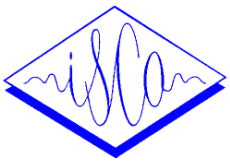
ISCA Archive
http://www.isca-speech.org/archive



ITRW on
Speech and Emotion
Newcastle, Northern Ireland, UK
September 5-7, 2000

DESCRIBING THE EMOTIONAL STATES EXPRESSED IN SPEECH

Roddy Cowie

Psychology, Queen's University, Belfast

ABSTRACT

Describing relationships between speech and emotion depends on identifying appropriate ways of describing the emotional states that speech conveys. Several forms of description are potentially relevant. The most familiar have obvious shortcomings, and alternatives urgently need to be explored.

1. INTRODUCTION

One of the obstacles to research on speech and emotion is the lack of a settled approach to describing emotion. Presumably part of the problem comes with the territory. Emotion belongs to a substrate of experience that existed long before humans began to use linguistic symbols, and it seems to actively resist attempts to capture it in symbols. On the other hand, part of the problem is probably self inflicted. Investigators have rarely invested serious effort in finding suitable descriptive systems.

This paper responds to that situation in two ways. First, it highlights key problems that go with the territory. They are reflected in the fact that no existing descriptive framework is wholly satisfactory. The problems surrounding everyday description are particularly revealing. Second, it aims to encourage the speech community towards standardisation of key terms and descriptive techniques.

The point of standardisation is not to narrow choice. On the contrary, it is a core aim of the paper to ensure that a range of descriptive tools and frameworks are recognised as legitimate options, so long as they are used appropriately.

One might expect descriptive frameworks to be provided ready-made by research on emotion *per se*. The view taken here is that speech raises particular problems that have to be worked through, and hopefully in the long run reintegrated into a rounded picture of emotional phenomena. In the meantime, other lines of research can certainly inform the choice of descriptive frameworks, but it would be wrong to expect them to provide definitive prescriptions.

Most of the paper is spent summarising alternative descriptive frameworks, and considering their uses and limitations. Before that, though, there is a preliminary task, which is to consider the bounds of the domain that is to be described.

2. THE DOMAIN: FULLBLOWN EMOTIONS AND EMOTIONAL STATES, CAUSE AND EFFECT

The word emotion is semantically treacherous. In everyday use, its reference shifts according to context. That makes it a very flexible tool, but it creates havoc when the word is taken out of context and used to describe a field of research. That point has been made, in various ways, by distinguished scholars from William James at the end of the nineteenth century to James Russell at the end of the twentieth [1],[2]. The subtleties of the issue are beyond the scope of this paper, but one simple step is useful. It involves distinguishing two senses, and proposing terms that allow them to be separated.

The first sense uses the word to refer to entities – natural units that have distinct boundaries, and that can be counted, so that the word has a perfectly straightforward plural. That is the sense that is involved when we say that fear and anger are two distinct emotions. The second sense uses the word to refer to an attribute of certain states. That is the sense that is involved when we say that somebody's voice is tinged with emotion.

Research on emotion in psychology and biology has tended to emphasise the first meaning. It has looked for discrete states that deserve to be called emotions in the fullest possible sense. It is entitled to do that, and to some extent it has been vindicated by success. Although disagreements persist, a core of agreement about those states seems to be emerging. The problem is that it is not at all clear whether that sense of the word is the most interesting one for research on speech. After all, it is commonly taken as a mark of emotion in the fullest sense that it leaves people either speechless or incoherent.

The issue can be put in terms of applications. It is not obvious how useful it would be to know how to synthesise and recognise the vocal signs associated with (for instance) pure unadulterated rage or pure unadulterated bliss. A simulated agent that showed either of those emotional states would be deeply disturbing. Similarly, it is doubtful whether there would be many applications for a recognition system that noticed nothing untoward until the user reached a state of pure unadulterated rage. From examples like those, it seems that shades of emotion are the obvious focus for speech research.

It is easier to address the issue if terms for the distinction can be agreed. The first sense is well captured by the phrase fullblown emotions. The second is more difficult, but the term 'emotional states' is serviceable. It is worth adding a third phrase, 'emotional systems'. It captures reasonably accurately what biologically oriented research has come to prioritise.

