each stimulus as many times as they wanted. Answers were given through a graphical interface on a computer screen.

2.6 Expert transcription and grouping

Each emotion was expressed through two prompted affect burst vocalisations and one that was freely produced by each speaker (see 2.3). The versions of a prompted affect burst produced by different speakers were considered as belonging to the same affect burst class. For the freely produced affect burst vocalisations, a decision had to be taken whether they belonged to one of the two established classes for the given emotion, or whether they belonged to a different class for that emotion. To that end, an approximate phonetic transcription of the 79 vocalisations used in the listening test was performed by the author. Segmental phonetic similarity was used as the criterion for grouping the vocalisations for each emotion into affect burst classes[2] (see Table 4). Due to the freely produced affect burst vocalisations, the number of distinct affect burst classes per emotion varies from two to four, with a total of 28 classes. Consequently, the number of vocalisations per class varies from one to six. For affect burst classes represented by more than one stimulus, the phonetic transcription given in Table 4 corresponds to a 'typical' pronunciation. Three voice qualities were distinguished: Modal ([aː]), breathy ([a̤ː]) and creaky ([a̰ː]). Intonation was transcribed using a simple level-based system

based on early American structuralist systems (Wells, 1945). Four pitch levels are distinguished where 1 is lowest and 4 highest. Pitch movements are described by indicating start and end level of the movement, e.g., “4-1” would be a fall from the highest to the lowest pitch. This transcription system was preferred to modern, more complex systems like ToBI because linguistically motivated concepts like accents and boundary tones were considered inadequate for the transcriptions of short affect bursts. In addition, this simple system allows at least a limited account of pitch range, which would be difficult in a system like ToBI distinguishing only high and low tones.

2.7 Non-expert orthographic transcription

Within each affect burst class, the vocalisation with the highest recognition rate in the listening test was identified. These 28 vocalisations (one per class) were presented in a transcription task. 10 non-expert transcribers that had not been involved in the preceding tests (5 male, 5 female, native German speakers, students of various topics) took part in the task. They listened to the 28 vocalisations individually via headphones in randomised order, with the possibility to hear each vocalisation as many times as they wanted to. Half the subjects heard the stimuli in the reverse order. They were asked to write down what they heard, using German letters.

From the ten resulting orthographic transcriptions of each affect burst, a representative transcription was chosen. This was either the most frequent transcription or, if none was most frequent, one as similar as possible to the transcriptions (i.e., minimising the transcription variability measure introduced below).

Transcription variability was calculated for each affect burst by summing up the distances between each individual transcription and the representative transcription. Distances were quantified as the minimal number of deletions, insertions or replacements of units needed to transform a given transcription into the representative transcription. An attempt was made to do this in a sensible way from the point of view of German pronunciation rules, e.g. by treating “sch” (pronounced [ʃ]) as one unit.

2.8 Written perception test

The representative transcriptions for the 28 affect burst classes were presented in a written perception test. Without hearing the original vocalisations, the subjects had to identify the emotion expressed by an affect burst, relying only on its orthographic transcription. The same ten emotion categories as before (see 2.2) were used in a forced choice setting. The 27 transcriptions[3] were presented in randomised order on paper, with boxes to tick for the ten emotions. 20 subjects that hadn’t taken part in any of the preceding tests (10 male, 10 female, native speakers of German, students of various topics) rated the transcriptions. To half of the subjects, the transcriptions were presented in reverse order. 3.1% of the answers were invalid (zero or more than one box ticked) and were treated as missing values.

3. RESULTS

3.1. Listening test

3.1.1. Recognition rates

In the listening test, that addressed Question (I), the fundamental question whether affect bursts can convey emotional meaning, the overall mean recognition rate is 81.1%. The mean recognition rates for the ten emotions are shown in Table 2.

---
insert Table 2 about here

Admiration, disgust and relief are recognised from affect bursts with more than 90% accuracy. The least recognised categories are threat and anger with just over 60% accuracy. In the cases where identification was not as intended, it is interesting to look for systematic confusion patterns between emotions (Table 3).

---
insert Table 3 about here

Bi-directional confusions can be found between threat and anger as well as between boredom and worry. The only other confusion worth mentioning is that elation is sometimes identified as relief.

3.1.2. Recognition of affect bursts

The recognition rates in section 3.1.1 do not take into account the fact that the vocalisations of a given emotion do not form a homogeneous set. In fact, each emotion is expressed by several affect burst classes (see 2.6 and Table 4).

---
insert Table 4 about here

It can be seen from Table 4 that in the listening test, only one affect burst class was not recognised as the intended emotion: the growls intended to express threat were reliably classified as anger. All intended emotions except anger are expressed reliably (with ≥80% recognition rate) by at least one affect burst class.

The confusions between emotions, observed in Table 3, are mainly due to particular affect burst classes within the concerned emotions. This can be seen from a reduced confusion matrix (Table 5) showing affect burst classes with a 'correct' recognition rate of 70% or less, and only selected emotions.

