



ELSEVIER

Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE @ DIRECT®

Speech Communication 40 (2003) 5–32

SPEECH
COMMUNICATION

www.elsevier.com/locate/specom

Describing the emotional states that are expressed in speech

Roddy Cowie^{a,*}, Randolph R. Cornelius^b

^a School of Psychology, Queen's University, BT7 INN Belfast, Northern Ireland, UK

^b Vassar College, Poughkeepsie, NY, USA

Abstract

To study relations between speech and emotion, it is necessary to have methods of describing emotion. Finding appropriate methods is not straightforward, and there are difficulties associated with the most familiar. The word emotion itself is problematic: a narrow sense is often seen as “correct”, but it excludes what may be key areas in relation to speech—including states where emotion is present but not full-blown, and related states (e.g., arousal, attitude). Everyday emotion words form a rich descriptive system, but it is intractable because it involves so many categories, and the relationships among them are undefined. Several alternative types of description are available. Emotion-related biological changes are well documented, although reductionist conceptions of them are problematic. Psychology offers descriptive systems based on dimensions such as evaluation (positive or negative) and level of activation, or on logical elements that can be used to define an appraisal of the situation. Adequate descriptive systems need to recognise the importance of both time course and interactions involving multiple emotions and/or deliberate control. From these conceptions of emotion come various tools and techniques for describing particular episodes. Different tools and techniques are appropriate for different purposes.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Speech; Emotion; Categories; Dimensions; Instruments

1. Introduction

The past few years have seen growing interest in what is generally described as ‘speech and emotion’. As the phrase indicates, the research deals with relationships between two domains. To understand the relationships between them, the developing research area needs access to appropriate ways of describing each one individually. This paper is about the ways of describing emotion that might be useful to research in the area. The task is not routine, because there is no generally accepted

methodology for describing emotion. Instead, the literature from related disciplines—primarily psychology and biology—contains a variety of alternative approaches and techniques (Cornelius, 1996). The alternatives persist because each of them has useful features, but none of them quite captures the whole picture, and there is as yet no truly integrative approach on the horizon.

The net result is that the literature on emotion offers the speech community a range of tools that could potentially relate to its concerns. The exact value of any given tool is rarely obvious, though. Our basic aim in this paper is to promote informed choice by ensuring that the key options are in full view. That includes making it clear that the best known options should not be regarded as

* Corresponding author. Tel.: +44-28-90-274354.

E-mail address: r.cowie@qub.ac.uk (R. Cowie).

automatic choices, because there are substantial problems associated with them.

Beyond that, we argue that it is worthwhile for the speech community to engage actively with the problem of describing emotion, rather than expecting to extract solutions ready-made from other disciplines. The point is not that speech research should isolate itself from other research on emotion, but, rather, that it should consider itself capable of making a distinctive contribution, which can and should be fed back into the mainstream of research on emotion.

The paper divides into three broad parts. The first (Sections 2–4) considers what is to be described, and why. The second (Sections 5 and 6) introduces the conceptual issues associated with key approaches to describing relevant states. The third describes specific techniques arising out of these approaches.

2. Definitions

This section is concerned with defining key terms, particularly those that describe the broad areas and themes associated with emotion. Definition is both important and difficult because the everyday word ‘emotion’ is notoriously fluid in meaning—as noted by distinguished scholars from James (1884) at the end of the nineteenth century to James Russell at the end of the twentieth (see Russell and Barrett-Feldman, 1999). As a result, phrases that use it may be open to quite different interpretations. ‘Speech and emotion’ is a case in point. Inappropriate ways of resolving the ambiguity can produce a mismatch between the topics that investigators intuitively consider interesting and those they feel bound to study.

In the meaning that people tend to feel is central, emotions are episodes that are relatively brief and highly distinctive. There is no generally agreed, compact term for these episodes. Terms such as ‘primary emotions’ (Plutchik, 1984), ‘basic emotions’ (Stein and Oatley, 1992), ‘modal emotions’ (Scherer, 1994), and ‘acute emotions’ (Lazarus, 1994) are related, but they carry specific theoretical implications (see Section 5.3). We have chosen Scherer’s term ‘full-blown emotion’ (1999) as a

neutral way of referring to episodes that would be widely regarded as prime examples of emotion, without prejudging theoretical issues (Cowie et al., 2001).

There is reasonable agreement that episodes of full-blown emotion are syndromes, which tend to involve elements of at least the following kinds.

Affect is ‘the irreducible aspect that gives feelings their emotional, non-cognitive character’ (Frijda, 1993, p. 383). It is used here in that specific sense, rather than as a near-synonym of emotion (e.g., Corsini, 1994).

Appraisal is a categorisation of an event or situation that marks it as having a distinctive kind of significance for the organism, including a valence (Arnold, 1960).

Valence is a quality of being (subjectively) charged with positive or positive or negative significance, which is inherent in emotional appraisal and experience.

Action readiness may take the form either of generalised activation or a disposition to follow preset courses of action related to the appraisal (Frijda, 1993, p. 383).

Communicative behaviour conveys information about affect, appraisal and action tendency to other individuals, often without deliberate intention.

Physiological adjustments may be related to arousal and/or autonomic nervous system activity.

There is debate about other aspects of emotion syndromes. Broadly speaking, the psychoevolutionary tradition (Darwin, 1872; Plutchik, 1984; Ekman, 1999) proposes that full-blown emotions are directly linked to systems that evolved to deal with biologically significant situations; and they are therefore universal. In contrast, social constructivism regards all emotions as cultural products that owe their meaning and coherence to learned social rules (e.g., Averill, 1980; Harré, 1986).

A different conception of emotion emerges in statements such as “every way of experiencing the world involves a stance that is emotional” (Lazarus, 1999b, p. 11); or “we are always ... happy, angry, and so forth, to some extent ... one can always ask how loving, attached, or affiliative one feels” (Buck, 1999, p. 325). In that context, the

word ‘emotion’ refers to an aspect of most (or all) mental states, rather than to a special kind of state. We have used the phrase ‘underlying emotion’ to convey that sense (Cowie et al., 2001). For example, it would be wrong to say that friendliness is an emotion; but saying that a person is friendly would usually be taken to imply positive underlying emotion.

It is useful to have a term to cover all states that are considered to involve true emotion, be it full-blown, underlying, or of any other variety (examples might be suppressed emotion, mixed emotion, etc.). We have used the term ‘emotional states’ for that purpose (Cowie et al., 2001). In a similar vein, we have used the phrase ‘emotion-related states’ for states that people feel are not emotions proper, but are sufficiently similar that ‘emotion’ and derived terms are applied loosely to them in everyday speech. Considering full-blown emotions as syndromes suggests why this makes sense psychologically: various states appear to include some of the elements associated with full-blown emotion, but not all of them. Key examples, such as moods, states of arousal, and attitudes, are considered in later sections.

When people talk informally about speech and emotion, their interest often seems to be in the possibility that speech conveys underlying emotion, or emotion-related states. It would be unfortunate if they felt that those interests had to be left outside the laboratory door because the term emotion properly referred only to full-blown emotion. It is a serious question whether research is better to address a narrowly defined field, consisting primarily of full-blown emotion; or a broader one, including emotional and emotion-related states. Definitions of the kind introduced in this section ensure that the decision is not preempted by the vagaries of semantics. Section 4 considers evidence that does appear relevant.

3. Choices of orientation

The descriptive tasks facing research on speech and emotion depend not only the subject’s boundaries, but also on the priorities that it adopts. This section lists key choices of this kind.

3.1. Practical and theoretical goals

The literature reflects at least three types of research goal that have a bearing on descriptive needs.

Psychological theory: Research may aim to help psychology and biology to understand the essential nature of emotion. It so, it is almost bound to use the same descriptive schemes as the other disciplines involved.

Applications: Research may be concerned with specific applied problems. In that case, the problem may strongly influence the choice of descriptive schemes—for instance, because it has to describe emotions in a way that is accessible to non-expert users, or because the task requires a coarse distinction that cuts across theoretically important lines (e.g., ‘any emotion or emotion-related state that might impair concentration’).

Ecological: Research may adopt an ‘ecological’ stance, in the sense advocated by Gibson (1979) and Neisser (1976); i.e., prioritising the study of commonplace, everyday expressions of emotion and emotion-related phenomena in speech. Such an approach requires descriptive systems that apply effectively to the states that are typical of everyday discourse. Section 4 takes up some of the implications of that point.

It would be ideal if a single descriptive system met the requirements of all these orientations. But that kind of system appears to be a future goal rather than an immediate reality; and for the present it seems better to accept a pragmatic, pluralistic approach than to insist, for instance, that applied research is deficient if it fails to observe distinctions drawn from psychological theory.

3.2. Cause- and effect-type orientation

Given a sample of speech that has features associated with emotion, research may be concerned either with the emotional state that led the speaker to produce that kind of speech, or with the impression of the speaker’s state that a listener is likely to form on the basis of the speech. We have described research with the first goal as

cause-oriented, and research with the second goal as effect-oriented (Cowie et al., 2001).

Again, the two concerns lead to different descriptive tasks. For cause-oriented research, descriptions need to capture distinctions that psychology or biology recognise as valid. In contrast, effect-oriented research needs descriptions that capture perceivers' impressions of emotionality. The two types of description need not correspond simply. One might assume that if they did, the task of understanding the objective nature of emotion would not have proved as difficult as it has.

3.3. *Identifying units of analysis*

The task of defining units of analysis can be approached from two contrasting directions. One approach is to look for speech correlates of a given type of emotional phenomenon; the other is to ask what kind of information about emotion is carried by a given type of speech phenomenon.

The choice would not matter if it could be assumed that there must be a simple mapping between boundaries in the domain of speech and boundaries in the domain of psychology. However, there is no logical guarantee that such a mapping exists, and there are specific reasons to suspect that it might not. If full-blown emotions involve multiple elements, speech features could well be associated with one element of the syndrome rather than with the whole. If so, the same speech features would appear in a range of states, emotional and quite possibly emotion related. An element that seems quite likely to have direct correlates in speech is activation (see Section 6.2). Conversely, opposite speech patterns have been reported in studies concerned with the same emotion category (Cowie et al., 2001). That could be due to flawed research techniques, but it could also indicate that states within a single standard category divide into subtypes that are more sharply distinguished at the level of speech than in other respects.

The most reasonable conclusion is probably that research on speech and emotion should consider both approaches, and see its task as discovering structure in the domain of emotion that

maps onto structure in the domain of speech. That is a basic reason why the paper emphasises the range of ways in which emotion can be described: it is not obvious in advance which will relate most satisfyingly to speech.

3.4. *Precision and tractability*

Clearly one requirement of an ideal descriptive system is sufficient precision to ensure that states which are genuinely different have different descriptors. However, a satisfactory descriptive system also needs to allow tractable research tasks to be formulated. The two tend to be in tension. For example, everyday language at least seems capable of capturing the emotion that people express in their speech quite accurately; but the number of categories involved would mean that very large amounts of data were needed to reach statistically reliable conclusions about the features associated with each one (see, e.g., Stibbard, 2001). Solving that problem is one of the central challenges in the development of descriptive systems.

3.5. *Consistency and innovation*

It is clearly desirable to achieve consistency in the way that emotion is described. Its absence currently poses problems at multiple levels: it is difficult to integrate data when different studies use different sets of categories, and those they share may have been operationalised in different ways. However, standardisation can be destructive if it means that genuinely unresolved issues may not be raised. As Salomon (2001, p. 6) has said, "there is a very real danger that a taxonomy that sorts all this out properly could also turn into an obstacle to progress". In our view, the natural goal at this stage is shared awareness of the uncertainties that surround description, and one of the aims of this paper is to encourage that awareness.

3.6. *Conclusion*

It is a truism that the value of a tool depends on the purpose for which it is needed. Descriptions are no exception to the rule, especially when the

material to be described has multiple dimensions, and can appear to be structured quite differently from different viewpoints. Emotion appears to be a phenomenon of that kind, and it seems unlikely that a single descriptive system to meet all needs will emerge in the near future.

4. Emotional states that are conveyed in speech

From an ecological perspective, one of the requirements for a satisfactory descriptive system is that it should apply well to the emotional and emotion-related states that are commonly expressed in speech. Emphasising applications has partly similar results: there may be practical reasons for considering some states that are rare, but on the whole, states that are common are more likely to be practically important.