The distinctions are intended to make it clear that the broad idea of research on emotion subsumes several more specific objectives, each of which is perfectly legitimate. There should be ample scope for research on one objective to support research on the others, provided that there is a clear recognition of the distinctions. To borrow a phrase from the poet Robert Frost, 'Good fences make good neighbours'.

The distinctions also have a major impact on the issue of description. If the aim of research is to understand fullblown emotions or emotional systems, then arguably the descriptive problem boils down to agreeing on a small set of categories. If the aim is to understand emotional states, then some way has to be found of describing emotional states in everyday life, with all the shades and ambivalences that we know distinguish them. That is a very different matter, and it is the task that takes up the bulk of this paper.

The second issue in this section also involves basic aims. There are two distinct types of description that an investigator might consider. One type of description would identify the emotion-related internal states and external factors that caused a person's speech to have particular characteristics. The other would describe what effect those characteristics would be likely to have on a typical listener. It is natural to call these cause- and effect-type descriptions respectively.

The two often coincide, because typical listeners are quite good at inferring the causes of emotion-related features in speech. Clearly they sometimes diverge, though. Listeners do misread vocal signs of emotion, either because they are misdirected, or because well-intentioned signals are misconstrued.

Both types of description are legitimate objects of study, and investigators are entitled to choose which they prioritise. It is important to recognise, though, that the choice can have far-reaching consequences. Prioritising cause-type descriptions tends to focus attention on verifiable properties of a speaker's internal state, which in turn tends to favour describing emotion in terms of physiological systems. It also tends to focus attention on signs that naïve listeners may not detect. Prioritising effect-type descriptions favours describing emotional states in terms of categories and dimensions that people find natural, and using signs that they are able to detect.

Again, the position taken here is that good fences make good neighbours. Both choices make sense in particular application areas. Research on vocal signs of stress is a clear case where cause-type descriptions are appropriate. It is all to the good if an automatic system can detect stress in a pilot's voice before the passengers can. On the other hand, virtual agents that are to interact with people probably need to understand emotion in human-like terms. There is no point in virtual agents giving signs that are objectively valid, but that human listeners are poor at detecting; and there would probably be limits to the acceptability of virtual agents that detected the user's true emotion however he or she tried to conceal it.

An important implication is that there is no absolute obligation on research to collect cause-type information about samples of emotional speech. For effect-oriented research, it is of secondary importance what a speaker's state actually was when a speech sample was recorded. The main issue is the impression that the speech creates in listeners. Note that adopting that orientation is not taking an easy option. It is a serious technical challenge to find ways of describing the kind of impression that speech creates in a listener. One of the themes in later sections is how that challenge might be met.

3. CATEGORY LABELS

Far the most obvious approach to describing emotion is to use the category labels that are provided by everyday language – labels such as fear, anger, contentment, and so on. This section looks at the various ways in which category labels may be used to describe emotions and emotional states.

A central theme in the section is that there are intimate links between everyday categorisation and the problems that go with the territory. Everyday categorisation is a deceptively sophisticated system that has developed to handle an exceedingly complex set of issues. Analysing it with due respect provides useful insights into the issues that research should ideally address in the long term. Attempting to use it immediately and uncritically is a recipe for trouble.

3.1 Basic emotion categories

Probably the best known theoretical idea in emotion research is that certain emotion categories are primary, others are secondary. Like so much accepted wisdom, the idea can be traced to Rene Descartes [3].

The idea of primary emotions has had an enormous effect on the description of emotion. It suggests that the natural starting point for research is to obtain a list of the primary emotions, and then to study how each of the emotions on that list is reflected in speech. It often seems to be assumed that once that has been done, secondary emotions will fall into place.

Evaluating that view depends on recognising that Descartes' idea has two very different components. One is that a few emotional states are pure and primitive in a way that the rest are not. The other is that the rest are derived from the primitive states by mixing them rather like primary colours. The second component has been called a palette theory of emotion.