---
insert Table 5 about here

Table 5 suggests that the two types of growl (intended as threat and anger, respectively) may actually form a single affect burst class, associated with anger, but also identified to a lesser extent as threat. The bi-directional confusions between threat and anger are due to this growl class and the ambiguous 'anger breath out' class. The 'anger Oh' class seems to lack clear perceptual cues, resulting in a more wide-spread ambiguity. The 'threat Hey' class (Table 4), however, conveys threat reliably.

Similarly, the bi-directional confusions between boredom and worry are due to three affect burst classes (Table 5) that are perceived as both boredom and worry, while for each of the two emotions, well-recognised affect burst classes exist (Table 4). In particular, the two types of 'Hmm' (boredom and worry) are only confused to a small degree, indicating that although they are phonetically similar, the perceptual differences are large enough to allow a distinction. A possible perceptual cue distinguishing them is the pitch contour which is rising for 'boredom Hmm' and falling for 'worry Hmm'.

Similarly, the two types of sigh, intended as boredom (Tables 4 and 5) and relief (Table 4), show very different perception patterns, indicating that despite their apparent segmental and supra-segmental similarity, they contain clearly distinct perceptual cues. Finally, the uni-directional confusion between elation and relief is due to the 'elation Ja' class (Table 5).

3.2. Orthographic transcription and written perception test

3.2.1. Transcription variability

The results of the orthographic transcription task are shown in Table 4. It can be seen that transcription variability varied strongly between the different vocalisations. This allows the application of criterion (1), developed in 1.3.3, for the distinction between the 'raw affect burst' extreme and the 'affect emblem' extreme on the burst-emblem continuum (see 4.3 for details).

3.2.2. Recognition rates

In the written perception test addressing Question (II), the role of the segmental structure for emotion recognition, the overall mean recognition rate in the written perception test is 65.1%. This is lower than in the listening test, but still far above chance level. The mean recognition rates for the ten intended emotions are shown in Table 6.

---
insert Table 6 about here

Clear differences can be seen between the emotions. For one group of emotions (admiration, disgust, elation, contempt, and to a lesser extent relief and worry), the recognition is roughly as high as in the listening test. This indicates that the segmental structure is sufficient for the recognition of these emotions. Another group of emotions (threat, boredom, startle, and anger) show comparatively low recognition rates, indicating that for these emotions, prosody and voice quality provided important cues in the listening test that were missing in the reading test. It is interesting to note that these emotions are among those with the most extreme arousal values (see 3.3).

A look at the individual affect burst classes (Table 4) reveals that the group of affect bursts easily recognised on the basis of their segmental transcription overlaps strongly with those affect burst classes showing low transcription variability (see 3.2.1), a point which will be discussed further in 4.3. On the other end of the spectrum, it is not surprising to find that the affect burst classes lacking an easily identifiable segmental structure (the yawn and the rapid intake of breath) are not recognised from the orthographic transcriptions.

3.3. Emotion dimensions

The scale data obtained in the listening test (see 2.5), reflecting the emotion dimensions arousal, valence and control (see 1.2.1), provides interesting information about the characteristics of emotions. In this context, it provides a means for assessing the degree and type of similarity or difference between the expressed emotions.

The means and standard deviations on the three scales were calculated for the ten emotion categories, based on *correct* categorical ratings (Table 7). This conservative

selection was made to ensure maximum consistency among the scale ratings of an emotion, although it does not seem a priori evident whether wrong categorisations actually do coincide with higher variance in scale ratings.[4] Figure 2 (a,b,c) shows the three 2-dimensional projections of the corresponding 3-dimensional space.

---
insert Table 7 about here

It is interesting to note that the emotion categories in confused pairs (threat/anger, boredom/worry) are relatively close to each other. This indicates a similarity of these emotion categories along the three dimensions investigated here. This type of information seems highly useful, e.g. in the context of applications: A confusion among semantically similar emotion categories is much less disturbing than one among semantically very different categories that happen to have a similar surface form. However, it is clear that the three dimensions cannot be expected to capture all relevant aspects of emotion properties. The most outstanding example for this is disgust and anger: While very close along all three dimensions (Figure 2), the two categories are practically not confused at all (Table 3). The feature allowing the clear-cut distinction between these categories seems to be a highly specialised one, that would only be captured in a richer “Schema” type description of emotion properties using a larger number of dimensions (Cowie et al., 1999).

The arousal-valence plane (Figure 2a) corresponds to what Cowie et al. (1999) call the activation-evaluation space. They had subjects locate emotion words on these two dimensions simultaneously, by clicking with a mouse on the appropriate location in the two-dimensional plane. Positions of emotion categories in Cowie et al. 1999 seem to be

quite close to roughly comparable emotion categories in this study (when pairing anger with ‘angry’ in Cowie et al. 1999, worry with ‘worried’, boredom with ‘bored’, relief with ‘satisfied’, and elation with ‘happy’). Although this is only a qualitative indication at the moment, it looks encouraging that very different methodologies in obtaining dimensional judgements seem to lead to similar results.