Traditionally, those points were overshadowed by what we have called ‘the benign interpolation hypothesis’ (Douglas-Cowie and Cowie, 2001). It is that the whole space of emotions can be reconstructed by relatively straightforward inference from information about a few well-selected categories, which serve as points of reference. On that hypothesis, there is no need for intensive research into states other than the reference categories. However, there are reasons to doubt the hypothesis. In the field of emotion research generally, attempts to identify the reference categories have encountered severe difficulties (see Section 5). In the particular field of emotion and speech, there is increasingly strong evidence that results based on laboratory research with archetypal states transfer poorly to real applications (see Batliner et al., 2003).

Under those circumstances, there is growing pressure to investigate the kinds of emotional phenomena that are regularly associated with speech; and with that, pressure to find descriptive machinery that applies well to those phenomena. There is still limited data on the prevalence and nature of emotional speech in natural settings, but pointers can be drawn from recent collections of natural emotional speech, the Leeds–Reading database and the Belfast database (Douglas-Cowie et al., 2003).

The aim of the Belfast project was to collect samples of speech that was clearly emotional from spontaneous discourse (mainly dialogue). Material of the kind that was expected proved strikingly elusive. Unexpectedly, small-group discussions on emotive topics produced very little speech suggesting anything approaching full-blown emotion (Cowie et al., 1999a). One-on-one discussions between friends about emotive life events produced a few examples, but still very few. Rather similarly, the Leeds–Reading project began with the expectation that pre-existing interviews on emotive topics would yield speech with emotional characteristics, and found so little that they turned to other sources.

The material that eventually formed the Belfast database came mainly from TV interviews and chat shows. The sources were used because they provided a substantial amount of speech involving spontaneous and relatively strong emotion. The database was composed of short extracts (‘clips’) chosen either because they provided comparatively clear examples of emotional discourse, or because they showed a speaker who had figured in an emotional clip in a relatively neutral state. Preliminary analysis indicates that even in that doubly selected material, speech expressing pure full-blown emotion was rather rare.

Three independent raters assigned verbal labels to the clips from the database. The procedure was that each rater was required to choose the one label from a list of 16 that best described the main emotion in the clip, and was free to choose up to two additional labels if the first was not felt to be sufficient. For each label, the strength of the emotion concerned was also rated strong, medium, or weak.

Despite the efforts made to find material showing strong emotion, the predominant emotion was rated ‘strong’ in only 34% of clips chosen for emotionality, ‘medium’ in 41%, and weakly emotional or neutral in 25%. A related point is made by the use of secondary labels, summarised in Table 1. The key observation involves the column headed “none”, which records cases where a rater regarded one label as a sufficient descriptor of the predominant emotion in a clip. These cases were quite rare. In fact, when a single label was

Table 1

Main labels assigned by three independent raters to clips in the PHYSTA database

First choice label		Second choice label			
Label	<i>n</i>	None	Negative—disapproving, annoyed, resentful, disgusted, hurt, despairing, angry, anxious, disappointed, sad	Control related—calm, confident, nervous, surprised	Positive—proud, happy, excited, pleased, embarrassed, relaxed, content, amused, hopeful, relieved, loving, serene
Neutral	272	43	89	29	55
Angry	114	7	89	9	1
Sad	94	8	59	5	4
Worried	19		15	3	
Disappointed	17		11		
Afraid	13	1	8	3	
Amused	26	10	1	1	14
Pleased	44	3	3	11	25
Happy	37			5	27
Excited	17		2	6	8
Confident	13		3	1	7
Interested	12		2	1	7
Affectionate	10		1	1	6
Loving	3				3
Content	3				2
Relaxed	3			1	1
Bored	4		3		

considered sufficient, it was usually “neutral”. It can be seen that the secondary label did not generally come from a completely different region of the emotional spectrum: it tended to have a similar broad thrust. Nevertheless, the first label was not often regarded as an adequate description.

Research using the Leeds–Reading database has reported related findings (Greasley et al., 2000). Subjects listened to naturalistic speech samples drawn from it, and applied one of five labels (happiness, sadness, anger, fear, disgust). Even with such a restricted forced choice task, only 54% of samples associated with negative emotions gave rise to general agreement (over 2/3 of ratings in the modal class); and even those gave rise to mixed responses when finer gradations in responding were allowed.

Such findings reinforce the suspicion that the emotion expressed in speech tends to be weak or moderately strong, and mixed rather than pure. The material also hints at a reason why that may be so.

There is a common stereotype that pure emotion leaves people either speechless or incoherent. If that were so, it would be unsurprising that

people who are actually talking tend to be in emotional states that are moderate and mixed. Theoretically, the stereotype makes some sense: it suggests that full-blown emotion competes for control with the cognitive systems that underpin fluent speech. That appears to be consistent with the kind of multi-level approach that various groups have been developing, in which emotional and propositional functions are associated with distinct systems (e.g., Teasdale, 1999).

To test the intuition that emotion is associated with partial or complete breakdown of speech as such, an independent rater evaluated each clip of the Belfast database, using a real time computer interface to indicate which of five categories the speech belonged to at any given instant: clear, non-overlapping speech; speech obscured by overlap with another person; “non-speech”, i.e., sound production that seemed not to be intended as verbal (sighs, sobbing, inarticulate exclamations, etc.); silence; and speech that was too incoherent to be easily understood. For want of a better term, the last four categories are described as “degraded communication”. Fig. 1 shows the ratio of time in each of the ‘degraded’ categories to time in clear,

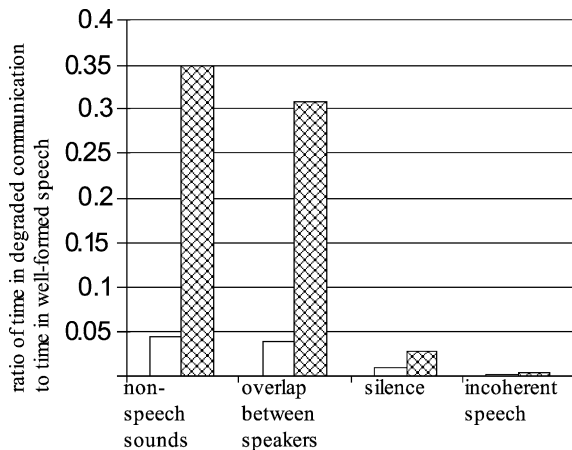


Fig. 1. Degraded communication in the PHYSTA database: ratios of time in various forms of degraded communication to time in clear, non-overlapping speech, for clips selected to represent relatively unemotional speech (white columns) and emotional speech (shaded columns).

non-overlapping speech. It was consistently higher for clips that were included as examples of emotional speech than for clips that were included as examples of relatively unemotional speech.

Such evidence does not prove that there is a conflict between strong emotion and fully fluent speech, but it indicates that the intuition is worth investigating. Its relevance here lies in a series of corollaries. If it is true, then when signs of emotion are embedded in speech rather than disrupting it, the emotion in question is likely to be moderate and controlled. Conversely, the speech patterns associated with strong emotions may offer rather little guidance to those that occur with moderate and controlled emotion, because they are dominated by a conflict that is not a major problem in moderate and controlled emotion.

A final point to be drawn from the Belfast database is that even if speech rarely expresses strong, pure emotion, it is not often emotionally neutral either. Although we set out to find emotionally neutral episodes for every speaker in the database, 84% of cases where a clip was rated neutral included a supplementary label indicating that some emotion was felt to be present.

Other sources extend the argument. There is extensive evidence that speech is a significant

source of information about emotion-related states that are both common and practically important.

Stress is a clear example. Lazarus (1999a, p. 35) has described it as an ‘absurdity’ that stress and emotion tend to be separated in the scientific literature, arguing that ‘When there is stress there are emotions ... when there are emotions, even positively toned ones, there is often stress too...’ It is well established that voice provides evidence of stress (Tolkmitt and Scherer, 1986; Womack and Hansen, 1996; Fernandez and Picard, 2003). Unsurprisingly, the signs of stress that are most regularly described (raised mean pitch and pitch range, increased intensity and hesitation) are also regularly cited as signs of emotion. At the other end of the spectrum there is also evidence that speech carries signs of states associated with low arousal, notably boredom (see Cowie et al., 2001 for a summary) and drowsiness (Hadfield and Marks, 2000).

Another large body of literature indicates that speech carries information about speakers’ attitude (see Schubiger, 1958; Crystal, 1969, 1975; O’Connor and Arnold, 1973). Linguists have often insisted that attitude should not be confused with emotion (e.g., Wichmann, 2000), and that is reasonable if the latter term is used to mean full-blown emotion. But there should be no doubt of the connection between at least some attitudes and underlying emotion—for example, ‘friendliness’ would at least be expected to imply some underlying emotional warmth. At first sight, the literature suggests that attitude and emotionality may well be associated with qualitatively different kinds of vocal expression (Cowie et al., 2001). If so, it is a result well worth verifying. However, the impression may arise mainly because different research traditions have tended to study the two areas, and have used different descriptors for speech.

In summary, it seems certain that there are speech-related phenomena to be studied in the area of mild and mixed emotion, and beyond that in emotion-related areas such as stress and attitude. It is less certain how much speech occurs during full-blown emotion, and whether it tends to involve rather special kinds of effect that limit the

prospects of extrapolating from it. Reasons may emerge for dividing research on speech and emotion proper from research on speech and stress or attitude. But sensible separation cannot take place until boundary areas have been studied to establish where any natural distinctions lie. That is a practical issue as well as a theoretical one. A system for detecting stress by voice needs to take account of the possibility that emotions such as happiness may give many similar signs.

Our title reflects the points made in this section. Research on speech and emotion needs descriptive tools that apply to the emotional and emotion-related states that speech actually conveys on a reasonably regular basis. That entails being wary of descriptive schemes that set obstacles to studying ecologically and practically significant states, either because they imply that those states ought not to be included in the research domain, or because they are not well suited to the task. Conversely, it provides an incentive to explore less familiar descriptive schemes that might apply well to them.

5. Category labels

This section is the first of two concerned with the conceptual basis of particular approaches to description. It considers the most familiar way of describing specific emotions, which is to use category labels derived from everyday language. The section after that considers alternative ways of describing emotion.

Because they are so familiar, it is easy to assume that category labels drawn from everyday language are self-evidently the correct way to describe emotion. However, transforming everyday categorical descriptions into an effective research tool is at the least a challenge. The aim of this section is to consider what kind of resource everyday categorization is, and what kinds of problem surround its use.

5.1. *Everyday terms for emotions*

Everyday language contains a large number of words for states that people tend to agree refer to

emotions in a strict sense. Research on these categories reveals a range of complexities and uncertainties that bear on their use in speech research.

The first problem is how to decide which terms qualify as genuinely emotional. Two main solutions have been proposed, one quantitative and the other qualitative. Quantitative approaches were pioneered by Fehr and Russell (1984). Following Rosch's account of semantics (Rosch and Mervis, 1975), they argued that a state counts as an emotion if the associated concept is close enough in conceptual space to the centre of the class, which is marked by prototypical examples of emotions. The boundary of the class is assumed to be fuzzy. From this conception have flowed attempts to measure the typicality of emotion-related concepts. In contrast, the qualitative approach looks for logical criteria that distinguish emotions from related states. Ortony et al. (1988) have been its most prominent champions.

The approaches lead to different views of what constitutes an emotion. For instance, Zammuner (1998) proposed on the basis of quantitative analysis that people classify terms as emotional in so far as they relate them to valence, and to arousal that differs from the default level. In contrast, Ortony et al. (1988) propose a stepwise partition. To say that a term refers to an emotion, they require it to describe an internal condition that involves affect. Within that, terms are counted emotional in a strong sense if affect is their focal attribute, as against terms that imply a combination of affect and behaviour (e.g., cheerful) or affect and cognition (e.g., bemused). In our terminology, the latter would count as emotional states.

Despite the differences, there is broad agreement on the scale of the emotion lexicon. The Semantic Atlas of Emotional Concepts (Averill, 1975) lists 558 words "with emotional connotations". The number that could be regarded as describing emotion proper is smaller, but still substantial. Fehr and Russell (1984) listed 196 emotion words in English; Plutchik (1980) listed 142; Whissell (1989) listed 107; Shaver et al. (1987) listed 213; Storm and Storm (1987) considered 135 to be clearly emotional. Ortony et al. (1988) classed about 250 words as referring to emotion in a

broad sense, and about half as many referring to emotion in a narrower sense. Languages other than English are comparable: Scherer (1984a) listed 235 emotion terms in German, and Zammuner (1998) listed 153 in Italian. A few studies list a subset of terms that are in common use, and relatively prototypical: Storm and Storm list 72, and Zammuner, 61.