The first component has had its ups and downs, but it is well regarded in modern psychology. A wide range of theorists agree that fullblown emotions take only a few forms, which are qualitatively distinct from each other. Each form is a syndrome, distinguished by the way a cluster of features come together.[4], [5], [6].These syndromes tend to be described as basic emotions, because the term primary is associated with the second part of Descartes' idea, and it has fared much less well. It would be rash to say that there is no support for palette theories, but if there is, it is a very select minority.

That situation has implications for the status of traditional lists of basic emotions. On one hand, if basic emotions are qualitatively distinct syndromes, then a list of categories is an appropriate way of describing them. On the other hand, that view of basic emotions offers no support for the idea that knowledge about the items on that list will transfer in any straightforward way to any other emotional state.

In short, a list of basic emotion categories is an appropriate starting point for research whose aim is to study the speech patterns associated with basic emotions. Studying the

relationship between speech and other, more commonplace emotional states is a different problem, and progress depends on finding forms of description that apply to those states.

It is worth adding for completeness that there is no definitive list of basic emotions. There is quite general agreement on the so-called 'big six' – fear, anger, happiness, sadness, surprise, and disgust [7]. Common ways of extending the group include distinguishing hot and cold anger, and adding contempt [8] and love – which may be divided into sexual and other types [5].

3.2 Second order emotion categories

Everyday language contains an abundance of emotion-related categories. To illustrate, a collection due to Whissell [9] lists 107 words describing emotional states, and one due to Plutchik [10] lists 142. The words cover a great range of emotional states, very few of which could be regarded as basic.

The only established term for the states that are not basic is 'secondary emotions'. The word carries an unfortunate implication that the states are less important, and it is worth establishing a more neutral alternative. 'Second order emotions' seems natural. It reflects the reasonable presumption that they are more complex than basic emotions in some sense, without too many other unwarranted overtones.

One option for research on speech is to exploit the power of everyday emotion language to the full by using the whole range of second order terms, and combining them to capture even subtler shades. Some research has followed that path – e.g. the database collected by the Reading group [11] describes an utterance as moving from 'hate' to 'vengeful anger'.

That approach deserves to be taken seriously, because it does not throw away very much information. If descriptions were detailed enough, investigators could expect to revisit data with fair assurance that they could identify the original state. However, research would be very unlikely to progress if it restricted itself to that level of description. If we treated every emotion word as an irreducibly distinct category, then the problem of accumulating information about speech correlates would be thoroughly intractable. There is very little prospect of accumulating a substantial body of data on the speech patterns associated with each of a hundred and forty categories, let alone the thousands that can be formed by combining terms.

It is reasonably clear what is needed. Somehow the fine-grained descriptive system provided by everyday categories needs to be embedded within a complementary representation that offers ways of drawing fewer, grosser distinctions, for which there is more chance of finding reliable speech correlates. That is task is taken up in later parts of the paper.

There is a second side to the argument for embedding everyday descriptions in a complementary representation, and it is in some senses more fundamental. Common experience indicates that even fine-grained linguistic categories do not capture every shade of emotion that people can distinguish. Pictorial art provides a neat way of making the 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*. One of the functions of a complementary representation is to define emotional configurations that are possible, and perhaps even important, but for which there is no word (as yet).

3.3 Emotion-related states: arousal

Everyday emotion terms are surrounded by terms that people feel bear a strong family resemblance to them, notwithstanding differences on various levels. It is natural to call them emotion-related terms. It is an important issue for speech research how it approaches these terms and the states associated with them.

One of the key issues is where the resemblances and distinctions lie. In particular, it matters whether states are perceived as emotion-related partly because they share vocal characteristics with emotions proper. If so, then it is natural for research on speech to consider them together, even if there are radical differences between the states at (for instance) the level of biological systems. People often resist the idea that different ways of drawing boundaries may be appropriate for different purposes, but it is not an uncommon situation – as people in Northern Ireland have particular reasons to know.

One of the most difficult boundaries to draw is between emotional states proper and states that involve arousal without some of the other characteristics of emotion. To illustrate the point, Frijda's definition [12] makes happiness a rather marginal example of emotion: he defines emotions in terms of readiness to take specific actions, and happiness involves generalised activation rather than a specific action tendency.