---
insert Figure 2 about here
---

4. DISCUSSION

4.1. Recognition

The recognition accuracy for ten emotions expressed through affect bursts, presented without context and audio only, is very high (81% in mean). For many affect bursts, there is very little ambiguity (accuracy > 90%). This suggests that affect bursts are a highly effective means of expressing emotion. The recognition rates are considerably higher than those found for the expression of emotion through prosody and voice quality: On pseudo-sentences consisting of logatomes, Banse & Scherer (1996) obtained 55% correct recognition for ten emotions; on a single word (a name), Leinonen et al. (1997) obtained 50% correct recognition, also for ten “emotional-motivational connotation” categories. The information conveyed by the segmental structure of the affect bursts seems to be the disambiguating factor leading to the higher recognition rates for affect bursts.

4.2. Suitableness and prototypicality

Differences between emotions are worth a closer look. It seems possible that affect bursts play a more important role in expressing some emotions than others. Disgust is the prime example of an emotion that seems to be typically expressed through affect bursts. In Banse & Scherer (1996), the recognition rate for disgust, when expressed through prosody and voice quality accompanying speech, is particularly low (15%). The possible explanation given by Banse & Scherer (1996) is that disgust is not likely to be expressed through long utterances, but rather through short affect bursts. The results in the present study, with disgust being conveyed through affect bursts with 93% accuracy, lend support to that assumption. Similarly well-recognised emotions are relief, admiration, and startle with recognition rates around 90%. Conversely, it seems probable that some emotions are less typically expressed through affect bursts. The lower recognition rates for anger may indicate that anger is such a case.

Johnstone et al. (1995), in a re-analysis of the data from Banse & Scherer (1996), came to the conclusion that some emotions (e.g., hot anger and boredom) are expressed through prototypical configurations of acoustic variables, and are consequently easily produced by actors and highly recognised by listeners, while other emotions lack typical acoustic configurations and are poorly recognised (e.g., disgust). Similarly, it may be that some emotions are expressed through prototypical affect bursts, while others are not. A simple criterion for prototypes could be that they are spontaneously produced by speakers and easily identified by listeners. In this study, at least ‘Buäh’ (disgust) and ‘rapid intake of breath’ (startle) seem to fulfill that criterion (5 out of 6 speakers spontaneously produced the mentioned affect bursts; recognition accuracy >90%). On

the other hand, emotions lacking prototypical affect bursts for their expression should show more diversity of spontaneously produced affect bursts, and recognition rates should be lower. In this study, elation (‘Ja’, see footnote 2), as well as boredom and contempt (large variety of affect bursts spontaneously produced) seem to fulfill at least the production part of this criterion. Thus, these emotions may lack clear affect burst prototypes.

4.3. Affect bursts and affect emblems

An answer to Question (III), the distinction between raw affect bursts and affect emblems, can be attempted by applying the two criteria formulated in 1.3.3 to the results of the transcription task and the written perception test (reported in 3.2 and Table 4). Criterion (1) said that transcription variability should be high at the ‘raw affect burst’ end and low at the ‘affect emblems’ end of the burst-emblem continuum. The highest variability was found for the ‘disgust Ih’ and ‘boredom yawn’ affect bursts (Table 4). According to criterion (1), these qualify as prime candidates for the ‘raw affect burst’ extreme of the burst-emblem continuum. Less extreme examples would be ‘boredom sigh’, ‘contempt laughter’ and ‘worry Oweh’[5]. Very low transcription variability, on the other hand, was found for a number of vocalisations, the most extreme examples being ‘disgust Buäh’, ‘disgust Igitt’, ‘elation Yippie’, ‘elation Hurra’, and ‘relief Puh’. Criterion (1) places these at the ‘conventionalised emblem’ end of the burst-emblem continuum. Vocalisations with medium transcription variability would be somewhere in the middle of the continuum.

Criterion (2) predicted high recognition rates in the written perception test for affect emblems. This confirms the emblem status for most of the candidates produced by criterion (1), but removes some. In particular, “disgust Buäh” must be considered less likely to be an extreme emblem despite its low transcription variability because of its medium recognition rate.

In summary, clear affect emblem candidates are ‘disgust Igitt’, ‘elation Yippie’, ‘elation Hurra’, and ‘relief Puh’. Less extreme, but still relatively clear are ‘admiration Wow’, ‘admiration Boah’, ‘contempt Pah’ and ‘contempt Tse’. Clear candidates for the opposite ‘raw affect burst’ extreme would be ‘disgust Ih’, ‘boredom yawn’, and to a lesser extent ‘boredom sigh’ and ‘contempt laughter’.

4.4. Emotion dimensions and affect bursts

The distribution of the emotion categories in the three-dimensional arousal-valence-control space is relatively widespread. This indicates that affect bursts are not restricted to expressing a particular type of emotion, that would correspond to a delimited region in the three-dimensional emotion space, but are suitable for expressing some variety of emotions.

However, the distribution of expressed emotions is not uniform. Nearly all emotions expressed show medium to high levels of arousal, the only exception being boredom. The majority of emotions are judged relatively negative. While for arousal and valence, values close to the extremes can be found, the emotions are positioned more closely to the mid value for control. Conclusions from these observations should be drawn with some caution, though, because the procedure for selecting the emotion categories used

in this study (see 2.2) cannot claim to represent the entire spectrum of emotions typically expressed through affect bursts. While not claiming generalisability, these observations can serve as a starting point for future research. For example, one hypothesis that could be deduced from the above observations would be that affect bursts typically express aroused emotions.