Numbers like these reveal a difficult problem. Languages do not generally multiply terms unless they are needed. Hence, if speech research intends to cover the range of emotions that occur with reasonable frequency in everyday life, and to describe them using the everyday category system, about 60 is a low estimate of the number of categories that it should expect to deal with. Working directly with that number of categories clearly raises problems of tractability. At the least, establishing speech correlates for that number of categories would require research projects on a very large scale.

The problem is exacerbated by the fact that despite their number, everyday categories do not capture every shade of emotion that people can distinguish. A vocabulary of 60 words or so may be too large to be tractable, but it is too small to capture distinctions that human observers make, and feel are important. Pictorial art provides a neat way of making this point. Artists revel in expressions that convey an emotional state which is very easy to identify with, and yet very hard to verbalise. That may well be why languages are so receptive to words from other cultures that capture a hitherto unlabelled emotional state—consider examples such as *chagrin*, *ennui*, *angst*, *hubris* (see Russell, 1997).

A related phenomenon is seen in the psychological literature, where theorists argue that everyday terms subsume quite different types of states. Anger is often divided into two types, hot and cold (Banse and Scherer, 1996). Lazarus (1999a) argues that a four-fold distinction is more appropriate: his subtypes are inhibited, righteous, pouting, and hostility. Sternberg (1988) has distinguished four types of love.

The multiplicity of emotion categories is related to the complexity of the meaning that everyday emotion words carry. As noted above, attributing

an emotion to a particular person tends to imply judgements about his or her affect, appraisal of the situation, action tendency, communicative behaviour, and physiological state. It may involve much more. In some cases, the complexity is uncontroversial. A word like “vengeful”, for example, does not simply describe the present situation. It carries implications about past events (for which revenge is sought), long term goals, and moral status of the parties. Some emotions seem more straightforward, e.g., anger. However, social constructivists have made a case against accepting that kind of judgement too readily. Averill (1982) argues that far from being primitive, anger rests on a complex pattern of socially determined appraisals, and serves important social functions. On that view, to say that someone is angry is to say that he or she is following a complex, culturally defined script.

The complexity of emotion categories raises questions about the part that speech is likely to play in categorisation. It is reasonable to suspect that the category system may be oriented towards situations where different kinds of information are available in several modalities. If so, speech might be expected to provide a particular kind of information rather than to suggest a full classification. The evidence to date tends to suggest that cues from face and voice carry essentially similar types of information (Mehrabian and Ferris, 1974; de Gelder and Vroomen, 2000a,b; Massaro and Cohen, 2000), but the audio–visual combinations involved may have been too simple to make divergent functions obvious.

Some research has followed the path of embracing everyday emotion language by using the whole range of emotion terms, and if necessary, combining them to capture even subtler shades. For example, the database collected by the Reading group (Douglas-Cowie et al., 2003) describes an utterance as moving from “hate” to “vengeful anger”. That approach has the advantage that it does not throw away very much information. However, it is most likely to be useful if it can be linked to a system that is more tractable, so that states identified by a fine description can be assigned for purposes of analysis to a smaller number of larger classes, or located in an ordered

framework. Potentially relevant techniques are considered below.

5.2. *Emotion-related states*

We have already expressed reservations about separating emotion proper from emotion-related states. Ideas introduced in the previous section reinforce the point. According to quantitative approaches, there is no intrinsic division, only a rough line drawn for the sake of convenience. According to qualitative approaches, the features that emotions proper and emotion-related states share are as significant, and as likely to be related to speech, as the features that distinguish them.

Uncertainty about the distinction is underlined by the fact that certain states are categorised by some authors as prime examples of emotion, whereas others deny that they are strictly emotions at all. The outstanding examples involve love and humour. Love featured in early lists of primary emotions, notably Descartes' (Anscombe and Geach, 1970) and has been reinstated by many recent groups (e.g., Hatfield and Rapson, 1993; Oatley and Jenkins, 1996), but others argue that it is not an emotion at all: For instance, Ekman (1999) argues that love differs from emotion proper in its time course, and so (with grief and jealousy), it is better classified as an emotional plot. Conversely, emotion research did not traditionally consider humour. That has changed recently, and emotions related to humour (amusement, 'exhilaration') are a major research area (Ruch, 1993). Note that amusement does meet Ekman's criteria, and he considers it a basic emotion. On the other hand, Ortony et al. (see below) consider it more cognitive than emotion proper.

Unsurprisingly, there is no agreed taxonomy of states that are generally classified as emotion-related, but the literature offers some ways of structuring them. A well known type of distinction involves the temporal dimension. Some emotion terms refer to transient states, others to enduring traits. That issue is taken up in Section 6.4. A second type of distinction, introduced earlier, involves states that seem to imply either valenced appraisal or arousal, but not the full emotional syndrome. One is attitude. A standard definition

of attitude is that it entails "categorisation of a stimulus object along an evaluative dimension" (Zanna and Rempel, 1988, p. 319), which suggests that attitude is like emotion in the sense that valence is pivotal, but differs in that arousal is not. Conversely, terms like "stressed" or "drowsy" imply arousal, but not valence (at least not to the same extent).

The system proposed by Ortony et al. develops a related conception. They identify a class of terms that relate to physical and bodily states, corresponding roughly to arousal. Contrasting with that is a class of terms referring to mental conditions that have "either a significant cognitive component or a significant affective component ... some have a significant behavioural component" (p. 351). These are subdivided according to which of the elements is focal. Emotion terms (considered earlier) focus on the affective component. Terms like "alert", "hopeful", and "sincere" focus on cognition. Terms like "aggressive", "funny", and "sarcastic" focus on cognitive and behavioural components.

Each of these classes subsumes a large number of everyday and theoretical terms. Lazarus (1999a) proposes a threefold division of stress: his subtypes are harm/loss, threat, and challenge. Articles on attitude by Schubiger (1958) and O'Connor and Arnold (1973), used nearly 300 labels between them. These covered states such as abrupt, accusing, affable, affected, affectionate, aggressive, agreeable, airy, amused, angry, animated, annoyed, antagonistic, apologetic, appealing, appreciative, apprehensive, approving, argumentative, arrogant, authoritative, etc.

Even a brief review indicates that the categorical descriptions readily available in these areas pose a challenge to speech research. On one hand, it seems very likely that speech variables carry some information about these states (see Section 4). On the other hand, the task of finding unique speech correlates for every category is even less tractable than the task of finding unique speech correlates for every emotion proper. Correspondingly, there is an even more urgent need to find how everyday categories can be organised into a smaller number of larger classes, or located in an ordered framework.

5.3. Lists of key emotion categories

The oldest approach to simplifying the description of emotion retains categorical description, but tries to identify a short list of categories that underpin emotional life as a whole. Such a list would be invaluable for research on speech and emotion, but attempts to formulate one have had limited success. They are linked to theoretical debates, since different theories suggest different criteria for inclusion or omission.

An early version of the idea, introduced by Descartes (Anscombe and Geach, 1970), assumes that the key categories are ‘primary’—i.e., other emotional states are derived from them by mixing them rather like primary colours. That has been called the palette theory (Scherer, 1984b). The term primary is still widely used, in lay parlance and in the speech literature (e.g., Murray and Arnott, 1993); but in fact, ‘palette’ theories have very little support in modern emotion research (Ekman, 1999).

The phrase ‘basic emotions’ expresses a related conception with wider support (e.g., Stein and Oatley, 1992; Ekman and Davidson, 1994). The term refers to qualitatively distinct states that are held to be universal at least in essence, and associated with brain systems evolved to cope with significant types of situation. Fuller lists of the characteristics associated with basic emotions have been proposed, notably by Ekman (1992, 1999), who lists nine. Other theorists use different specific criteria, and so arrive at different lists of the states that they consider basic. Section 7 gives some of the main examples.

Lists of basic emotions may in a sense span the whole of emotional life. Ekman (1999, p. 57) proposes that ‘all emotions are basic’. However, lists of basic emotions need to be treated with caution for two reasons. First, each entry actually stands for a family of emotions, some with quite diverse members. Second, Ekman’s proposal amounts to a prescription that the term emotion should be reserved for states that satisfy the stringent criteria that he lists. If so, it simply means that other terms need to be found for the broader range of interesting and important states that the word emotion currently stretches to cover.

Although the basic emotion framework has widespread support, there are substantial challenges to it. As noted earlier, social constructivism regards all emotions as cultural products that owe their structure to learned social rules. From that perspective, there could be no fixed list of emotions. New social roles may always be added to a culture’s repertoire, or old roles modified (see, e.g., Wierzbicka, 1992). Ortony and Turner (1990) have questioned the concept on different grounds. They argue that emotional life does rest on a small number of primitives, but that the primitives are not themselves emotions. All emotions correspond to combinations of the primitives. The idea is taken up in the next section.

It is beyond the scope of this paper to attempt a resolution of these issues. A fuller discussion is given in (Cornelius, 1996). In our judgement, it seems best to accept that some emotions are more primitive and universal than others, allowing that it is not straightforward to specify the sense in which they are primitive or universal. The term ‘basic’ seems the most acceptable way to refer to those emotions. There is no generally accepted contrasting term. ‘Secondary’ is unsatisfying because of its association with primary. The term ‘higher level’, used by Buck (1999), avoids that problem.

A contrasting approach to formulating brief, comprehensive lists has become increasingly prominent. It considers the idea that emotion terms fall into a few superordinate categories. Several of the studies described in the previous subsection offer structures of that kind. Shaver et al. (1987) proposed that five prototypes—anger, love, joy, fear, and sadness—underlie all emotion categories. Subdivisions of joy, for instance, would be pride, contentment, and zest. Storm and Storm (1987) proposed an essentially similar division of terms with a definite implication of positive or negative experience, but added two categories for relatively neutral terms emphasizing cognition and arousal. One included passive terms such as sleepy, apathetic, and contemplative; the other included active terms such as aroused, surprised, and understanding. Lazarus (1999a) proposes five super classes that partially overlap with those of Shaver et al.: nasty; empathic; provoked by favourable

conditions; provoked by unfavourable conditions, and existential. Buck (1999) distinguishes biological affects (happiness, sadness, anxiety, fear, anger, disgust) from social affects (pride, guilt, envy, pity, arrogance, shame, jealousy, scorn); cognitive affects (curious, interested, surprised, bored, burned out); and moral affects (moral rapture, moral indignation).

Relationships between speech and these superordinate categories seem to be unexplored. *Prima facie*, it seems unlikely that there is one type of speech per category. Pride, contentment, and zest may be conceptually close, but it seems unlikely that they have similar vocal expressions. However, there could be more abstract correspondences (what is expressed prosodically, what through voice quality, etc.).

This is a brief review of a notoriously complex area. It has tried to indicate why research on speech and emotion should not depend too much on any given list of key categories. However, that does not mean compact lists do not have uses. The last section of the paper sets out some significant candidates.

5.4. *Everyday emotion terms: summary*

Everyday categories are much more interesting than people tend to assume—and much harder to work with. It may be useful to think of applying them appropriately as the final examination for an intelligent person-sensitive system, not as the first step. On the other hand, the attempt to apply them automatically, on the basis of evidence such as speech, provides a perspective that may be useful to the field as a whole.

6. The structures underlying categorical descriptions

It has been proposed that category labels specify where a person's state falls in some kind of underlying structure. If so, the solution to the problems associated with categorical descriptions may be to look beyond category terms to the structure that underpins their meaning.

Several approaches to that task have been explored to a greater or lesser extent. They highlight

descriptive tools and dimensions that are available to research on speech and emotion as a result. The case to be made is not that their value is proven, but that it ought to be explored systematically. This section introduces the various approaches and considers their status. Details of techniques and tools based on them are considered in the next section.

6.1. *Biological representations*

Biologically oriented descriptions have a peculiar status, partly because there is an influential tradition that regards them as the essence of emotion. Psychologists have devoted an enormous amount of time and energy to testing the proposition, derived from James (1884), that emotion can be equated with awareness of a visceral response; and therefore each emotion should be accompanied by a unique pattern of physiological response. The idea offers the attractive prospect that physiological measurement might offer a way of accessing a person's emotional state directly, without relying on his or her honesty or self-knowledge (Blascovitch, 2000). That background suggests that the solution to problems with verbal descriptions is to replace them with physiological parameters. This is not a realistic position, however. The real discriminative power of physiological measures is limited both in principle and in practice.

The limitation in principle involves a style of argument that has already been introduced. At least some emotion terms refer implicitly beyond the immediate situation. The validity of a claim to experience remorse, for instance, depends partly on the actions that follow it, not purely in the physiological state that accompanies it. The point is one of a group related to Wittgenstein's private language argument (Wittgenstein, 1953; Kenny, 1973), and is not easy to dismiss.

