Outline

Tables (2)

Neural Networks
Volume 18, Issue 4, May 2005, Pages 371-388
#
2005 Special Issue
Beyond emotion archetypes: Databases for emotion modelling using neural
networks
Author links open overlay panelRoddy Cowie, Ellen Douglas-Cowie, Cate Cox
Show more

https://doi.org/10.1016/j.neunet.2005.03.002Get rights and content
Abstract
There has been rapid development in conceptions of the kind of database that
is needed for emotion research. Familiar archetypes are still influential, but
the state of the art has moved beyond them. There is concern to capture
emotion as it occurs in action and interaction (?pervasive emotion?) as well
as in short episodes dominated by emotion, and therefore in a range of
contexts, which shape the way it is expressed. Context links to modality-
different contexts favour different modalities. The strategy of using acted
data is not suited to those aims, and has been supplemented by work on both
fully natural emotion and emotion induced by various technique that allow more
controlled records. Applications for that kind of work go far beyond the
?trouble shooting? that has been the focus for application: ?really natural
language processing? is a key goal. The descriptions included in such a
database ideally cover quality, emotional content, emotion-related signals and

signs, and context. Several schemes are emerging as candidates for describing pervasive emotion. The major contemporary databases are listed, emphasising those which are naturalistic or induced, multimodal, and influential.
* Previous article in issue
* Next article in issue
Keywords
Database
Emotion
Multimodal
Elicitation
Naturalistic
It is a truism that a neural network can only achieve what its training data allows it to. In the area of research on emotion, the quality and quantity of data available is a major issue. This paper aims to inform the research community about the collections of data that are available, and to inform it about ongoing developments in the area.
One of the key developments is growing recognition that the task of assembling adequate databases is a major challenge, intellectual as well as practical, with links to fundamental issues in the theory of emotion. In effect, a database expresses a tacit theory of the domain it deals with, and learning extracts it (selectively). To choose a database, and still more to construct one, research teams need some sense of the theoretical issues in the area. That includes knowing when ideas that intuitively seem appealing?or even compelling?are old science or ?folk psychology? (Place, 1996) rather than credible contemporary science. The paper aims to draw out the scientific issues that are relevant to decisions about databases as well as describing the actual material available.
Reviewing the issues leads to a broad sense of the ground that database research might ideally aim to cover. The best convenient summary is ?emotion in action and interaction??indicating that the challenge is not to find the emotional equivalent of pure elements, but to represent emotion as it appears woven through everyday activities.
From that we move to list the significant data sources reported in the literature. For each source, we give the information most likely to be relevant to research teams aiming to identify useful sources or to evaluate work that has been done with them.
1\. Emotion and archetypes
?Emotion? and related words are notoriously elusive (Russell & Barrett-Feldman, 1999). One of the key issues relates to a well known characteristic of words in general. When they are used in isolation, they tend to call up images that people regard as archetypal examples of the category in question (Mervis, Catlin, & Rosch, 1976). That is a useful device in its place, but it needs to be treated with proper caution. Certainly research should not slip into assuming that the empirical data needed to understand a topic equate to a collection of cases that reflect the archetypal images we associate with it.
In the domain of emotion, there are several kinds of pressure to make exactly that assumption. Understanding recent work on databases depends on recognising that it has been concerned to break their grip.
The first type of pressure to focus on archetypes comes from well-known theoretical concepts. The concept of primary emotions is particularly influential. The underlying idea, which goes back at least to Descartes (1911?1912), is that emotion is like colour. Certain emotions, like certain colours, are ?pure? examples of the phenomenon: others are ?secondary?, made by combining them in different proportions. If that theory were correct, then

the agenda for emotion research would be clear: investigate the primaries, and the secondary forms will fall into place.

The primary/secondary distinction has enduring appeal, but Scherer et al. (2004, p. 10) sum up the consensus within the emotion research community: ?This distinction is especially problematic ? and should therefore be used only with caution.?. The same warning applies to the idea that the ideal basis for training is a database stocked with examples of ?primary emotions?.

More recent is the concept of ?basic emotions?. It expresses the idea that emotion consists of qualitatively distinct types of state, each related to a biologically important scenario; and (in the best known version of the idea) each with a direct biological link to particular expressive behaviours, vocal and more particularly facial. A famous list due to Ekman's group (Ekman et al., 1987), is sometimes called the ?big six? (Cornelius, 1966) However, even Ekman no longer advocates that kind of attractively short list (Ekman, 1999). Longer alternatives are summarised by Cowie and Cornelius (2003). Correspondingly, databases composed of a few hypothetically basic states need to be treated with caution.