In the recognition rates of the written perception test, an interesting pattern occurs. The emotions with the lowest recognition rates based on the segmental structure alone (threat, boredom, startle, and anger) are among the emotions with the most extreme (high or low) arousal values. The mean distance from the mean arousal value (i.e., 4) is 1.3 for these emotions, and 0.7 for the other six emotions. This observation is in line with findings in the literature showing that arousal correlates with prosodic parameters in speech, such as F0 mean and range (Banse & Scherer, 1996; Pereira, 2000; Schröder et al., 2001).

A cursory inspection of the intonation labelling in Table 4 confirms that the most aroused emotions, startle and elation, also have the highest pitch levels, while boredom, with a low level of arousal, also has a relatively low pitch level. In order to investigate that question more seriously, acoustic analyses, including F0 measures, would be the appropriate starting point, rather than the crude pitch descriptions presented in Table 4. Such acoustic measures could be correlated to dimensional ratings, as has been done by Schröder et al. (2001) for emotional speech. The outcome of such measures on affect bursts could shed some light on the question, briefly raised in 1.3.1, whether speech prosody and affect burst prosody behave similarly for emotion expression.

5. CONCLUSION

The high overall recognition rate of 81% indicates that affect bursts, when presented audio only and without context, seem to be an effective means of expressing emotion. Moreover, ten different emotion categories can be distinguished quite reliably. The grouping of individual vocalisations into affect burst classes, on the grounds of phonetic similarity, showed that confusions between emotion categories tend to be due to individual, ambiguous affect burst classes. For all emotions except anger, affect burst classes exist that express the intended emotion with more than 80% accuracy.

The degree to which emotion recognition is due to the segmental structure of the affect bursts has been assessed for each affect burst class. Along with transcription variability, this was used as an experimental criterion for the degree of conventionality in the use of an affect burst class. This allowed the establishment of a tentative list of candidates for the extreme ‘raw affect burst’ vs. ‘affect emblem’ poles of the burst-emblem continuum.

The properties of the ten emotions with respect to the three emotion dimensions arousal, valence and control were measured and reported. These data showed that confused categories were also emotionally similar. In addition, evidence was found that suggests a more important role of prosody and voice quality in conveying emotions with more extreme (high or low) arousal ratings compared to emotions with mean arousal ratings.

6. SUGGESTIONS FOR FUTURE RESEARCH

The study of affect bursts is at its very beginning. The current study has maybe opened a pathway, showing experimentally that affect bursts are highly recognisable and can convey a number of different emotions. From here, many questions can be investigated, including:

- Which emotions are typically expressed through affect bursts?
- In which contexts (situations, speaking styles, type of utterance, location with respect to speech utterances) do affect bursts naturally occur?

These questions could be addressed through the analysis of large corpora of spontaneously occurring emotional speech such as the one built by Douglas-Cowie et al. (2000).

- How universal are raw affect bursts, from a production and perception perspective?

Cross-cultural studies could verify the postulated universal recognisability of raw affect bursts, and compare the cross-lingual recognition rates of raw affect bursts and affect emblems. Comparisons of spontaneously produced affect bursts from speakers of different cultures might address the question from a production point of view.

- How phonemic is the segmental structure of raw affect bursts and of affect emblems? Can the use phonemic vs. non-phonemic segments serve as a criterion for raw bursts vs. emblems?
- Does the prosody of affect bursts show the same variation with emotion as the prosody of speech?
- How do segmental and supra-segmental properties of affect bursts interact with each other and with other channels of emotion expression?

In the medium term, the use of affect bursts might become interesting for application in spoken language interfaces. If affect bursts are found to be effectively used in spontaneous human communication, their use in speech recognition and speech synthesis systems, especially in emotionally charged situations, might be worth thinking about. The need for appropriate emotional interaction with speech interfaces is likely to become more obvious when systems move from the current formal scenarios into more casual environments. An emerging area with potential use for this type of emotion expression is the field of Conversational Agents (André et al., 2000; Dietz & Lang, 1999), where especially in agent-to-agent interactions, the convincing expression of emotions, including negative emotions, seems to become increasingly necessary.

ACKNOWLEDGEMENTS

Thanks to Ralf Benzmüller for sharing his observations of affect bursts. Thanks to Roddy Cowie, Jürgen Trouvain and three anonymous reviewers for very valuable feedback and suggestions.

7. REFERENCES

André, E., Rist, T., van Mulken, S., Klesen, M., & Baldes, S., 2000. The Automated Design of Believable Dialogs for Animated Presentation Teams. In: Cassell, J. et al. (Eds.), Embodied Conversational Agents, MIT Press, Cambridge, MA, pp. 220-255.

Banse, R. & Scherer, K. R., 1996. Acoustic Profiles in Vocal Emotion Expression. Journal of Personality and Social Psychology, 70(3), pp. 614-636.