The limitation in practice rests on the kind of discrimination that physiology can actually achieve. This is an empirical problem, and a highly technical one. Visceral signs of emotion do show some statistical discrimination, as Levenson, Ekman and their colleagues have demonstrated (Levenson, 1992; Cornelius, 1996). The work of

Picard and her colleagues (Picard, 2000) conveys the state of progress on a task with a particular appeal to the speech community, i.e. classifying emotional states on the basis of physiological measurements. Early systems distinguished anger from peaceful emotions with about 90% accuracy, and high and low arousal states with about 80% accuracy; but not positive and negative emotions (Healey and Picard, 1997). Subsequent improvements have raised discrimination rates to around 80% for a set of eight emotions (Vyas and Picard, 1999). The performance is impressive, but very far from the level of discrimination involved in everyday classification. Note, too, that the performance is on a single subject inducing strong emotional states. Transfer to more challenging tasks remains to be seen.

Brain imaging techniques potentially add another dimension to physiological observation. Research on brain mechanisms of emotion is a rapidly growing field (for recent reviews, see Rolls, 1999; Davidson et al., 2000a). It has identified a number of different brain systems that are strongly associated with emotion. The amygdala have rich inputs from sensory systems, and are involved in learning the reward values of stimuli. It is natural to interpret them as a key site in evaluating situations as positive or negative. Orbitofrontal cortex is involved in preparing behavioural responses and autonomic responses. It is natural to link its function to action tendencies. The basal forebrain has widespread effects on cortical activation, and direct links to autonomic nuclei: that suggests a role in arousal. The prospect of observing activation in these systems is intriguing. It is becoming possible to think of activity in the amygdala as an indicator of emotional involvement. But as with older physiological techniques, resolution between fine shades of emotion is not an immediate prospect.

Paradoxically, research in the Jamesian tradition has tended to treat physiological variables as a means to categorical descriptions. Recent research has highlighted an alternative perspective, which is primarily concerned with correlation (and decorrelation) among various emotion-related variables. In particular, Reisenzein (2000) has highlighted the “decoupling” of physiological with self-report,

expressive, and behavioural measures of emotion; that is, the measures from these different systems tend not to correlate very highly with one another. From that perspective, the natural strategy is to explore patterns of correlation and decorrelation among vocal and physiological indicators, without assuming that either will relate simply to categorical descriptions.

That kind of approach is almost bound to look beyond emotion in a narrow sense—for example, to look at vocal correlates of physiological measures that indicate arousal, whether or not the arousal involves emotion proper; or at states where physiological changes diverge from signs that are under voluntary control (either because voluntary behaviour is meant to mask underlying emotions, or because voluntary behaviour tries to give the appearance of an emotion that is not felt). The approach can also benefit from the fact that many of the physiological changes associated with emotion have effects that can have an impact quite directly on voice, such as drying of the mouth, tension, and tremor. Stevens and his colleagues drew attention to speech variables that might reflect those changes (Williams and Stevens, 1972), and the approach has been elegantly developed by the Geneva group (Banse and Scherer, 1996). It seems to be relevant to strong emotions associated with preparation for “fight or flight”. It is less clear how far beyond this it applies.

It may be that the only reliable relationships between voice and emotion are those that reflect more basic relationships between emotion and emotion-related changes in physiology. However, it is also conceivable that either evolution or social factors have ensured that voice communicates more information about an organism’s general orientation to the world than basic physiological changes require it to. That kind of issue becomes easier to address if the traditional preoccupation with emotion categories is set aside.

6.2. *Abstract dimensions*

A natural goal for systematic description is to find ways of representing emotional states as coordinates in a space with a small number of dimensions. Research on that problem has a long

history. Wundt (1874/1903) argued that feelings could be located in a space with three dimensions, pleasant–unpleasant, excited–quiet, and strain–relaxation. The idea was taken up by a series of investigators in the mid twentieth century (see, e.g., Izard, 1993). An influential formulation by Schlosberg (1954) underlined the importance of two concepts that have already been introduced and used, valence and arousal. Schlosberg demonstrated that a space in which they form the axes captures a considerable amount of information about emotion. We have called it activation-evaluation space (Cowie et al., 2001).

Activation and evaluation dimensions can be thought of as minimal representations of themes whose centrality to emotion is not in doubt. The link between valence and emotional states is well put in Tomkins' (1982, p. 355–356) description of affect as what gives things value: "without its amplification, nothing else matters, and with its amplification, anything else can matter". Similarly, research from Darwin on has recognised that emotional states are characterised by dispositions to act in certain ways. A well known extension is Frijda's (1986) proposal that emotions are action tendencies. The concept of activation can be thought of as a coarse way of reflecting this notion: activation level summarises the strength of the person's disposition to take some action rather than none.

One of the attractions of activation-evaluation space is that it provides a way of describing emotional states which is more tractable than using words, but which can be translated into and out of verbal descriptions. Translation is possible because emotion-related words can be understood, at least to a first approximation, as referring to positions in activation-emotion space. A variety of techniques converge on this conclusion, including factor analysis, direct scaling, and others (Russell, 1997).

A curious feature of the space is that words describing full-blown emotions are not evenly distributed in it. Instead, they tend to form a roughly circular pattern. Russell, Plutchik and others have argued that the circularity captures fundamental intuitions about the structure of emotion (Plutchik and Conte, 1997)—for instance,

there are natural opposites, and progressing through the gamut of emotions eventually returns us to the starting point.

Activation-emotion space is an elegant tool, and it has proved attractive to computationally oriented research (Cowie and Douglas-Cowie, 2001; Schröder et al., 2001; Tsapatsoulis et al., 2002). However, it has to be emphasised that the representation depends on collapsing the structured, multidimensional space of possible emotional states into a homogeneous space of two dimensions. Information is inevitably lost; and worse still, different ways of making the collapse lead to substantially different results.

Particularly awkward is the fact that fear and anger lie close together in activation-evaluation space, too close to be effectively distinguished. That problem can be met by adding a third dimension, which is sometimes identified as perceived control (positive in anger, negative in fear) and sometimes as inclination to engage (also positive in anger, negative in fear). One of the best known dimensional treatments of emotion, due to Osgood et al. (1975), incorporates control as a third dimension. The difficulty is that no one extension allows very many additional states to be discriminated; and once one begins to add dimensions for the sake of a few discriminations, it is difficult to know where to stop.

A different reformulation was introduced by Watson and Tellegen (1985). They proposed that the natural axes of a two-dimensional emotion space were not activation and emotion, but positive and negative affect. The major advantage of the reformulation is that it may relate more directly to physiological substrates (see Davidson, 1999 for a review).

A number of investigators have used descriptions based on activation-evaluation space to study speech. Of particular interest is the suggestion that speech variables could be correlated with dimensions rather than with discrete categories. There has been an accumulation of evidence suggesting that activation in particular relates directly to vocal variables (Bachorowski, 1999; Pereira, 2000; Schröder et al., 2001). Looking across studies (Cowie et al., 2001), high activation appears to be associated with increases in the mean and range of

F0, and tense voice quality. These have been reported in connection with happiness, fear, anger, and to a lesser extent surprise, excitement and puzzlement. Low activation appears to be associated with decreases in the mean and range of F0, as suggested by studies of sadness, grief, and, to a lesser extent, boredom. Vocal correlates of evaluation have been reported, but the findings are less consistent, and the effects may be subtler. (Bachowski, 1999; Schröder et al., 2001).

6.3. *Structural models*

The techniques described in the previous section consider one type of framework in which emotional states might be located. It is defined by a few continuous dimensions. An alternative approach is to consider a framework that takes the form of a tree, in which states are characterised by the presence or absence of various logical elements. That approach has been associated with theorists who argue that distinct types of emotion correspond to distinct ways of appraising the situation that evokes the emotion. Work in that tradition attempts to identify logical primitives from which descriptions of appraisals can be constructed, and combinations of elements that capture the distinctive appraisals underlying particular emotional states.

Scherer (1999) provides a wide-ranging review of that approach. To illustrate it, consider two proposals, one from Roseman (1991) and the other from Ortony et al. (1988). Both include two distinctions that can be regarded as basic—whether the key elements of the situation are positively or negatively evaluated in themselves, and whether or not they help the agent to achieve his or her goals. Roseman identified additional distinctions based on the way agents appraise key elements of the perceived situation—whether they are of the agent's own making, whether they are known or unknown, and whether the agent regards him- or herself as powerful or powerless. Ortony et al. explored a different approach, using the concept of focus, which was introduced in Section 5. The focus may be on different agents—the person experiencing the emotion, or someone else. It may also be on different levels—"objects" (including people or

things), actions (of people or animals), or sequences of causally related actions or events. Broadly speaking, the range of emotions that can be associated with an object as such is much narrower than the range of emotions that can be associated with a sequence of events involving various agents.

Accounts of this kind suggest that speech research might look for variables relevant to the distinctions underlying a system of appraisals. Many of those distinctions are of a kind that one might imagine having vocal correlates—consider, e.g., themes like power or weakness, knowledge or uncertainty, guilt or satisfaction, and perhaps focus on immediate surroundings as against scenarios that are remembered or anticipated. It should be noted that similar distinctions apply very naturally to attitude, reinforcing the point made earlier that affinities between attitude and emotion should not be taken lightly.

6.4. *Temporal issues*

Timing is an issue whose relevance to emotion is clear, and yet it is often not fully integrated into descriptions of emotion categories. It is important that it should be, on several levels. The issue is signalled by the way everyday emotion language distinguishes among states that have similar instantaneous qualities but different time courses. The word sadness can describe a relatively short-lived state. Grieving, on the other hand, is a process, and if it does not extend over a period then it is debatable whether the word properly applies. Depression is also intrinsically likely to be an extended phenomenon, and likely to continue until something happens to end it. Gloominess as a personality trait is expected to last a lifetime.

A straightforward implication for speech research is that there may be issues worth considering at relatively long timescales. For example, one could envision a system that used speech to accumulate evidence on shifts in a user's mood over a period of hours or days (cf. Nezlek et al., 1983). A natural extension is that many issues may be best addressed in terms of a dual timescale, involving a long term average to act as a reference, and short term departures from it to signal emotionally

marked events. It is well-known that people find at least some individuals difficult to ‘read’ emotionally unless they have enough experience to know the relevant baselines (cf. Mischel and Peake, 1982).

With regard to full-blown emotional episodes, timing may well be diagnostic in several senses. An effective synthesiser needs to release and sustain signals of emotion on the right timescale. Conversely, an effective recognition system needs to be prepared to capture departures from baseline that last for a relatively brief time. Both depend on collecting data that are capable of reflecting the real time course of emotional signals in speech. Our impression is that timing is an area where acted emotionality may be very far from the real thing. A real possibility is that there may be multiple scales at work even in the short term, with some signs building up over a period of seconds or minutes and others erupting briefly but tellingly. In all of these respects, research concerned with speech and emotion needs to be clear that its decisions relate to states with characteristics in the domain of time as well as in the domains of feeling, appraisal, action tendency or social consequences.

6.5. *Emotion and social interaction*

If speech research is concerned with understanding the way emotion appears in everyday life, then it has to deal systematically with the interactions that determine how underlying emotional tendencies are expressed. The issue may be particularly acute in the context of speech if, as Section 4 has suggested, there is a conflict between speech and unrestrained emotion.

The most obvious type of interaction is restraint. Ekman and his co-workers (e.g., Ekman and Friesen, 1969) introduced the term ‘display rules’ to describe the constraints that govern socially acceptable expressions of emotion. Rather little seems to be known about the display rules for speech (but see Goffman, 1981, for some characteristic insights into such matters). Our exploratory work convinces us that they are a vital topic. Strong underlying emotion is often signalled by unnatural behaviour arising from a determination not to release socially unacceptable signs. It is even

more revealing when socially unacceptable signs surface briefly in spite of one’s determination to control them.

A second type of interaction is ambivalence. Well-known phrases highlight the effect—‘parting is such sweet sorrow’, ‘I don’t know whether to laugh or cry’, ‘love–hate relationship’, etc. The arts have a fascination with these ambivalent states. The kind of data summarised in Table 1 suggests that they may actually be what speech usually expresses. Certainly the issue needs to be addressed systematically. Temporal issues may be central to doing that—the lead role often seems to shift back and forth between mixed emotions.