A second pressure to focus on archetypes is precisely that many databases do deal with a few archetypal states. The outstanding example is Ekman and Friesen, 1975a, Ekman and Friesen, 1975b collection of still photographs, which is grounded in an early version of his ?basic emotion? theory. Various other collections, still or video, have been modelled more or less explicitly on it (e.g. http://www.media.mit.edu/).

These databases are intuitively appealing, but, as noted above, the theory behind them is dated. And even in its strongest form, Ekman's account included qualifications that set limits to the likely uses of his database and others like it. It specified that in everday life, the expressive gestures rooted in biology were overlaid, modified, and mimicked according to socially defined ?display rules? (Ekman & Friesen, 1969). Hence, the ?pure? examples of Ekman's database were explicitly not the patterns that we should expect to see when humans actually interact, with each other or with machines; and training systems to deal with those interactions would need data that showed the effects of display rules.

A third type of pressure favours a different kind of archetype. There is a tradition of believing that there is a peculiarly direct correspondence between emotion terms and physiological measurements. One version of the idea was introduced by William James (1884), who suggested that the basis of emotion was a set of bodily responses to a significant type of stimulus?change in heart rate, breathing, skin state, and so on. More recently there has been intense interest in relationships between emotion and brain states, such as activation of the amygdala (LeDoux, 1987) and change in the balance of electrical activity between the two cerebral hemispheres (Davidson, 1999). It is a natural step from those ideas to infer that some set of physical measurements represents the ?ground truth? behind the confusing surface of emotional behaviour, and it is the business of a database to ensure that those measures are available.

The Jamesian idea seems to contain some truth (Cornelius, 1966), but it is far from the whole truth. There is a large literature on the extent to which emotions can be discriminated by bodily signs (e.g. Cacioppo, Klein, Berntson, & Hatfield, 1993). Some discriminations certainly can be made, but they are quite coarse, and they almost always depend on very carefully controlled experimental conditions. The reason is that most of the variables which are affected by emotion are also affected, as much or more, by commonplace activities such as moving, speaking, or even thinking. As a result, their value as discriminators diminishes rapidly as one moves towards free situations where people are likely to be moving, speaking, or thinking.

On one level, the jury is still out on the measurement of brain states. The reason is that many of the key measurement techniques require activity to be severely restricted, making it impossible to study whether normal activity would interfere with the detection of emotion or change the patterns associated with it. However, there are good reasons to doubt that reductionist approaches based on brain states will fare better than any others.

Saying that a person is in a particular emotional state carries implications about the way he or she attends to and perceives things, people, and events, which may be present, past, anticipated, or imagined; about his or her feelings; about his or her capacity to weigh options dispassionately; about the actions he or she is disposed or likely to take; about the moral status of his or her thoughts and actions; and at least in some cases about the reality of his or her situation (e.g. Sabini and Silver, 2005, Cowie, 2005).

Such a statement strongly suggests that ascriptions of emotionality implicate many systems, at least some of them of rather a high order, and not all of them contained within the person to whom the emotion is attributed; and that they refer to subtle features of the ways in which systems operate rather than simply whether they are active or not. It is instructive that systems such as attention, which are part of the emotional pattern, are themselves based on highly complex and distributed neurophysiological substructures (Schneider, 1995). In such a situation, research cannot realistically expect brain imaging to identify tractable correlates of emotion terms as they are used in natural language. It may find correlates of something simpler that plays a role in emotionality, but that is another matter.

A common feature of the ideas covered in this section is that they suggest a two-stage approach to understanding emotion. In the first stage, a set of relatively simple core correspondences is established. It is assumed?perhaps implicitly?that analysing the core correspondences will pave the way to deal with the complexities of emotion as it appears in everyday action and interaction, and the attribution of everyday emotion words. It is a short step to the assumption that the business of databases is to supply data that allow these core correspondences to be recovered.

It would be misleading to say that two-stage approaches are discredited. Rather, it is no longer obvious that attempting to bypass the apparent complexity of the emotion domain is the best way to proceed. Emotion-related phenomena have refused to fall into place around a succession of plausible simplifying assumptions. It may be that current assumptions will pay off with additional research, or that assumptions yet to be articulated will pay off. But the alternative has to be taken seriously?that the phenomena are essentially complex, and will resist simplification indefinitely.