Campbell, N., 2000. Databases of emotional speech. In: Proc. ISCA Workshop on Speech and Emotion, Northern Ireland, pp. 34-38.

Cauldwell, R. T., 2000. Where did the anger go? The role of context in interpreting emotion in speech. In: Proc. ISCA Workshop on Speech and Emotion, Northern Ireland, pp. 127-131.

Cowie, R., 2000. Describing the emotional states expressed in speech. In: Proc. ISCA Workshop on Speech and Emotion, Northern Ireland, pp. 11-18.

Cowie, R., Douglas-Cowie, E., Apolloni, B., Taylor, J., Romano, A., & Fellenz, W., 1999. What a neural net needs to know about emotion words. In: Mastorakis, N. (Ed.), Computational Intelligence and Applications, World Scientific & Engineering Society Press, pp. 109-114.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W. & Taylor, J., 2001. Emotion Recognition in Human-Computer Interaction. In: IEEE Signal processing Magazine, 18 (1), p. 32-80.

Dietz, R. B. & Lang, A., 1999. Æffective Agents: Effects of Agent Affect on Arousal, Attention, Liking & Learning. In: Proc. 3rd Intl. Cognitive Technology Conference, http://www.added.com.au/cogtech/CT99/Dietz.htm.

Douglas-Cowie, E., Cowie, R., & Schröder, M., 2000. A new emotion database: Considerations, sources and scope. Proc. ISCA Workshop on Speech and Emotion, Northern Ireland, pp. 39-44.

Drosdowski, G.,1989. Duden Herkunftswörterbuch, Etymologie der deutschen Sprache [Duden etymological dictionary of German], 2nd ed. Bibliographisches Institut & Brockhaus, Mannheim.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vrecken, O., 1996. The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In: Proc. ICSLP 96, pp. 1393-1396.

Ehlich, K., 1986. Interjektionen. Max Niemeyer Verlag, Tübingen.

Ekman, P., 1982. Emotion in the human face. New York: Cambridge University Press.

Gobl, C. & Ní Chasaide, A., 2000. Testing affective correlates of voice quality through analysis and resynthesis. In: Proc. ISCA Workshop on Speech and Emotion, Northern Ireland, pp. 178-183.

Johnstone, T., Banse, R., & Scherer, K. R., 1995. Acoustic profiles in prototypical vocal expressions of emotion, In: Proc. ICPhS 95, Stockholm, Vol. 4, pp. 2-5.

Kienast, M., Paeschke, A., & Sendlmeier, W., 1999. Articulatory reduction in emotional speech. In: Proc. Eurospeech 1999.

Ladd, D. R., Silverman, K. E. A., Tolkmitt, F., Bergmann, G., & Scherer, K. R., 1985. Evidence for the independent function of intonation contour type, voice quality, and F0 range in signalling speaker affect. In: J. Acoust. Soc. Am. 78(2), pp. 435-444.

Leinonen, L., Hiltunen, T., Linnankoski, I., & Laakso, M.-L., 1997. Expression of emotional-motivational connotations with a one-word utterance. Journal of the Acoustic Society of America, 102(3), pp. 1853-1863.

McQueen, J. M., & Cutler, A., 1997. Cognitive Processes in Speech Perception. In: Hardcastle, W. J. & Laver, J. (Eds.), The Handbook of Phonetic Sciences, Blackwell, Oxford, UK, Cambridge, MA, pp. 566-585.

Mozziconacci, S. J. L., 1998. Speech Variability and Emotion: Production and Perception, PhD Thesis, Technical University Eindhoven.

Murray, I. R. & Arnott, J. L., 1993. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. In: J. Acoust. Soc. Am. 93(2), pp. 1097-1108.

Ohala, J. J., 1994. The frequency code underlies the sound-symbolic use of voice pitch. In: Hinton, L., Nichols, J., & Ohala, J. J. (Eds.), Sound Symbolism, Cambridge UP, Cambridge, pp. 325-347.

Ortony, A. & Turner, T. J., 1990. What's Basic About Basic Emotions? In: Psychological Review 97(3), pp. 315-331.

Paeschke, A. & Sendlmeier, W. F. (2000). Prosodic Characteristics of Emotional Speech: Measurements of Fundamental Frequency Movements. In: Proc. ISCA Workshop on Speech and Emotion, Northern Ireland, pp. 75-80.

Pereira, C., 2000. Dimensions of emotional meaning in speech: In: Proc. ISCA Workshop on Speech and Emotion, Northern Ireland, pp. 25-28.

Scherer, K. R., 1986. Vocal Affect Expression: A Review and a Model for Future Research. Psychological Bulletin, Vol. 99, No.2, pp. 143-165.

Scherer, K. R., 1988. On the symbolic functions of vocal affect expression. Journal of Language and Social Psychology, 7, pp. 79-100.

Scherer, K. R., 1994. Affect Bursts. In: van Goozen, S. H. M., van de Poll, N. E., Sergeant J. A. (Eds.), Emotions, Lawrence Erlbaum, Hillsdale, NJ, pp. 161-193.