States of arousal are among the strongest candidates for vocal overlap with emotion proper. Research has studied various states involving arousal, and described vocal correlates that are at least broadly similar to variables associated with emotion. Examples are excitement, agitation, and lethargy. One with particular practical significance is stress, which has become a sub-topic in its own right [13].

The main point to be made here is that an adequate descriptive framework ought to identify states of arousal as at least near neighbours of emotionality, at least insofar as vocal expression is concerned. They are sufficiently close that it is a research question whether the categories can be distinguished vocally. For instance, it is not obvious whether the arousal associated with stress can be distinguished reliably from the arousal associated with happiness. It would be reasonable to find out before installing equipment that would initiate emergency procedures every time a pilot received good news. The only way to ensure an answer is to ensure that speech research does not divide itself into watertight compartments that may not be appropriate for its purposes.

3.4 Emotion-related states: attitude

Another difficult boundary is between emotion terms and terms that refer to attitude. The term attitude is widely used both in linguistics and in social psychology. Various definitions have been proposed in psychology, but similarities to emotion are common ground. A standard summary is that attitude entails 'categorisation of a stimulus object along an evaluative dimension' [14]. Since evaluation is accepted as a fundamental dimension of emotionality (see section 5), that implies overlap between attitude and emotion. Some theorists go further and explicitly link attitude to affect (i.e. what distinguishes emotional states from dispassionate rationality)[15] [16].

Linguists' use of the term may extend beyond psychologists'. Speech is said to convey attitudes in which evaluation and/or affect are at least not salient. Examples might be businesslike, or inquisitive, or formal. States of that kind are too important for speech research to let a nominal barrier marginalise them. A definition that covers both those and more overtly 'affective' attitudes is that a person who exhibits a particular attitude approaches a situation prepared to find certain kinds of problem or opportunity, and to take certain kinds of action.

The key issues surrounding attitude are of familiar types.

The overlap in definition suggests that an adequate descriptive framework should mark attitude as at least a near neighbour of emotionality. Prima facie, it seems they may well be particularly close in terms of vocal expression. Research in either domain is lacking unless it clarifies which vocal signs it shares with the other, and which distinguish the domains.

Attitude terms are even more numerous than terms describing second order emotion. Articles by Schubiger [17] and O'Connor and Arnold [18], for example, used nearly 300 labels between them. These cover states such as 'abrupt, accusing, affable, affected, affectionate, aggressive, agreeable, airy, amused, angry, animated, annoyed, antagonistic, apologetic, appealing, appreciative, apprehensive, approving, argumentative, arrogant, authoritative …'. The prospects of finding unique speech correlates for every one are even more remote than the prospects of finding unique speech correlates for every second order emotional state. Correspondingly, there is even more need to find ways of embedding the fine-grained descriptive system offered by everyday language in a complementary representation that allows broader distinctions to be drawn.

3.5 Emotion & everyday terms: reprise

Reviewing everyday descriptions highlights the complexity of what people do when they apply an emotion-related term to another person. They judge that the person's state conforms to one of several hundred recognisable patterns. Each pattern involves many types of variable. Most involve degree of arousal, orientation towards certain aspects of the situation, evaluation of those aspects, and disposition to act in certain ways. Some involve much more. For example, a word like 'vengeful' does not simply describe a feeling. It carries implications about past events – there must have been some past action for which revenge is sought. It carries information about long term goals – vengefulness means that action will be taken against the person who carried out the past action, not necessarily in the short term. It carries overtones of moral judgement – someone who is vengeful claims a kind of moral justification, in a way that someone who is jealous does not.

The fullest examples of that kind of pattern are second order emotions. Other types seem to share some of the structure that occurs in those cases. Basic emotions involve strong changes in arousal, with appraisal reduced to a focused minimum. Attitude is diametrically opposite, with limited arousal and potentially complex appraisal. Arousal terms suggest an attempt to retain normal appraisal patterns alongside changes related to primary emotion.

It is clear that speech has rich interfaces with that system. Speakers modulate their voices in ways that reflect where they are currently positioned in it, and listeners use vocal signs to infer where speakers are located in it. The research task that interests our group is to understand how that is possible, and preferably how it can be simulated.