A final type of interaction is simulation. People do simulate emotion. It is sometimes obvious, and sometimes not. Some styles, such as expressive reading, seem to use emotion-like features in tandem with signals that they are not to be taken literally. The issue is one that speech research would be unwise to leave off its agenda. People respond negatively to displays of emotion that are perceived as simulated, and this is a real issue for agents that are intended to convey emotion.

6.6. *Conclusion*

As should be evident from the discussion above, emotion has many aspects that may shape vocal patterns. It makes sense to explore relations between speech variables and these potentially relevant aspects rather than concentrating exclusively on their relationship to a category system that may pose intractable problems. Work on these different aspects is not mutually exclusive, and they should not be regarded as competing approaches. We turn now to a discussion of a set of tools that may prove useful in understanding speech emotion relationships.

7. *From concepts to resources*

This section moves from conceptual issues to specific resources that are available to describe the emotional states of individual speakers. It is clearly not possible to give full descriptions of complex apparatus and procedures: the aim is simply to

indicate that key resources exist and where more information can be found.

7.1. Lists of key emotions

We have argued against over-reliance on short lists of emotion categories. However, it would be unrealistic not to consider lists as significant descriptive tools.

The best known list of key emotions is the so-called “Big Six”—fear, anger, happiness, sadness, surprise, and disgust (Cornelius, 1996). Minimal ways of extending the group include distinguishing hot and cold anger, and adding contempt (Ekman and Heider, 1988) and love (Hatfield and Rapson, 1993) or desire (Johnson-Laird and Oatley, 1992). Greasley et al. (2000, p. 364) have directly examined the use of a short list to code the emotions expressed in speech, and concluded that “giving subjects choice of just one label from a small range of basic emotions may not allow subjects an adequate level of discrimination”.

During the 1990s several theorists advanced lists of basic emotions that were considerably longer, mainly because they included more positive and/or interpersonal emotions. Table 2 gives significant examples, from Lazarus (1999a) and Ekman (1999). Beside them are lists that reflect different perspectives. Buck (1999) offers a typology of what he calls affective states rather than emotions per se; the table lists the states that he describes. The list from Lewis and Haviland (1993) can be taken to reflect a pragmatic consensus: it consists of the emotions that are allocated chapters in their influential handbook. The list from Banse and Scherer (1996) is the most systematic in the literature concerned explicitly with speech.

The final column represents our own attempt to develop a “basic emotion vocabulary”, that is, a set of emotion-related categories that is small enough to be tractable, but that covers the range of emotion-related states that commonly occur (Cowie et al., 1999a,b). Subjects were given a list of terms that recur in the literature, and asked to select a subset that they felt would be sufficient to describe everyday emotional occurrences. The terms used in Table 2 were the 16 most commonly chosen. Note that a substantial proportion of

terms describe emotion-related states (e.g., ‘confident’, ‘relaxed’, ‘bored’): this reflects subjects’ sense that these were significant parts of everyday emotional life.

Overall, the table shows 38 emotional or emotion-related states. However their theoretical status may eventually be resolved, most of them have a reasonable claim to be regarded as significant landmarks in the domain of emotional life. A list of that kind provides a useful orientation, in the sense that it provides a basis for asking what proportion of the key discriminations in the domain can be based on speech alone.

7.2. Authentication of cause-oriented emotion descriptions

For cause-oriented research, a core part of the descriptive task is establishing the true state of the speaker. We use the term authentication to describe that task. It is tempting simply to call it description, but that is misleading: authentication is only one of several possible descriptive aims. It should be noted that not all emotions are equally easy to authenticate. A fast, reliable method of authenticating love would be interesting indeed.

7.2.1. Self report

Emotions researchers have developed a variety of self-report measures to index participants’ emotional experience. Among the most widely used of such measures are the Discrete Emotions Scale (Izard, 1972), the Multiple Affective States Adjective Checklist (Zuckerman and Lubin, 1985), and the Positive and Negative Affect Scales (PANAS) (Watson et al., 1988; Roemer, 2001). Key tools aimed at emotion-related states are the Spielberger State-Trait Anxiety Inventory (Spielberger, 1988) and the Profile of Mood States (POMS) (see Jacobson et al., 1978).

Self-report data are notoriously prone to be unreliable. Some of the problems are associated with the way measures of these kinds are administered and scored (Green et al., 1993), and these need to be considered when self report is used. Nevertheless, it is a legitimate question whether subjective feeling correlates with other indicators of emotion, including vocal indicators; and it is

Table 2
Recent lists of key emotions

Lazarus (1999a)	Ekman (1999)	Buck (1999)	Lewis and Haviland (1993)	Banise and Scherer (1996)	Cowie et al. (1999b)
Anger	Anger	Anger	Anger/hostility	Rage/hot anger Irritation/cold anger	Angry
Fright	Fear	Fear	Fear	Fear/terror	Afraid
Sadness	Sadness/distress	Sadness	Sadness	Sadness/dejection Grief/desperation	Sad
Anxiety	Sensory pleasure	Anxiety	Anxiety	Worry/anxiety	Worried
Happiness		Happiness	Happiness	Happiness	Happy
			Humour	Elation (joy)	Amused Pleased Content Interested
		Interested Curious Surprised			
	Excitement	Bored		Boredom/indifference	Excited Bored Relaxed
Disgust	Disgust	Burnt out			
	Contempt	Disgust	Disgust	Disgust	
Pride	Pride	Scorn		Contempt/scorn	
		Pride	Pride		
Jealousy		Arrogance			
Envy		Jealousy			
Shame	Shame	Envy			
Guilt	Guilt	Shame	Shame	Shame/guilt	
	Embarrassment	Guilt	Guilt		
			Embarrassment		
Relief	Relief				Disappointed
Hope					
					Confident
Gratitude					
Love			Love		Loving Affectionate
Compassion		Pity Moral rapture Moral indignation			
Aesthetic					

difficult to answer the question without using self report.

7.2.2. Antecedent conditions

Cognitive theorists have argued that there is an intimate correspondence between emotions and the antecedent situations that evoke them (Scherer and Walbott, 1986; Scherer, 1988). Hence, arguably, a key element of authentication is arranging

antecedent circumstances whose effect is well documented (Scherer, 1988). Various standard ways of doing this exist (see Gerrards-Hesse et al., 1994).

In one of the first experimental investigations of the appraisal process, Speisman et al. (1964) had volunteers view a gruesome film under different instructional set conditions. Films and videos are now widely used for the induction of a variety of

emotions (see Gross and Levenson, 1995; Hagemann et al., 1999; Luminet et al., 2000). Still photographs, such as the “Philadelphia Morgue” set, have been used extensively to elicit emotions such as fear and disgust (see Averill et al., 1978). A standard set of still photographs to be used in eliciting a range of emotions was recently developed by Bradley, Lang and their colleagues (Center for the Study of Emotion and Attention, 1999; Lang et al., 1999).

Music, in particular, sad music, has been found to be a reliable means of inducing sadness in the laboratory (Clark, 1983; Kenealy, 1988) and has been extensively used to induce depressive-like affect in non-clinical samples of undergraduate students (Morrow and Nolen-Hoeksema, 1990).

Researchers have also elicited an impressively wide range of both simple (Ekman et al., 1983) and complex emotions (Shackelford et al., 2000) simply by having participants imagine situations involving the desired target emotions. A variant of this technique, presenting participants with written scenarios with systematically varied content, has been used by cognitively oriented emotion researchers to test hypotheses about the specific appraisal components underlying particular emotions (see Scherer, 1988; Cornelius, 1996). Note, though, that physiological evidence suggests that states induced by imaging are substantially different from states induced by external events (Stemmler et al., 2001).

A pioneering set of experiments by MacDowell and Mandler (1989) used participants’ reactions to positive and negative unexpected events programmed into a video game to test propositions from Mandler’s (1975) theory of the antecedents to emotion. More recently, Reisenzein (2000) used a computerized science quiz to examine the relationships among the various expressive and experiential components of surprise. A related, older methodology involves exposing participants to various “threats” such as a painful injection (Schachter, 1959), a difficult intelligence test (Braunstein-Bercovitz et al., 2001), electric shock (Cornelius and Averill, 1980), or speaking in public (Davidson et al., 2000b).

Most of these techniques were not developed with speech elicitation specifically in mind, and it is

not obvious how easily they can be adapted for that purpose. However, a number of speech-oriented procedures have been reported. One of the simplest and yet most effective means of inducing both positive and negative affect in laboratory settings is the Velten Mood Induction Procedure (Velten, 1968), in which participants are asked to read a number of brief self-referential statements about their moods and emotions, e.g., “I feel happy”, “I feel sad”. Although the procedure seems almost too easy, it has nevertheless been found to be remarkably effective in inducing both happiness and depressive-like affect (see Frost and Green, 1982; Kenealy, 1986; Larsen and Sinnett, 1991; Finegan and Seligman, 1995), as well as anger (Engelbreton et al., 1999).

A few induction techniques have been specifically designed to elicit speech; for example, Bachorowski (1999) used a difficult spelling task to elicit negative emotion; Batliner et al. (2003) used a Wizard-of-Oz-scenario where a malfunctioning system is simulated in order to evoke anger; and Fernandez and Picard (2003) used mental arithmetic tasks to elicit speech under stressed conditions. It is noticeable that the states involved tend to be moderate or emotion-related rather than full-blown emotions.

All this is not to say that there are no problems with the various procedures for inducing moods and emotions experimentally and that they have no critics. Westermann et al. (1996) conducted a meta-analysis of 250 experimental mood/emotion induction effects and found that such effects tend to be larger for negative than positive moods and emotions and when participants are given explicit instructions to enter a particular mood or emotion. The former finding echoes an earlier review of the effectiveness of a variety of mood/emotion induction procedures by Gerrards-Hesse et al. (1994) that indicated that negative mood (e.g., depression) inductions were more effective than positive mood (e.g., elation) inductions. The latter finding suggests that at least some mood/emotion induction procedures may be contaminated by demand, and, indeed, Westermann et al. found that mood induction effects tend to be smaller when experimental demand is controlled (see also Polivy and Doyle, 1980). Finegan and Seligman

(1995), however, explicitly tested the effectiveness of one particular method of inducing mood, the Velten Mood Induction procedure (Velten, 1968), to produce effects over and above those attributable to demand. They concluded that the Velten procedure, even though prone to demand, produces “genuine” mood effects. Caution regarding the use of mood/emotion inductions was also urged by Polivy (1981), who, using three different laboratory mood/emotion induction procedures (angering participants by an experimenter, the Velten Mood Induction procedure, and the threat of electric shock), found evidence that such inductions tend to produce multiple affective states rather than single, discrete moods or emotions.

The available evidence on the effectiveness of a wide variety of mood and emotion induction procedures suggests that there is considerable variation in the effectiveness of such procedures and how “pure” the moods and emotions they elicit are. Nevertheless, there does seem to be a consensus that such procedures work and that they may be (in the language we are using here) important and occasionally quite powerful tools to use in the authentication of cause-oriented emotion descriptions. We offer a simple prescription. Using established induction procedures can make it likely that samples of a target emotional state will be authentic; but additional checks are needed to make it certain.

7.2.3. *Physiology*

As noted in Section 6.1, research offers limited support for the hypothesis that emotions are associated with unique patterns of physiological activity. The distinctions that are generally accepted can be drawn using a relatively small set of measures. Reviews (see Levenson, 1992; Cacioppo et al., 1993; Cornelius, 1996) consistently indicate that there is specificity at the level of a small set of indicators of autonomic arousal, e.g., heart rate, blood pressure, finger temperature, and some measures of facial electromyographic activity, for a small number of emotions, most notably, anger, fear, sadness and disgust. In addition, there is evidence that anxiety may be marked by very simple, emotion-like responses such as the startle eye blink (Yartz and Hawk, 2001; Koukounas and McCabe,

2001). “Stress” is associated with cardiovascular changes that may be confounded with effects of emotion per se (Feldman et al., 1999), but also more specifically with elevated serum cortisol and lowered salivary secretory immunoglobulin A levels (see Hucklebridge et al., 2000). A much wider range of measures is used by research that distinguishes within traditional categories: for instance, Stemmler et al. (2001) consider 29.

In the central nervous system, there is a growing body of evidence that patterns of EEG activity associated with different emotions, especially those that can be characterized as approach- versus withdrawal-related, are quite different (see Davidson et al., 1990; Davidson, 1999). More sophisticated brain imaging techniques such as PET scanning are now widely used, but it is not obvious at this stage how they might apply to speech research.