Scherer, Klaus R. & Ceschi, G., 2000. Criteria for Emotion Recognition From Verbal and Nonverbal Expression: Studying Baggage Loss in the Airport. In: Personality & Social Psychology Bulletin, 26(3), pp. 327-339.

Scherer K. R., Ladd, D. R. & Silverman, K. E. A., 1984. Vocal cues to speaker affect: Testing two models. In: J. Acoust. Soc. Am. 76(5), pp. 1346-1356.

Schröder, M., 1999. Can emotions be synthesized without controlling voice quality? In: Phonus 4, Research Report of the Institute of Phonetics, University of the Saarland, pp. 37-55.

Schröder, M., Aubergé, V., & Cathiard, M.-A., 1998. Can we hear smiles? In: Proc. ICSLP 98, Sydney.

Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M. & Gielen, S., 2001. Acoustic Correlates of Emotion Dimensions in View of Speech Synthesis. In: Proc. Eurospeech 2001, Aalborg, Vol. 1, pp. 87-90.

Stallo, J., 2000. Simulating Emotional Speech for a Talking Head. Honours Thesis, School of Computing, Curtin University of Technology, Australia.

Wells, R., 1945. The pitch phonemes of English, Language, 21, pp. 27-40.

Wichmann, A. (2000). The attitudinal effects of prosody, and how they relate to emotion. In: Proc. ISCA Workshop on Speech and Emotion, Northern Ireland, pp. 143-147.

Zerling, J.-P., 1995. Onomatopées et interjections en français. Travaux de l’Institut de Phonétique de Strasbourg, 25, pp. 95-109.

TABLES

Table 1

Contributions of the individual speakers to the ten emotion categories. Speakers MA, MI and SB are male, speakers CG, SJ and GS are female.

	Actors				Non-actors		Total
	CK	MA	MI	SJ	GS	SB	
admiration	2	2	2		2		8
threat	3	2	2[6]	1			8
disgust	2	1	1	2	2		8
elation	1	2	2	2		1	8
boredom	2	2	1	1		2	8
relief	2	2	2		1	1	8
startle	1	1	2	3		1	8
worry	2	1	2	1	2		8
contempt	1	1	1	2	2	1	8
anger	2	3		1	1	1	8
Total	18	17	15	13	10	7	80

Table 2

Mean recognition rates in the listening test for the ten emotion categories.

Admiration	Threat	Disgust	Elation	Boredom
90.6%	62.9%	93.1%	79.4%	72.5%
Relief	Startle	Worry	Contempt	Anger
92.5%	88.8%	84.4%	84.4%	60.6%

Table 3

Listening test confusion matrix for the ten emotion categories.

Intended Emotion	Perceived Emotion									
	admiration	threat	disgust	elation	boredom	relief	startle	worry	contempt	anger
admiration	**91**%			5%		3%				1%
threat	2%	**63**%		1%			3%	1%	5%	**24**%
disgust			**93**%		1%	1%	1%	1%	3%	1%
elation	2%	1%		**79**%		**15**%	2%	1%		
boredom			1%		**73**%	5%	1%	**16**%	4%	1%
relief	1%				3%	**93**%		4%	1%	
startle	1%		1%	1%		4%	**89**%	3%	1%	
worry	1%				9%	1%	4%	**84**%	1%	
contempt	1%			6%	2%	6%			**84**%	1%
anger	3%	**14**%	1%	1%	2%	3%	3%	6%	6%	**61**%

Table 4

Affect burst classes within each intended emotion. The 'emotion recognised' columns indicate the most frequent answer for that affect burst in the respective test ('✓' = intended emotion). Recognition rates are given for that most frequent answer. 'int. breath' designates a rapid intake of breath.