The idea that the phenomena are essentially complex suggests two conclusions. First, it is the responsibility of research on databases to supply data that fully reflect the complexity of the phenomena. Second, techniques like neural nets, which are inherently suited to addressing complex relationships, are likely to have a key place in describing the patterns that exist. The two define complementary parts of an emerging strategy.

2\. The scope of databases

It is not self-evident what range of material emotion-related databases should contain, particularly if the decision is not pre-empted by archetypes. This section outlines major themes in recent thinking about the subject.

2.1. Types of emotionality

At least two broad types of phenomenon are described as emotional in common parlance. We have described them as ?episodic? and ?pervasive? emotion (Cowie & Schroeder, 2005).

The term ?episodic emotions? describes states where emotion dominates people's awareness, usually for a relatively short time, by affecting their perceptions, feelings, and dispositions to act. The emotion may not determine the action that the person takes, but it requires effort to prevent it from doing so. Clear-cut episodic emotional states are generally brief, and relatively rare.

In contrast, ?pervasive emotion? refers to something that is an integral part of most mental states, including states where rationality seems subjectively to be in control. It is integral in the sense that emotional colouring is part and parcel of the way people experience situations that they are in, or remembering, or anticipating. It may be associated with the situation as a whole, or with individual elements of it, or with possible courses of action. However, the colouring is part of the background unless something triggers a kind of phase shift and propels people into a state of episodic emotion.

It is a major issue whether the business of databases is to assemble instances of episodic emotion, or to take a broader remit and represent pervasive emotion. Database users have the right to look for sources that consist of emotion in the strong sense formulated by Scherer. But if they do, they should be aware of the reasons why recent collections have moved away from that pattern. These are bound up with prevalence and application.

The key question about prevalence is whether episodic emotion is intermittent, but relatively common, or distinctly rare. Naturalistic studies indicate that there are strikingly few clear instances of episodic emotion even in material that would usually be considered emotional in a global sense. For our Belfast naturalistic database, we identified TV talk shows which previews suggested were rich in genuine emotion (Douglas-Cowie, Campbell, & Roach, 2003). Analysis showed that clear-cut emotional episodes were unexpectedly rare even in that material (Cowie & Cornelius, 2003). Instead, the sense of emotionality derived mainly from emotional colouring?fluctuating, often apparently conflicting, concerned with situations that were remembered or anticipated as much as with present surroundings, and with rare exceptions thoroughly subordinated to the demands of civil discourse. Other analyses of naturalistic data present similar pictures, though in different language (Ang et al., 2002, Batliner et al., 2004, Batliner et al., 2004; Campbell, 2002a, Campbell, 2002b, Campbell, 2004; Craggs and Wood, 2004, Greasley et al., 2000, Roach, 2000, Tsukahara and Ward, 2001, Ward, 2003).

Application is discussed later, but it can be said at this stage that one might expect to find much more use for systems that respond to a very prevalent feature of human communication than for systems that respond to one that is much rarer than people think.

A purely verbal point sometimes complicates the issue. Some theorists prefer to reserve the term ?emotion? for episodic phenomena. For instance, Scherer proposes to define emotions as ?episodes of massive, synchronized recruitment of mental and somatic resources to adapt or cope with a stimulus event subjectively appraised as being highly pertinent to the needs, goals and values of the individual? (Scherer et al., 2004, p. 10). The definition captures very precisely what we have called episodic emotion, but it implies that another name should be found for what we have called pervasive emotion. Ekman (1999) takes a related position.

We prefer the episodic/pervasive terminology because it allows experts and non-experts to be roughly agreed about how much of everyday life might be called emotional, whereas the Scherer/Ekman usage implies that experts will focus on a much smaller part of life (and may actively encourage them to focus on a smaller part). Which usage wins depends on the politics of language rather than science.

What is a matter for science and technology is that very large parts of everyday life are emotionally coloured, and yet are not what we call episodic emotion and Scherer calls emotion. It is important to have databases that represent them, whatever they are called. We will continue to use the term pervasive emotion.

2.2. Contexts and modalities of emotion

The context of emotion becomes a major issue as soon as one considers pervasive emotion. It is plausible to assume that clear-cut bursts of episodic emotion will look and sound somewhat similar in most contexts?whatever the person was doing will be interrupted and replaced by emotion-specific behaviour. Thinking of pervasive emotion raises different issues. It will be integrated into the way ongoing actions and interactions are executed. The results will depend on the nature of those ongoing actions and interactions, and will be very different in different contexts?watching a football match, driving a car, buying groceries, standing waiting for a friend, chairing a meeting, writing a paper.