Putting the problem in that way is meant to make it clear how complex the task is, and how limited progress has been to date. It is also meant to underline uncertainty about the natural boundaries of research on the vocal signs of emotion. The signs that convey attitude may conceivably be quite different from those that convey second order emotion. Equally, though, the same system of signs may be involved in conveying aspects of appraisal whether or not they are part of a second order emotion, or in conveying arousal whether or not it is part of an emotion proper. The issue ought not to be prejudged. It is part of research on emotion in speech to establish whether the speech variables involved in signalling emotion do so as part of a wider function.

A complementary point is that everyday descriptions of emotional states are not designed to be based on speech variables alone. Judgements about emotion usually integrate information from vision, speech content, perception of the context, and often prior knowledge about the person involved. It seems quite likely that speech variables alone support a far more limited range of range of distinctions. Representing the kind of information that they carry is key problem.

Considering only basic emotions certainly limits the number of distinctions to be drawn. The problem is that the states involved are very special, and it is not obvious either that they are directly relevant to many applications, or that there are good prospects of generalising from them to states that are relevant to applications.

The next sections consider an alternative kind of response. 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 problem of category numbers may be to look beyond category terms to the structure that underpins their meaning. There are several interesting approaches to that task.

4 BIOLOGICAL REPRESENTATIONS

It is widely assumed that everyday descriptions of emotions are effectively surrogates for descriptions of physiological states. On that account, the ideal response to problems with verbal descriptions is to replace them with physiological parameters.

There are well known philosophical objections to that idea. The validity of a claim to experience remorse, it is argued, lies in the actions that follow it, not in the physiological state that accompanies it. That kind of argument should be taken seriously, but it is overshadowed for the time being by a simpler problem, which is that physiological states cannot be measured with anything like the resolution required to discriminate fine shades of emotion.

To illustrate the situation, consider work by Picard and her colleagues at MIT. They have used state of the art computing to classify emotion on the basis of physiological measurements. In 1997 they were able to distinguish anger from peaceful emotions with about 90% accuracy, and high and low arousal states with about 80% accuracy; but positive and negative emotions were not well distinguished [19]. More recently, additional measurements and improved computing have raised discrimination rates to around 80% for a set of 8 emotions [20]. Note, though, that the performance is on a single subject inducing strong emotional states. Transfer to more challenging tasks remains to be seen.

In the long term, there may well be interesting correlations between speech variables and measures like those of the Picard team. However, there is some way to go before physiologically based description provides the kind of objective, reliable evidence about emotions that people often assume it can.

New brain imaging techniques add another dimension to physiological observation. Research on brain mechanisms of emotion is a rapidly growing field (for a recent review, see [21]). 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. Nevertheless, it remains to be seen what level of discrimination the techniques offer.

In general, the use of physiology to describe emotion fits a familiar pattern. It is an approach that investigators are entitled to take, but not obliged to. It is particularly appropriate where the aim is to achieve cause-type description, as (for instance) in applications such as stress detection. It may be important in effect-type description when human judgement is mediated by recognising signs related to physiology – such as a dry mouth or rapid breathing – and use them to infer emotional states. That approach came to prominence through the work of Stevens and his colleagues [22], and it has been elegantly developed by the Geneva group [23]. It seems to be relevant to strong emotions associated with preparation for 'fight or flight'. It is less clear how far beyond that it applies.

5. CONTINUOUS REPRESENTATIONS

A widely used approach to describing the domain of emotional states is to assume that they correspond to co-ordinates in a space with a small number of dimensions [24], [25]. From it derives a type of representation that is both simple and capable of capturing a wide range of significant issues in emotion. We have called it activation-evaluation space [26]. It rests on a simplified treatment of two themes.

Valence Emotional states are characteristically 'valenced', i.e. they are permeated by positive or negative evaluations of people or things or events. The link between emotion and valence is widely agreed, although authors describe it in different terms. Arnold [27] refers to the "judgement of weal or woe"; Rolls [21] sees emotional processing as where "reward or punishment value is made explicit in the representation" (p.6); Tomkins [28] describes affect as what gives things value – "without its amplification, nothing else matters, and with its amplification, anything else can matter".