Intended Emotion	Affect Burst Class	Expert transcription		Listening test			Written perception test			
		segments, voice qual.	intonation	No. of Stimuli	Emotion recognised	Recogn. Rate	Orthographic transcription	Transcr. variability	Emotion recognised	Recogn. Rate
admiration	Wow	[waːw]	3-1	4	✓	91%	wow	3	✓	90%
	Boah	[bo̤aː]	1	4	✓	90%	boah	2	✓	90%
threat	Hey	[hɛɪ]	3-2	5	✓	81%	ej	3	✓	65%
	growl	[m̰ː]	1	2	anger	80%	mrr	8	anger	50%
disgust	Buäh	[bʊæ̤ː]	3-2	6	✓	92%	uäh	1	✓	63%
	Igitt	[iːgɪtʰ]	3	1	✓	100%	igitt	1	✓	100%
	Ih	[ɪːə̤]	3-2	1	✓	95%	irgh	29	✓	84%
elation	Ja	[jaː]	3	4	✓	69%	jaaa	2	✓	47%
	Yippie	[jɪpiː]	4-3	2	✓	100%	jippii	1	✓	100%
	Hurra	[huʁaː]	4-3	2	✓	80%	hurra	0	✓	95%
boredom	yawn		3-1	4	✓	81%	uuahh	20	startle	53%
	sigh	[ə̤ː]	2-1	2	✓	45%	hmm	12	✓	63%
	Hmm	[mː]	1-2	2	✓	83%	mmh	7	✓	60%
relief	sigh	[a̤ː]	2-1	3	✓	85%	ahh	5	✓	50%
	Uff	[ʊfː]	2-1	3	✓	98%	uff	6	✓	80%
	Puh	[pʰṳɸː]	3-1	2	✓	95%	puh	1	✓	85%
startle	int. breath		3	6	✓	92%	he	8	threat	40%
	Ah	[a]	3	2	✓	80%	a	8	relief	37%
worry	Oje	[o̤jeː]	2-1	4	✓	96%	ujeh	6	✓	75%
	Oh-Oh	[ʔoʔːoː]	3-2	2	✓	85%	o-oh	6	✓	67%
	Oweh	[o̤ːβe̤ː]	3-1	1	✓	50%	oh jee	14	✓	85%
	Hmm	[m̰m]	2-1	1	✓	70%	hmm	9	boredom	63%
contempt	laughter	[ħə̤ħ]	1	5	✓	77%	hähä	10	✓	74%
	Pha	[phaʔ]	1	2	✓	95%	pah	4	✓	95%
	Tse	[tsʰə]	3	1	✓	100%	tse	5	✓	85%
anger	growl	[m̰ː]	2-1	4	✓	69%	ahr	2	✓	39%
	breath out	[ħː]	1	3	✓	55%	chrr	8	✓	39%
	Oh	[ɔ̤ː]	2-1	1	✓	45%	ooh	4	admiration	53%

Table 5

Major confusions for less well recognised affect burst classes (recognition as intended in ≤ 70% of the cases).

Affect burst	Perceived emotion						
	threat	elation	bored	relief	worry	anger	other
threat growl	**18**%				3%	**80**%	
anger growl	**10**%		1%	6%	3%	**69**%	11%
anger breath out	**23**%		5%		7%	**55**%	10%
anger Oh	5%	**10**%			**15**%	**45**%	**25**%
boredom sigh			**45**%	10%	**40**%		5%
worry Oweh			**40**%	5%	**50**%		5%
worry Hmm			**25**%		**70**%		5%
elation Ja		**69**%		**29**%			3%

Table 6

Mean recognition rates in the written perception test for the ten emotion categories.

Admiration	Threat	Disgust	Elation	Boredom
90.0%	44.7%	82.8%	81.4%	33.3%
Relief	Startle	Worry	Contempt	Anger
71.7%	23.1%	63.6%	84.5%	30.9%

Table 7

Means and standard deviations for correct categorical ratings of the ten emotions, on the three seven-point scales arousal (from 1=calm to 7=excited), valence (from –3=negative to 3=positive) and control (from 1=subordinate to 7=dominant).

	arousal		valence		control	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
admiration	4,8	1,1	1,6	0,9	4,5	1,4
threat	5,0	1,1	-1,3	1,1	5,5	1,2
disgust	5,0	1,1	-2,0	0,9	4,0	1,2
elation	6,1	0,8	2,4	0,8	5,0	1,2
boredom	2,5	1,2	-0,8	1,1	4,2	1,0
relief	4,1	1,4	1,0	1,3	3,9	1,1
startle	6,0	0,9	-1,5	0,9	2,9	1,2
worry	4,0	1,4	-1,5	0,9	3,1	1,3
contempt	3,9	1,2	-0,9	1,5	5,3	1,2
anger	5,2	1,3	-1,8	1,0	4,4	1,4

FIGURES

Figure captions:

Figure 1. Illustration of the relationship between affect bursts, affect emblems, and raw affect bursts, and the fuzzy borders with verbal interjections and physiological sounds.

Figure 2. Mean position of the ten emotion categories on the arousal, valence, and control scales (based on categorically correct ratings, see text).

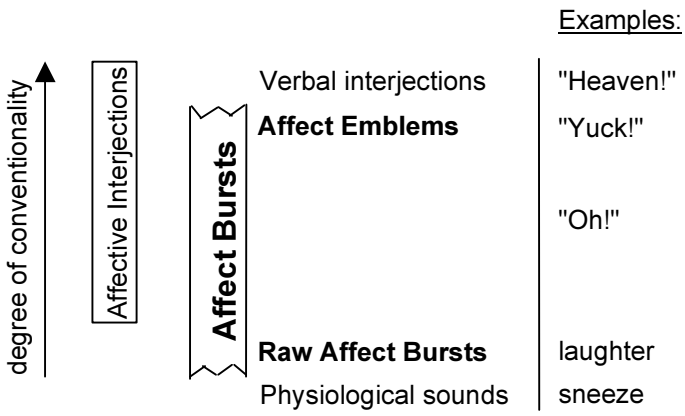
Examples:
degree of conventionality
Affective Interjections
Affect Bursts
Verbal interjections
"Heaven!"
Affect Emblems
"Yuck!"
"Oh!"
Raw Affect Bursts
laughter
Physiological sounds
sneeze