7.2.4. *Facial expression*

A natural alternative to physiologically based authentication is to make still photographs or video recordings, and to use visual signs to identify the emotion being experienced or to elicit particular moods/emotions (for the latter, see, for example, Schneider et al., 1994). Facial expressions and other visible bodily indicators of emotion may be used from either a cause-oriented or an effect-oriented viewpoint. If the research is cause-oriented, the Facial Action Coding System (FACS), developed by Ekman and Friesen (1978), can be used to supplement intuitive judgement.

In FACS, each muscle group on the face is given a code number and so every emotional expression may be precisely described using a small set of numbers. By using the FACS, researchers can reliably encode (that is, have an actor produce) and decode (that is, have a naive participant recognise) a fairly impressive but still small number of “basic” emotions (the ‘Big Six’, plus, perhaps, contempt).

The FACS is widely used, but it is worth recognising that the underlying rationale has been challenged, notably by Russell (1994), Carroll and Russell (1996) and by Fridlund (1994). Russell’s challenge stresses issues of ecological validity. Ekman and his colleagues have generally used

posed expressions and forced choice responses in their research. The high recognition rates that they report do not carry over to facial expressions of emotion that have been spontaneously produced, or when subjects have a free choice of descriptors (For commentaries see (Ekman, 1994; Cornelius, 1996)). Hence there are questions about the use of faces to authenticate emotional states in just the situations where speech research might most obviously want supporting evidence.

In a more fundamental challenge, Fridlund (1994) disputed Ekman's implicit distinction between "natural" or pure expressions of emotion uncontaminated by display rules, and expressions produced or constrained by such rules. For Fridlund, it is misleading to think of human and other primate expressive displays as *expressions* of inner states. Instead, *all* expressive displays, even those that take place in private, involve social motives. Complementing Fridlund's contentions about the social nature of facial displays of emotion, Hess et al. (1995) found that facial expressions of emotion as indexed by four muscle groups was influenced by the type of eliciting stimulus involved (two films that differed in their degree of funniness), sociality (whether participants viewed the film alone or with another person), and the relationship between participants (friends versus strangers).

These points bear on the way facial expression might be used in speech-oriented research, not on the idea that it is potentially relevant, and hence the value of audio–visual data is not in dispute. Despite the problems surrounding authentication, it is clearly possible to ask both about correlation and non-correlation between facial and vocal signs, and about the way voice and face combine to induce particular effects in observers. However, it is very much in the spirit of this article to consider whether voice might convey the kind of information highlighted by Hess, Banse and Kappas rather than classical emotion categories.

7.2.5. *Long term outcome*

A few studies have considered whether emotion-related signs recorded at a particular time can predict practically important outcomes in the longer term. The most striking example was re-

ported by France et al. (2000). They studied recordings of psychiatric interviews, and found attributes of voice that distinguished patients who went on to commit or attempt suicide (specifically male patients).

Similar possibilities are suggested by the work of Gottman and his colleagues on predictors of divorce (Gottman and Levenson, 1992, 2000). They made video recordings of interactions between couples, and identified the attributes that characterised the couples who proved most likely to divorce (Gottman and Levenson, 1992). Some of the key indicators seem likely to have vocal elements, notably incidents of "unrequited" interest and related emotions in which each partner's expression of excitement over what he or she was discussing was not reciprocated by his or her partner (Gottman and Levenson, 2000).

These studies illustrate a kind of authentication that is logically important for at least some kinds of claim about emotionality; that is, claims that imply, if they are correct, that certain kinds of action are likely to follow. Smaller scale versions of the approach may well be appropriate for states such as boredom or irritation, whose significance is partly that they are likely to lead to unwanted behaviour.

7.3. *Effect-oriented emotion assignments*

Effect-oriented description is a pressing issue for the speech community because it needs to understand the likely emotional effect of synthesised speech. But by comparison with authentication, it has received very little attention from other disciplines, and so it is much more difficult to describe or illustrate good practice.

Central to the area is the need to handle ambiguity. From the naturalistic databases available (see Section 4), it appears that a sample of natural speech tends to have several possible emotion categories associated more or less strongly with it. Cauldwell (2000) has shown that the predominant response to an isolated extract may not be the predominant response in context. One might expect similar shifts in modal response to be associated with information from different modalities (see Section 5.1) and prior knowledge on the

listener's part (see Section 6.4). Progress towards capturing that kind of fluid, context-sensitive understanding is in its early stages.

7.3.1. *Categorisation by observers*

Techniques for obtaining categorical descriptions differ in the way they handle the fact that speech samples tend to evoke a spread of response.

Spread can be reduced by presenting a forced choice task with a small number of options. However, Greasley et al. (2000) have shown that apparent agreement produced by a five-choice discrimination dissolves when free choice is allowed. Forced choice tasks do have large advantages, but it is difficult to justify using lists much shorter than those shown in Table 2.

Banse and Scherer (1996) suggest an upper limit for the kind of agreement that can be expected with a list of that scale. Their observers classified about 50% of extracts in the expected category. The proportion was high because the extracts had been preselected as convincing portrayals. That is a rational strategy for certain purposes (e.g., providing exemplars on which to model synthesis).

Truly free choice produces much lower agreement. Using samples that appeared relatively unambiguous on a forced choice task, Greasley, Sherrard and Waterman found a spread of free choices, with few categories used by more than 15% of observers.

Multi-level responding may provide the best compromise between precision and tractability. The strategy used for the Belfast database, described in Section 4, is an example. Raters first selected from a short list (given in Table 2), but could then supplement that description from a longer one (containing the terms in the top row of Table 1). A rating of strength is associated with each choice, to reduce the sense that the category label needs to convey that information.

That kind of approach partially captures the range of responses that everyday speech is likely to evoke. However, it suffers from two related problems. First, it does not reflect the intuition that distinct category labels may or may not point to the same region of an underlying structure. Second, it is not well suited to representing the fluctuation of emotion from moment to moment.

7.3.2. *Agreed structural assignment*

One approach to defining underlying structure is offered by cognitive theorists, particularly by Ortony et al. (1988). Greasley et al. (1995) have translated that approach into a coding system. They proposed two levels of descriptor, one characterising the antecedent events associated with the emotion, the other describing valenced responses to it (e.g., 'reproach emotion', indicating disapproval of another person's actions). The system was applied to the Leeds-Reading database (Douglas-Cowie et al., 2003), but there is currently no detailed description of its characteristics.

7.3.3. *Static dimensional*

A second approach to defining underlying structure is to characterise emotions in terms of abstract dimensions (see Section 6.2). That approach provides a neat way of representing uncertainty, as a region of a space within which responses normally fall.

The problem with standard versions of the approach, as we see it, is that they are static and do not allow for capturing the meaning of emotion discourse in real time. Below, we describe a variant of the dimensional approach to the study of emotion meaning, FEELTRACE, that gives researchers the ability to track changes in the meaning of a person's emotional experience as they occur.

7.3.4. *Kinetic dimensional*

A system called FEELTRACE provides time-sensitive dimensional representations. It uses activation/evaluation space to let observers record their impression of a person's emotional state as it fluctuates in real time (Cowie et al., 2001). Its effectiveness depends on several features designed to convey the idea of emotion as a point in a two-dimensional space. Fig. 2 illustrates how FEELTRACE may be used to describe a person's changing emotional state.

The space is represented by a circle on a computer screen. Within it, the main axes are drawn and labelled: one (activation) running from very active to very passive; the other (evaluation) running from very positive to very negative. The circular form reflects the view of Russell, Plutchik

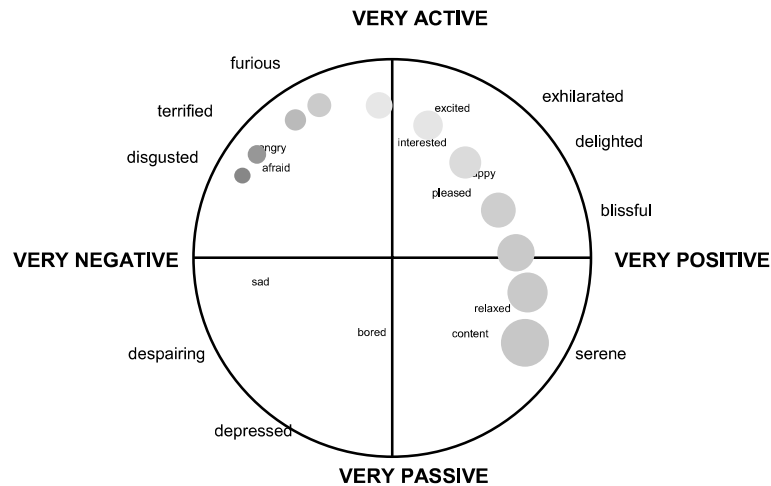


Fig. 2. Example of a FEELTRACE display during a tracking session. Cursor colour changes from red/orange at the left hand end of the arc, to yellow beside the active/passive axis, to bright green on the negative/positive axis, to blue-green at the right hand end of the arc.

and others that emotional space is naturally circular. We have considered the obvious alternative, a rectangular space, but responses are generally more consistent with the circular form. Subjects move a cursor within the space, so that at any given time its position signals the levels of activation and evaluation perceived by him or her.

The cursor is controlled by a mouse. Additional features provide users with fast, effective feedback. First, the cursor is colour coded using a scheme derived from Plutchik, which subjects find reasonably intuitive. The cursor is coloured pure red when its position signifies the most negative evaluation possible, and pure green when its position signifies the most positive evaluation. Similarly, it is pure yellow when its position signifies the most active state possible, and pure blue when its position signifies the least active state possible. Colour at intermediate positions was set by a straightforward additive rule. It is white when the cursor is at the origin of the space.

The colour coding is supplemented by verbal landmarks. Within the circle, emotion words from the core vocabulary in Table 2 are placed at coordinates obtained from a study where subjects placed them in a blank circle (Cowie et al., 1999b). The function of the words is to ensure consistency by directing subjects towards the average under-

standing of what a given position in activation-evaluation space means.

The dimension of time is represented indirectly by keeping the circles associated with recent mouse positions on screen, but having them shrink gradually over time (as if the pointer left a trail of diminishing circles behind it). The effect is to provide a visual image of the way ratings change over time.

We have reported experiments designed to validate the system with these features (Cowie et al., 2000). They indicate that in effect, FEELTRACE has the ability to discriminate static emotional states comparable to a vocabulary of around 20 non-overlapping emotion words. However, it also provides systematic representation of intermediate states, and allows the course of emotion to be mapped over time.

We regard that work as a prototype for the kind of development that the speech community as a whole can usefully undertake. It is a major challenge to find ways of expressing what emotionally coloured speech seems to convey. Our techniques address an interesting subset of the issues raised in Section 6 above, but by no means all. The conceptual resources exist to make progress in other areas if the speech emotion community chooses to accept the challenge.

8. Conclusion

Various ways of describing emotion have been outlined. One possible agenda for the speech community is to set about selecting from the range of possibilities those which relate most directly to speech. However, there is also a more radical possibility, which relates to the recurring theme of attempting to understand how emotion tends to be expressed in everyday interactions.

It seems not unlikely that a typical emotional episode will involve vocal elements related to many or all of the different aspects of emotion that have been considered. Consider, for example, an interaction where one partner expresses anger towards another. One might expect to find a close interlacing of vocal signs related not only to anger itself, but also to emotion-related states including hostility and distress. At peaks of anger, vocal patterns related to physiological changes might be expected. More generally, one might expect to find patterns associated with heightened activation and negative valence, and with attempts to achieve socially acceptable control. An overall assessment of the interaction would depend critically on the distribution and sequence of these elements in time. The outcome would depend not only on the signs themselves, but also on the way that they were detected, overlooked, or misinterpreted by the other party, and on the way he or she categorised the speaker's behaviour.

If that is a fair picture, then recovering the emotional content of a speech episode will depend on co-ordinating all the forms of description that have been considered, and understanding their various relationships to speech. It is not likely that that will be technically feasible in the short term, but the attempt to achieve it is a genuinely interesting challenge.