Activation level Research from Darwin on has recognised that emotional states involve dispositions to act in certain ways. A well known extension is Frijda's [12] proposal that emotions equate with action tendencies. A basic way of reflecting that theme turns out to be suprisingly useful. States are simply rated in terms of the associated activation level, i.e. the strength of the person's disposition to take some action rather than none.

The axes of activation-evaluation space reflect those themes. The vertical axis shows activation level, the horizontal axis evaluation. A basic attraction of that arrangement 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 that conclusion, including factor analysis, direct scaling, and others [25].

Words describing fullblown emotions are not evenly distributed in the space. Instead they tend to form a roughly circular pattern. From that and related evidence, Plutchik has argued that there is a circular structure inherent in emotion. That opens interesting avenues. For example, it suggests that points can be described in terms of an angular measure, which we have called emotional orientation; and distance from the centre, which we have called emotional strength. The concept of a fullblown emotion can then be translated roughly as a state where emotional strength has passed a certain limit. An interesting extension is to think of primary or basic emotions as cardinal points on the periphery of an emotion circle.

Activation-emotion space is a powerful tool, and it has proved attractive to computationally oriented research [26]. However, it has to be emphasised that the representation depends on collapsing the structured, high-dimensional 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). The difficulty is that neither 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.

Descriptions based on activation-evaluation space open various avenues for research on speech. A neat application is described by Schroeder [29]. He has used ratings in a dimensional space to measure the distance between the emotions involved when people misclassify an affect burst. The distances between the emotions indicate how serious confusions are.

More radically, speech variables could be correlated with dimensions rather than with discrete categories. If necessary, categorical descriptions could be recovered via a look-up table giving the categories associated with specified co-ordinates.

In the case of activation, there is some empirical support for that approach. Positive activation appears to be associated with increased 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, all of which involve positive activation. Negative activation appears to be associated with decreased mean and range of F0 – as suggested by studies of sadness, grief, and to a lesser extent boredom [26].

6. STRUCTURAL MODELS

Dimensional techniques represent one systematic approach to describing possible emotional states in a coherent framework, The natural alternative is associated with the approach to emotion described as cognitive. It has been argued that distinct types of emotion correspond to distinct ways of appraising the situation that evokes the emotion. That has prompted attempts to set out logical primitives that can be used to generate appraisals coresponding to distinct emotional states.

Scherer [30] provides a wide ranging review of that approach. To illustrate it, consider two substantial proposals, one due to Roseman [31] and the other to Ortony et al [32]. 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, based on the idea that appraisals may emphasise different kinds of element. 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 oneself and various others.

Accounts of that 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 in the mind (remembered or anticipated or whatever). The accounts also suggest hypotheses about states that voice alone might not be sufficient to identify, because the relevant appraisals involve multiple agents and events in complex relationships – e.g. pity, or remorse, or gratitude.

It should be noted that that kind of framework applies very naturally to attitude, reinforcing the point made in section 3 that its affinities with emotion are too strong to take lightly.

7. MATTERS OF TIMING

Timing is an issue whose relevance to emotion is increasingly accepted, and yet it is often not fully integrated into descriptions. 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 timecourses. 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 envisage a system that used speech to accumulate evidence on shifts in a user's mood over a period of hours or days. Most people can think of individuals who might benefit if they had feedback from a system of that kind.

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 notorious that people find at least some individuals difficult to read emotionally unless they have enough experience to know the relevant baselines.

With regard to fullblown emotional episodes, timing may well be diagnostic in several senses. An effective synthesiser needs to release and sustain signals of emotion on the right time scale. 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 timecourse of emotional signals in speech. Our impression is that that 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 domain of feeling or appraisal or action tendency.

8. MIXING AND MASKING

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 most obvious type of interaction is restraint. Ekman and his co-workers [33] 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 our exploratory work convinces us that they are a vital topic. Strong underlying emotion is often signalled by unnatural behaviour arising from determination not to release socially unacceptable signs. It is even more revealing when socially unacceptable signs surface briefly in spite of determination to control them.