Figure 1

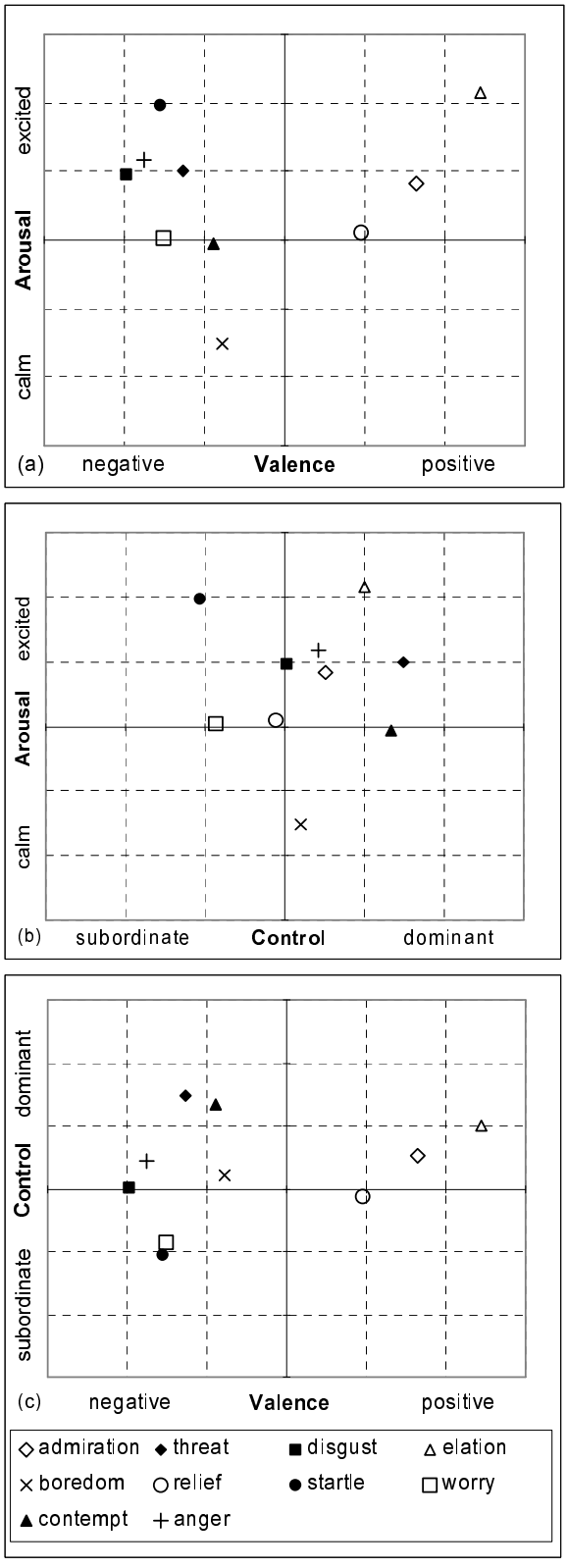
excited
Arousal
calm
(a) negative Valence positive
excited
Arousal
calm
(b) subordinate Control dominant
dominant
Control
subordinate
(c) negative Valence positive
admiration
threat
disgust
elation
boredom
relief
startle
worry
contempt
anger

Figure 2

Footnotes

---

[1] The themes of the frame stories were as follows: For admiration, the speaker is delighted at the sight of the beautiful evening dress of a good friend. For threat, the speaker has to watch children in the schoolyard, chasing the boys who are once more trying to bother the girls. For disgust, the speaker discovers a large, hairy, black, moving worm in his/her food[1]. For elation, the speaker's team has just won the gold medal in a sports competition. For boredom, the speaker has been sitting for two hours with someone talking about an uninteresting subject. For relief, the speaker relaxes in an armchair after a day of hard but successful work. For startle, the speaker, lost in thought, suddenly becomes aware of the tall, dark silhouette of a person standing behind him/her. For worry, the speaker has just heard about the financial difficulties of his/her business. For contempt, the speaker rejects the apology of a former friend who had shamefully betrayed the speaker. Finally, for (hot) anger, the speaker is furious about a person he/she dislikes, because the person is late once more and shows no sign of regret.

[2] The vocalisation 'Ja' ("yes") for elation does not actually fulfill the criteria for the working definition of affect bursts, given in the introduction, because of its verbal nature. However, it was spontaneously produced by 4 out of 6 speakers.

[3] Two affect burst classes, 'worry Hmm' and 'boredom sigh', were identically transcribed as "hmm". Therefore, in the written perception test, only one occurrence of "hmm" was used.

[4] This question, although interesting, is beyond the scope of this study.

[5] A different explanation for the high transcription variability may hold for ‘worry Oweh’, however: It contains a bilabial approximant [β] which is not a German phoneme, probably an imperfect realisation of the verbal interjection “Oh weh” /ove:/. The variability would be due to the slurred realisation of the phonemic structure rather than the absence of an appropriate mental pattern in the transcribers’ mind.

[6] After the listening test, it was discovered that due to a programming error, one of the threat stimuli from speaker MI had not been used in the listening test, reducing the number of stimuli for the threat category to 7 and the total number of stimuli to 79. The unused threat stimulus was a ‘Hey!’ affect burst (see Table 4).