References

- Anscombe, E., Geach, P. (Eds.), 1970. *Descartes' Philosophical Writings*. The Open University, Nelson, UK.
- Arnold, M.B., 1960. *Emotion and Personality: Vol 1. Psychological Aspects*. Columbia University Press, New York.
- Averill, J.R., 1975. A semantic atlas of emotional concepts. *JSAS Catalog of Selected Documents in Psychology*, 5330 Ms. No. 421.
- Averill, J.R., 1980. A constructivist view of emotion. In: Plutchik, R., Kellerman, H. (Eds.), *Emotion: Theory, Research and Experience*, Vol. 1. Academic Press, New York, pp. 305–339.
- Averill, J.R., 1982. *Anger and Aggression: An Essay on Emotion*. Springer, New York.
- Averill, J.R., DeWitt G.W., Zimmer, M., 1978. The self-attribution of emotion as a function of success and failure. *Journal of Personality* 46, 323–347.
- Bachorowski, J., 1999. Vocal expression and perception of emotion. *Current Directions in Psychological Science* 8, 53–57.
- Banse, R., Scherer, K., 1996. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology* 70 (3), 614–636.
- Batliner, A., Fischer, K., Huber, R., Spilker, J., Nöth, E., 2003. How to find trouble in communication. *Speech Communication* 40 (1–2), this issue. PII: S0167-6393(02)00079-1.
- Blascovitch, J., 2000. Using physiological indexes of psychological processes in social psychological research. In: Reis, H.T., Judd, C.H. (Eds.), *Handbook of Research Methods in Social and Personality Psychology*. Cambridge University Press, Cambridge, pp. 117–137.
- Braunstein-Bercovitz, H., Dimentmann-Ashkenazi, I., Lubow, R.E., 2001. Stress affects the selection of relevant from irrelevant stimuli. *Emotion* 1, 182–192.
- Buck, R., 1999. Biological affects: a typology. *Psychological Review* 106, 301–336.
- Cacioppo, J.T., Klein, D.J., Berntson, G.G., Hatfield, E., 1993. The psychophysiology of emotion. In: Lewis, M., Haviland, J. (Eds.), *Handbook of Emotions*. Guilford Press, New York, pp. 119–142.
- Carroll, J.M., Russell, J.R., 1996. Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology* 70, 205–218.
- Cauldwell, R.T., 2000. Where did the anger go? The role of context in interpreting emotion in speech. In: Cowie, R., Douglas-Cowie, E., Schröder, M. (Eds.), *Proc. ISCA ITRW on Speech and Emotion: Developing a Conceptual Framework*, Newcastle, N. Ireland, 5–7 September 2000, Textflow, Belfast, pp. 127–131.
- Center for the Study of Emotion and Attention (CSEA-NIMH), 1999. *The International Affective Picture System: Digitized Photographs*. The Center for Research in Psychophysiology, University of Florida, Gainesville, FL.
- Clark, D.M., 1983. On the induction of depressed mood in the laboratory: evaluation and comparison of the Velten and musical procedures. *Advances in Behavioral Research and Therapy* 5, 27–49.
- Cornelius, R.R., 1996. *The Science of Emotion. Research and Tradition in the Psychology of Emotion*. Prentice-Hall, Upper Saddle River, NJ.

- Cornelius, R.R., Averill, J.R., 1980. The influence of various types of control on psychophysiological stress reactions. *Journal of Research in Personality* 14, 503–517.
- Corsini, R.J. (Ed.), 1994. *Encyclopaedia Psychology*. John Wiley, New York.
- Cowie, R., Douglas-Cowie, E., Romano, A., 1999a. Changing emotional tone in dialogue and its prosodic correlates. In: *Proc. ESCA Workshop on Dialogue and Prosody*, Eindhoven, The Netherlands, pp. 41–46.
- Cowie, R., Douglas-Cowie, E., Apolloni, B., Taylor, J., Romano, A., Fellenz, W., 1999b. What a neural net needs to know about emotion words. In: *Mastorakis, N. (Ed.), Computational Intelligence and Applications*. World Scientific Engineering Society, pp. 109–114.
- 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: *Proc. ISCA ITRW on Speech and Emotion: Developing a Conceptual Framework*, Newcastle, N. Ireland, 5–7 September 2000, Textflow, Belfast, pp. 19–24.
- Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J., 2001. Emotion Recognition in Human–Computer Interaction. *IEEE Signal Processing Magazine* 18 (1), 32–80.
- Crystal, D., 1969. *Prosodic Systems and Intonation in English*. Cambridge University Press, Cambridge.
- Crystal, D., 1975. *The English Tone of Voice*. Edward Arnold, London, UK.
- Darwin, C., 1872. *The Expression of Emotion in Man and Animals*. Murray, London (reprinted by University of Chicago Press, 1975).
- Davidson, R.J., 1999. Neuropsychological perspectives on affective styles and their cognitive consequences. In: *Dalgleish, T., Power, M. (Eds.), Handbook of Cognition and Emotion*. John Wiley, New York, pp. 103–123.
- Davidson, R.J., Ekman, P., Saron, C.D., Senulis, J.A., Friesen, W.V., 1990. Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology. *Journal of Personality and Social Psychology* 58, 330–341.
- Davidson, R.J., Jackson, D.C., Kalin, N.H., 2000a. Emotion, plasticity, context and regulation: Perspectives from affective neuroscience. *Psychological Bulletin* 126, 890–909.
- Davidson, R.J., Marshall, J.R., Tomarken, A.J., Henriques, J.B., 2000b. While a phobic waits: Regional brain electrical and autonomic activity in social phobics during anticipation of public speaking. *Biological Psychiatry* 47, 85–95.
- de Gelder, B., Vroomen, J., 2000a. The perception of emotions by ear and eye. *Cognition and Emotion* 14, 289–311.
- de Gelder, B., Vroomen, J., 2000b. Bimodal emotion perception: integration across separate modalities, cross-modal perceptual grouping, or perception of multimodal events? *Cognition and Emotion* 14, 321–324.
- Douglas-Cowie, E., Cowie, R., 2001. Speech and emotion: Reflections from the ISCA workshop. *Emotion Researcher* 15 (2), 8–10.
- Douglas-Cowie, E., Campbell, N., Cowie, R., Roach, P., 2003. Emotional speech: Towards a new generation of databases. *Speech Communication* 40 (1–2), this issue. PII: S0167-6393(02)00070-5.
- Ekman, P., 1992. An argument for basic emotions. *Cognition and Emotion* 6, 169–200.
- Ekman, P., 1994. Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. *Psychological Bulletin* 115, 268–287.
- Ekman, P., 1999. Basic Emotions. In: *Dalgleish, T., Power, M. (Eds.), Handbook of Cognition and Emotion*. John Wiley, New York, pp. 301–320.
- Ekman, P., Davidson, R.J. (Eds.), 1994. *The Nature of Emotion. Fundamental Questions*. Oxford University Press, Oxford.
- Ekman, P., Friesen, W., 1969. The repertoire of non verbal behavior: categories, origins, usage and coding. *Semiotica* 1, 49–98.
- Ekman, P., Friesen, W.V., 1978. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto.
- Ekman, P., Heider, K., 1988. The universality of a contempt expression: a replication. *Motivation and Emotion* 12, 303–308.
- Ekman, P., Levenson, R.W., Friesen, W.V., 1983. Autonomic nervous systems activity distinguishes between emotions. *Science* 221, 1208–1210.
- Engelbreton, T.O., Sirota, A.D., Niaura, R.S., Edwards, K., Brown, W.A., 1999. A simple laboratory method for inducing anger. *Journal of Psychosomatic Research* 47, 13–26.
- Fehr, B., Russell, J.A., 1984. Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General* 113, 464–486.
- Feldman, P.J., Cohen, S., Lepore, S., Matthews, K., Kasmark, T., Marsland, A.L., 1999. Negative emotions and acute physiological responses to stress. *Annals of Behavioral Medicine* 21, 216–222.
- Fernandez, R., Picard, R.W., 2003. Modeling drivers' speech under stress. *Speech Communication* 40 (1–2), this issue. PII: S0167-6393(02)00080-8.
- Finegan, J.E., Seligman, C., 1995. In defense of the Velten mood induction procedure. *Canadian Journal of Behavioral Science* 27, 405–419.
- France, D.J., Shiavi, R.G., Silverman, S., Silverman, M., Wilkes, D.M., 2000. Acoustical properties of speech as indicators of depression and suicide risk. *IEEE Transactions on Biomedical Engineering* 47, 829–837.
- Fridlund, A.J., 1994. *Human Facial Expression. An Evolutionary View*. Academic Press, San Diego.
- Frijda, N.H., 1986. *The Emotions*. Cambridge University Press, Cambridge.
- Frijda, N.H., 1993. Moods, emotion episodes, and emotions. In: *Lewis, M., Haviland, J. (Eds.), Handbook of Emotions*. Guilford Press, New York, pp. 381–403.
- Frost, R.O., Green, M.L., 1982. Velten mood induction procedure effects: duration and postexperimental removal. *Personality and Social Psychology Bulletin* 8, 341–347.

- Gerrards-Hesse, A., Spies, K., Hesse, F.W., 1994. Experimental inductions of emotional states and their effectiveness. *British Journal of Psychology* 85, 55–78.
- Gibson, J.J., 1979. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston.
- Goffman, E., 1981. *Forms of Talk*. University of Pennsylvania Press, Philadelphia.
- Gottman, J.M., Levenson, R.W., 1992. Marital processes predictive of later dissolution: behavior, physiology, and health. *Journal of Personality and Social Psychology* 63, 221–233.
- Gottman, J.M., Levenson, R.W., 2000. The timing of divorce: predicting when a couple will divorce over a 14-year period. *Journal of Marriage and the Family* 62, 737–745.
- Greasley, P., Setter, J., Waterman, M., Sherrard, C., Roach, P., Arnfield, S., Horton, D., 1995. Representation of prosodic and emotional features in a spoken language database. In: *Proc. XIIIth ICPHS, Stockholm*, Vol. 1, pp. 242–245.
- Greasley, P., Sherrard, C., Waterman, M., 2000. Emotion in language and speech: methodological issues in naturalistic approaches. *Language and Speech* 43, 355–375.
- Green, D.P., Goldman, S.L., Salovey, P., 1993. Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology* 64, 1029–1041.
- Gross, J.J., Levenson, R.W., 1995. Emotion elicitation using films. *Cognition and Emotion* 9, 87–108.
- Hadfield, P., Marks, P., 2000. This is your captain dozing. *New Scientist* 1682267, 21.
- Hagemann, D., Naumann, E., Maier, S., Becker, G., Luerken, A., Bartussek, D., 1999. The assessment of affective reactivity using films: Validity, reliability, and sex differences. *Personality and Individual Differences* 26, 627–639.
- Harré, R. (Ed.), 1986. *The Social Construction of Emotions*. Basil Blackwell, Oxford.
- Hatfield, E., Rapson, R.L., 1993. *Love, Sex and Intimacy: Their Psychology, Biology and History*. Harper Collins, New York.
- Healey, J., Picard, R., 1997. Digital processing of affective signals. Available from <http://www-white.media.mit.edu/cgi-bin/tr_pagemaker#TR444>.
- Hess, U., Banse, R., Kappas, A., 1995. The intensity of facial expression is determined by underlying affective state and social situation. *Journal of Personality and Social Psychology* 69, 280–288.
- Hucklebridge, F., Lambert, S., Clow, A., Warburton, D.M., Evans, P.D., Sherwood, N., 2000. Modulation of secretory immunoglobulin A in saliva; response to manipulation of mood. *Biological Psychology* 53, 25–35.
- Izard, C., 1972. *Patterns of Emotions: A New Analysis of Anxiety and Depression*. Plenum Press, New York.
- Izard, C., 1993. Organisational and motivational functions of discrete emotions. In: Lewis, M., Haviland, J. (Eds.), *Handbook of Emotions*. Guilford Press, New York, pp. 631–641.
- Jacobson, A.F., Weiss, B.L., Steinbock, R.M., Brauser, B., Goldstein, B.J., 1978. The measurement of psychological states by use of factors derived from a combination of items from mood and symptom checklists. *Journal of Clinical Psychology* 34, 677–685.
- James, W., 1884. What is emotion? *Mind* 9, 188–205.
- Johnson-Laird, P., Oatley, K., 1992. Basic emotions, rationality, and folk theory. *Cognition and Emotion* 6, 201–223.
- Kenealy, P., 1986. The Velten mood induction procedure: a methodological review. *Motivation Emotion* 10, 315–335.
- Kenealy, P., 1988. Validation of a music mood induction procedure: some preliminary findings. *Cognition and Emotion* 2, 41–48.
- Kenny, A., 1973. *Wittgenstein*. Penguin, London.
- Koukounas, E., McCabe, M.P., 2001. Emotional responses to filmed violence and the eye blink startle response: A preliminary investigation. *Journal of Interpersonal Violence* 16, 476–488.
- Lang, P.J., Bradley, M.M., Cuthbert, B.N., 1999. *International affective picture system IAPS: Technical manual and affective ratings*. The Center for Research in Psychophysiology, University of Florida, Gainesville, FL.
- Larsen, R.J., Sinnett, L.M., 1991. *Personality and Social Psychology Bulletin* 17, 323–334.
- Lazarus, R.S., 1994. The stable and the unstable in emotion. Fundamental questions. In: Ekman, P., Davidson, R.J. (Eds.), *The Nature of Emotion. Fundamental Questions*. Oxford University Press, Oxford, pp. 79–85.
- Lazarus, R.S., 1999a. *Stress and Emotion: A New Synthesis*. Springer, New York.
- Lazarus, R.S., 1999b. The cognition–emotion debate: a bit of history. In: Dalgleish, T., Power, R. (Eds.), *Handbook of Cognition and Emotion*. John Wiley, New York, pp. 3–19.
- Levenson, R.W., 1992. Autonomic nervous system differences among emotions. *Psychological Science* 3, 23–27.
- Lewis, M., Haviland, J. (Eds.), 1993. *Handbook of Emotions*. Guilford Press, New York.
- Luminet, O., Bouts, P., Delie, F., Mansted, A., Rim, B., 2000. Social sharing of emotion following exposure to a negatively valenced situation. *Cognition and Emotion* 14, 661–688.
- MacDowell, K.A., Mandler, G., 1989. Construction of emotion: discrepancy, arousal, and mood. *Motivation Emotion* 13, 105–124.
- Mandler, G., 1975. *Mind and Emotion*. Wiley, New York.
- Massaro, D., Cohen, M.M., 2000. Fuzzy logical model of bimodal emotion perception: Comment on ‘The perception of emotions by ear and eye’ by de Gelder and Vroomen. *Cognition and Emotion* 14, 313–320.
- Mehrabian, A., Ferris, S., 1974. Inference of attitude from nonverbal communication in two channels. In: Weitz, S. (Ed.), *Nonverbal Communication*. Oxford UP, New York, pp. 291–297.
- Mischel, W., Peake, P.K., 1982. Beyond déjà vu in the search for cross-situational consistency. *Psychology Review* 89, 730–755.
- Morrow, J., Nolen-Hoeksema, S., 1990. Effects of responses to depression on the remediation of depressive affect. *Journal of Personality and Social Psychology* 58, 519–527.
- Murray, I., Arnott, J., 1993. Towards the simulation of emotion in synthetic speech: a review of the literature on human