Research can only begin to address that whole system of signals if it acknowledges that the attempt to observe display rules is an integral part of emotional life. The issue is particularly acute in the context of speech, because completely unrestrained emotion seems to be incompatible with speech as such – which raises interesting questions for research on emotion in general.

A second type of interaction, also highlighted in our exploratory work, 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. Our impression is that some kinds of mix may be commoner than the pure versions – for instance, sadness is often tinged with anger or a kind of pleasure. The implication is straightforward. If mixes are common, then speech research needs to acknowledge them as a feature of the domain that it deals with. Temporal issues may be central to doing that – the lead role often seems to shift back and forth between mixed emotions.

A major topic is raised here for want of a better place. It is humour. Humour appears to have strong links to both control and emotional mixture. It may express anger or bleakness or happiness, and our explorations suggest that it is very often used as the preferred way of signalling these emotions without violating display rules. A useful way of making the point is in terms of artificial agents. If they are going to show emotion, we would surely hope that they would show a little humour too.

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 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 that is a real issue for agents that are intended to convey emotion.

9. FROM PRINCIPLES TO PRAGMATICS

The issues reviewed in this paper are not abstractions. They translate into ideas about the tools needed to describe emotion for the purpose of analysing its relationship to speech. Our group has begun a range of developments in that area, and reviewing them is a useful way of flagging key issues.

Naturalistic databases are fundamental to studying emotion as it spontaneously occurs - mostly second order, and strongly constrained by display rules. We have developed one that is described in another paper in this conference [34].

Dimensional approaches are well suited to studying the time course of emotion, because they allow a few parameters to be followed over time. We have exploited that possibility in a system called Feeltrace, which uses activation/evaluation space to let observers record their impression of a person's emotional state as it fluctuates in real time [26].

Alongside Feeltrace, we have tried to develop a 'basic emotion vocabulary'; that is, a set of emotion categories that is small enough to be tractable, but that covers the range of emotional states that commonly occur [35]. Naïve subjects chose a vocabulary that met the requirement. The result was very different from lists of basic emotions. The approach needs to be refined, but the underlying idea is well worth pursuing.

The same study set out to elicit structural models that subjects regarded as capturing the meaning of selected emotion words. Again, the approach needs refinement, but it points to a significant ideal – to provide a dictionary in which emotion words are systematically explicated in terms of dimensional co-ordinates and logical primitives. That would provide inter-translation between the various forms of description that have been sketched here, and allow investigators to use whichever best fitted the demands of a particular study.

In all of those areas, consensus is essential. The complexity of the domain demands databases on a scale that is likely to require massive co-operation. Systems like Feeltrace can generate dire confusion if the procedures for using them are not thoroughly standardised. Selected emotion vocabularies need to be agreed, or else the problem of innumerable labels simply transmutes into one of incompatible vocabularies. Various types of structural model could be useful, but a plethora of competing models would only deepen existing confusion.

One use of a meeting like this is to begin movement towards consensus on the topics raised here, from issues of domain and basic vocabulary to technical implementations. We hope that discussion will consider the possibility.

10. REFERENCES

1. James, W .(1884) What is emotion? *Mind* 9, 188-205
2. Russsell, J Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the

elephant.*Journal of Personality and Social Psychology*. 76 805-819.

3. Anscombe, E. & Geach, P. (eds) (1970*) Descartes Philosophical Writings*. Nelson: The Open University.

4. Ekman, P. (1999) Basic Emotions. In T. Dalgleish & M. Power (eds) *Handbook of Cognition and Emotion*. New York: John Wiley.

5. Oatley, K. & Jenkins, J. (1996*) Understanding Emotions*. Oxford: Blackwell.

6. Scherer, K. R. (1994) Towards a concept of 'modal emotions'. In P. Ekman & R. Davidson (eds*) The Nature of Emotion: Fundamental Questions*.Oxford: OUP.

7. Cornelius, R. (1996) *The Science of Emotion*. New Jersey: Prentice-Hall.

8. Ekman, P. & Heider, K. (1988) The universality of a contempt expression: a replication *Motivation & Emotion* 12, 303-308.