Database research has not generally paid systematic attention to the questions about context, but the issue is gaining recognition. In particular, the ongoing HUMAINE project has set out a programme involving two types of context-sensitivity?labelling for context in naturalistic data, and carrying out induction studies chosen to represent a range of strategically different contexts.

Context is inescapably linked to modality. Emotion is strongly multimodal in the sense that signs may appear in various different channels. However, not all types of sign tend to be available together, because context will affect the signs that are relevant or accessible. One of the most challenging tasks for database work is to capture records that clarify when and how which modalities come into play.

Facial signs are potentially available in a wide range of contexts, but they are also subject to a wide range of influences. These include culture-specific display rules and interactions with speech, involving both the lips and the eyebrows. They are reduced in the absence of social context, and can be feigned or dramatised, consciously or unconsciously. Databases provide systematic coverage of a small part of this range.

Gesture-related signs are not generally available when the hands are occupied by a task, and they are constrained when the arms are constricted (e.g. by a narrow space) or being used (e.g. to lean on or to lift something). They are also socially constrained (as every good child knows, it is rude to point), and the violation of social constraints may be a sign of emotion in itself. Databases showing emotion-related gestures reflect a very small part of that situation.

Similar comments apply to global posture (?body language?). The signs that can occur in mobile, standing, seated, and reclining contexts are very different, as are postures shaped by engagement with an object and/or other people. The emotional significance of a posture may also depend on its orientation to another person or congruence with his or her posture.

The classical vocal signs of emotion involve prosody, voice quality, and timing. Their context is subtle. They are clearly linked to interactive contexts, but there seems to be antagonism between overt emotionality and well-formed speech communication, as reflected in phrases like ?speechless with emotion? (Lee & Narayanan, 2005). That suggests the natural setting for

vocal signs of emotion is in a transition zone between controlled speech and inarticulate vocalisation. Intrusion of non-speech sounds (Schroeder, 2003) makes sense as part of a transitional state like that.
Capturing that kind of setting is an intriguing challenge in itself. Like the face, the voice can easily be dramatised. That presumably involves mimicking some of the signs that occur in the transition zone. In any case, dramatised speech is an integral part of the way people communicate their emotional states, and well worth representing in a database.
Traditional research on speech and emotion focused on vocal signs and deliberately excluded verbal variation by requiring subjects to express various emotions using the same word of phrase. In fact, though, verbal content naturally covaries with vocal signs, in several respects. Acted databases almost never capture that aspect of emotionality, and relatively little has been done to draw it out of the naturalistic databases that do reflect it.
Some attention has been paid to cues at the level of discourse (broadly speaking). Lee and Narayanan (2003) classified utterances into rejections, repeats, rephrases, ask-start overs, and others. The discourse measures were less informative than acoustic measures, but nevertheless they did relate systematically to emotional state.
Beyond that, emotion affects the classical linguistic variables of syntax and vocabulary. A simple demonstration comes from the paper by Athanaselis et al. in this special issue, which shows that the verbal content of emotional speech can be recognised much better if the language model is based on a corpus biased towards utterances with some signs of emotional content. A large literature dating back several decades provides a rich source of ideas about the way emotion might influence choice of vocabulary?immediacy of expression, concreteness of terms, use of expletives, and so on (Berger & Bradac, 1982). Intuitively, it seems obvious that emotion will usually be intimately linked to the semantic content of people's speech, and that the most powerful cues to emotionality will usually come from content. Rich datatabases are the key to pursuing that idea systematically, and they have only begun to reach the point where that is possible.
Effects involving attention are similarly under-represented in mainstream databases, with the partial exception of extreme inattention in boredom (Cowie, McGuiggan, McMahon, & Douglas-Cowie, 2003) and drowsiness (Hadfield & Marks, 2000). That is curious in the light of contemporary theory, which understands emotion in terms of appraisal processes concerned with forming selective, valenced representations of the environment (Scherer, 1999). Eye movement, head movement and posture all offer potential sources of information about the focus and manner of a person's attention, and hence of the appraisals governing his or her emotional state. Again, rich datatabases are the key to pursuing the idea systematically.
Action ranks with attention as a variable that theory suggests should be pivotal. Aristotle understood emotion in terms of action tendencies (such as the impulse to revenge in anger), and the idea has remained an integral part of modern theory since Frijda (1986) re-established it. Very few databases are designed to record action patterns that might be symptomatic of emotion. An exception is work on driving (McMahon, Cowie, Kasderidis, Taylor, & Kollias, 2003), which offers a context for measuring behaviours associated with risk-taking and aggression.
Sensitivity to context is particularly crucial in the case of physiological measurements, peripheral and central. The optimal setting for discriminating emotional states on the basis of physiological measurements is a sedentary