- vocal emotion. *Journal of Acoustical Society of America* 93, 1097–1108.
- Neisser, U., 1976. *Cognition and Reality*. W.H. Freeman, San Francisco.
- Nezlek, J.B., Wheeler, L., Reis, H.T., 1983. Studies of social participation. *New Directions for Methodology of Social and Behavioral Science* 15, 57–7.
- Oatley, K., Jenkins, J., 1996. *Understanding Emotions*. Blackwell, Oxford.
- O'Connor, J.D., Arnold, G., 1973. *Intonation of Colloquial English*. Longman, London.
- Ortony, A., Turner, T.J., 1990. What's basic about basic emotions? *Psychological Review* 97, 315–331.
- Ortony, A., Clore, G., Collins, A., 1988. *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge.
- Osgood, C., May, W.H., Miron, M.S., 1975. *Cross-cultural Universals of Affective Meaning*. University of Illinois Press, Urbana.
- Pereira, C., 2000. Dimensions of emotional meaning in speech. In: *Proc. ISCA ITRW on Speech and Emotion: Developing a Conceptual Framework*, Newcastle, N. Ireland, 5–7 September 2000, Textflow, Belfast, pp. 25–28.
- Picard, R.W., 2000. Toward computers that recognize and respond to user emotion. *IBM Systems Journal* 39, 705–719.
- Plutchik, R., 1980. *Emotion: A Psychoevolutionary Synthesis*. Harper and Row, New York.
- Plutchik, R., 1984. Emotions: A general psychoevolutionary theory. In: Scherer, K.R., Ekman, P. (Eds.), *Approaches to Emotion*. Erlbaum, Hillsdale, NJ, pp. 197–219.
- Plutchik, R., Conte, H.R. (Eds.), 1997. *Circumplex Models of Personality and Emotions*. American Psychological Association, Washington, DC.
- Polivy, J., 1981. On the induction of emotion in the laboratory: Discrete moods or multiple affect states? *Journal of Personality and Social Psychology* 41, 803–817.
- Polivy, J., Doyle, C., 1980. Laboratory induction of mood states through the reading of self-reference mood statements. *Journal of Abnormal Psychology* 89, 286–290.
- Reisenzein, R., 2000. Exploring the strength of association between the components of emotion syndromes: the case of surprise. *Cognition and Emotion* 14, 1–38.
- Roemer, L., 2001. Measures for anxiety and related constructs. In: Antony, M.M., Orsillo, S.M. (Eds.), *Practioner's Guide to Empirically Based Measures of Anxiety*. AABT Clinical Assessment Series, Kluwer Academic/Plenum, New York, pp. 49–83.
- Rolls, E.T., 1999. *The Brain and Emotion*. Oxford University Press, Oxford.
- Rosch, E., Mervis, C.G., 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7, 573–605.
- Roseman, I.J., 1991. Appraisal determinants of discrete emotions. *Cognition and Emotion* 5, 161–200.
- Ruch, W., 1993. Exhilaration and humour. In: Lewis, M., Haviland, J. (Eds.), *Handbook of Emotions*. Guilford Press, New York, pp. 605–616.
- Russell, J.A., 1994. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychology Bulletin* 115, 102–141.
- Russell, J.A., 1997. How shall an emotion be called. In: Plutchik, R., Conte, H. (Eds.), *Circumplex Models of Personality and Emotions*. APA, Washington, pp. 205–220.
- Russell, J., Barrett-Feldman, L., 1999. Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology* 76, 805–819.
- Salomon, R., 2001. Defining our terms. *Emotion Researcher* 152, 3–6.
- Schachter, S., 1959. *The Psychology of Affiliation*. Stanford University Press, Palo Alto, CA.
- Scherer, K.R., 1984a. Emotion as a multicomponent process: a model and some cross-cultural data. *Review Personality and Social Psychology* 5, 37–63.
- Scherer, K.R., 1984b. On the nature and function of emotion: a component process approach. In: Scherer, K.R., Ekman, P. (Eds.), *Approaches to Emotion*. Erlbaum, Hillsdale, NJ, pp. 293–317.
- Scherer, K.R., 1988. Criteria for emotion-antecedent appraisal: a review. In: Hamilton, V., Bower, G.H., Fridja, N.H. (Eds.), *Cognitive Perspectives on Emotion and Motivation*. Nijhoff, Dordrecht, pp. 89–126.
- Scherer, K.R., 1994. Toward a concept of modal emotions. In: Ekman, P., Davidson, R. (Eds.), *The Nature of Emotion: Fundamental Questions*. Oxford University Press, Oxford, pp. 25–31.
- Scherer, K.R., 1999. Appraisal theory. In: Dalglish, T., Power, M. (Eds.), *Handbook of Cognition and Emotion*. John Wiley, New York, pp. 637–663.
- Scherer, K.R., Walbott, H.G., 1986. How universal and specific is emotional experience? Evidence from 27 countries on five continents. *Social Science Information* 25, 1–14.
- Schlosberg, H., 1954. Three dimensions of emotion. *Psychology Review* 61, 81–88.
- 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.
- Schubiger, M., 1958. *English Intonation: Its Form and Function*. Niemeyer, Tubingen, NL.
- Shackelford, T.K., LeBlanc, G.J., Drass, E., 2000. Emotional reactions to infidelity. *Cognition and Emotion* 14, 643–659.
- Schneider, F., Gur, R.C., Gur, R.E., Muenz, L.R., 1994. Standardized mood induction with happy and sad facial expressions. *Psychiatry Research* 51, 19–31.
- 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.
- Speisman, J.C., Lazarus, R.S., Mordkoff, A., Davison, L., 1964. Experimental reduction of stress based on ego-defense theory. *Journal of Abnormal and Social Psychology* 68, 367–380.

- Spielberger, C.D., 1988. State-trait anger expression inventory (revised research edition). Psychological Assessment Resources, Odessa, FL.
- Stein, N., Oatley, K., 1992. Basic emotions: Theory and measurement. *Cognition and Emotion* 6, 161–168.
- Stemmler, G., Heldmann, M., Pauls, C.A., Scherer, T., 2001. Constraints for emotion specificity in fear and anger: the context counts. *Psychophysiology* 69, 275–291.
- Sternberg, R.J., 1988. Triangulating love. In: Sternberg, R.J., Barnes, M.L. (Eds.), *The Psychology of Love*. Yale UP, Yale.
- Stibbard, R., 2001. Vocal expression of emotions in non-laboratory speech. Unpublished PhD thesis, University of Reading, UK.
- Storm, C., Storm, T., 1987. A taxonomic study of the vocabulary of emotion. *Journal of Personality and Social Psychology* 53, 805–816.
- Teasdale, J.D., 1999. Multi-level theories of cognition–emotion relations. In: Dalgleish, T., Power, M. (Eds.), *Handbook of Cognition and Emotion*. John Wiley and Sons, Chichester, pp. 665–681.
- Tolkmitt, F., Scherer, K., 1986. Effect of experimentally induced stress on vocal parameters. *Journal of Experimental Psychology, Human Perception and Performance* 12 (3), 302–313.
- Tomkins, S.S., 1982. Affect theory. In: Ekman, P. (Ed.), *Emotion in the Human Face*. Cambridge University Press, New York, pp. 353–395.
- Tsapatoulis, N., Raouzaoui, A., Kollias, S., Cowie R., Douglas-Cowie, E., 2002. MPEG-4 FBA usage in emotion recognition. In: Pandzic, I.S., Forchheimer, R. (Eds.), *MPEG-4 Facial Animation – The Standard, Implementations, Applications*. John Wiley, Chichester.
- Velten, E., 1968. A laboratory task for induction of mood states. *Behaviour Research and Therapy* 6, 473–482.
- Vyas, E., Picard, R., 1999. Offline and online recognition of emotional expression from physiological data. Available from <http://www-white.media.mit.edu/cgi-bin/tr_page-maker#TR444>.
- Watson, D., Tellegen, A., 1985. Toward a consensual structure of mood. *Psychology Bulletin* 98, 219–235.
- Watson, D., Clark, L.A., Tellegen, A., 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology* 54, 1063–1070.
- 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.
- Whissell, C., 1989. The dictionary of affect in language. In: Plutchik, R., Kellerman, H. (Eds.), *Emotion: Theory, Research and Experience*, Vol. 4, *The Measurement of Emotions*. Academic Press, New York, pp. 113–131.
- Wichmann, A., 2000. The attitudinal effects of prosody, and how they relate to emotion. In: *Proc. ISCA ITRW on Speech and Emotion: Developing a Conceptual Framework*, Newcastle, N. Ireland, 5–7 September 2000, Belfast, Textflow, pp. 143–147.
- Wierzbicka, A., 1992. Talking about emotions: semantics, culture and cognition. *Cognition and Emotion* 6, 285–319.
- Williams, C.E., Stevens, K.N., 1972. Emotions and speech: some acoustical correlates. *Journal of Acoustical Society of America* 52 (2), 1238–1250.
- Wittgenstein, L., 1953. *Philosophical Investigations*. Blackwell, Oxford.
- Womack, B., Hansen, J., 1996. Classification of speech under stress using target driven features. *Speech Communication* 20 (1–2), 131–150.
- Wundt, W., 1903. *Grundzuge der Physiologischen Psychologie* Vol 2. Engelmann, Leipzig. (Originally published 1874.)
- Yartz, A.R., Hawk, L.W., 2001. Psychophysiological assessment of anxiety: tales from the heart. In: Antony, M.M., Orsillo S.M. (Eds.), *Practitioner's Guide to Empirically Based Measures of Anxiety*. AABT Clinical Assessment Series, Kluwer Academic/Plenum, New York, pp. 25–30.
- Zammuner, V.L., 1998. Concepts of emotion: “emotionness” and dimensional ratings of Italian words. *Cognition and Emotion* 12, 243–272.
- Zanna, M., Rempel, J.K., 1988. Attitudes: A new look at an old concept. In: Bar-Tal, D., Kruglanski, A.W. (Eds.), *The Social Psychology of Knowledge*. Cambridge University Press, Cambridge, pp. 315–334.
- Zuckerman, M., Lubin, B., 1985. *Manual for the Multiple Affective Adjective Checklist – Revised Edits*, San Diego, CA.