9. Whissell, C. The dictionary of affect in language. In R. Plutchnik & Kellerman, H. (eds) (1989*) Emotion: Theory, research and experience: vol 4, The measurement of emotions*. New York: Academic press.

10. Plutchik, R. (1980) *Emotion: A psychoevolutionary synthesis*. New York: Harper and Row.

11. Greasley. P. et al. (1995) Representation of prosodic and emotional features in a spoken language database. *Proc XIII ICPhS*, Stockholm.

12. Frijda, N. H. (1986) *The emotions*. Cambridge: CUP.

13. Murray, I. , Baber, C. & South, A. (1996) Towards a definition and working model of stress and its effects on speech. *Speech Communication* 20, 1-12.

14. Zanna, M. & Rempel. (1988) Attitudes: A new look at an old concept. In D. Bar-Tal & A. W. Kruglanski (eds) *The social psychology of knowledge*. Cambridge: CUP.

15. Fishbein, M. & Ajzen, I. (1975) *Belief, attitude, inention and behavior: An introduction to theory and research*. Reading, MA: Addison Wesley.

16. Breckler, S. J. & Wiggins, E. C. (1989) On defining attitude and attitude theory. Once more with feeling. In A. R. Pratkanis, et al (eds) *Attitude structure and function*. Hillsdale, NJ: Erlbaum.

17. Schubiger, M. (1958*) English intonation. Its form and function*. Tubingen: Niemeyer.

18. O'Connor, J. D. & Arnold, G. (1973*) Intonation of colloquial English*. London: Longman.

19. Healey, J. & R. Picard (1997) Digital processing of affective signals. http://www-white.media.mit.edu/cgi-bin/tr_pagemaker#TR444

20. Vyas, E. & Picard, R. (1999) Offline and online recognition of emotional expression from physiological data.http://www-white.media.mit.edu/cgi-bin/ tr_pagemaker#TR444

21. Rolls, E. T. (1999) *The brain and emotion*. Oxford: Oxford University Press.

22. Williams, C. & Stevens, K. Emotions and speech: some acoustic correlates. *JASA* 52, 1238-1250.

23. Banse, R. & Scherer, K. (1996) Acoustic profiles in vocal emotion expression. *Journ of Personality & Social Psychol* 70 (3), 614-636.

24. Schlosberg, H. (1954) A scale for judgement of facial expressions. *Jour of Experimental Psychol* 29, 497-510

25. Russell, J. A. (1997) How shall an emotion be called? In R. Plutchik & H. Conte (eds*) Circumplex Models of Personality and Emotions*. Washington: APA.

26. Cowie, R., Douglas-Cowie, E. et al. (at press) Emotion Recognition in Human-Computer Interaction. To appear in *IEEE Signal Processing Magazine*.

27. Arnold , M. B. (1980*) Emotion and personality, vol 2: Physiological aspects* New York:Columbia Univ Press.

28. Tomkins, S. S. (1982) Affect theory. In P. Ekman (ed) *Emotion in the human face*. New York: CUP.

29. Schroeder, M. (2000) Experimental study of affect bursts. *This volume*.

30. Scherer, K. R. (1999) Appraisal theory. In T. Dalgleish & M. Power (eds) *Handbook of Cognition and Emotion*. New York: John Wiley.

31. Roseman, I. J. (1991) Appraisal determinants of discrete emotions. *Cognition and Emotion* 5, 161-200.

32. Ortony, A., Clore, G. & Collins, A. (1988*) The cognitive structure of emotions*. Cambridge: CUP.

33. Ekman, P. & Friesen, W. (1969) The repetoire of non verbal behavior: categories, origins, usage and coding. *Semiotica* 1, 49-98.

34. Douglas-Cowie, E., Cowie, R. & Schroeder, M. (2000) A new emotion database: considerations, sources and scope. *This volume*.

35. Cowie, R. et al. (1999) What a neural net needs to know about emotion words. In N. Mastorakis (ed) *Computational Intelligence and Applications*. World Scientific Engineering Society. pp. 109-114