individual who is inducing pure and intense emotional states by acts of imagination (Picard, Vyzas, & Healer, 2001) or being exposed passively to powerful emotion-inducing images. Such experiments clarify points of theory by showing that connections exist between emotion and certain somatic and neural systems. It is a different issue to establish how they feature and may be used in the more complex contexts that are relevant to most applications. Failure to engage with that issue underlay the running scandal of ?lie detection? using the polygraph (Lykken, 1984), and it is a mistake that should not be repeated.

There are contexts where it is reasonable to think that physiological measurement could be useful. Driving is a prime example, because simple physical effort is limited, and relevant environmental variables (temperature, humidity, etc.) can be monitored. There is some data on physiological correlates of emotion-related states in that context, and more work is under way (McMahon et al., 2003).

Forming an overview of contexts, and the ways in which emotion might be expressed within them, is a first step towards developing an analysis of pervasive emotion. The next step is to assemble records that establish what actually does happen in contexts that have been identified as significant. That is considered in the next section.

2.3. Obtaining samples of emotion

Most technological research on emotion continues to be based on recordings of actors, skilled or unskilled. The fundamental reason is linked to the previous sections. It is assumed that the proper target is episodic emotion; episodic emotion is much rarer than people tend to assume, and difficult to elicit; therefore, research teams turn to actors.

Blanket rejection of acted data is not a sensible option, for various reasons. However, there are good reasons to move away from uncritical reliance on acted data, and that has been a major theme in recent database work.

Several kinds of standard can be used to evaluate acted data. The one which is most often observed is that people can recognise the emotion being portrayed (Banse & Scherer, 1996; Ekman and Friesen, 1975a, Ekman and Friesen, 1975b; Laukka, 2004). It is not a criterion that should be over-emphasised. On one hand, it would tend to rule out a good deal of genuine emotional material, because inter-observer agreement on it is often not particularly high. On the other, it does not rule out material which is identifiable, but not at all like natural emotion?in short, ?ham? acting.

We have explored a technique which addresses the second problem. As a recording is played (audio, visual or audio?visual), observers use a sliding scale to indicate whether they think the material is spontaneously expressed emotion or acting, and how confident they are in their judgment. We compared ten extracts from the Belfast naturalistic database with renditions of the same material by actors in the Belfast structured database (Douglas-Cowie et al., 2003). On the basis of pilot studies, we chose the most convincing rendition of each extract, and one which was about average. Two measures were of interest. The first was the percentage of trials which eventually favoured the right choice (acted or not). 68% concluded that the less convincing renditions were acted, and 67% that more convincing renditions were. In contrast 79% concluded that the ones which were actually spontaneous were not acted. The second measure was how quickly the difference between responses to the corresponding real and acted extracts reached statistical significance.

For both convincing and unconvincing renditions, the difference reached significance within the first three seconds. Both measures indicate that even well acted material differs very noticeably from spontaneous emotionality.

To the best of our knowledge, no databases of acted material currently include measures of naturalness. However, it is easy to carry out the kind of study we have described above, and it would make sense to carry out checks of that sort before committing to using an acted database.

Even if acted material is far from natural, it could conceivably be well suited for training, if it presented the features that are relevant to everyday behaviour in sharper focus. There seem to be no direct tests of the idea, but the evidence that exists is not encouraging. Batliner, Fischer, Huber, Spilker, & Nöth (2003) have shown that the vocal signs of emotionality used by an actor simulating a particular human?machine interaction were different from, and simpler than, the signs given by people genuinely engaged in it. Kollias (2005) have used the Ekman and Medialab databases of acted facial expressions as a basis for generating less extreme facial profiles, and are applying the approach to recognition in induced emotion data. Results to date suggest that the approach has some success with individuals who express themselves in a dramatic way, but not with individuals behaving more naturally.

Note that for some applications, acted material is actually the optimal source. For example, an artificial newsreader should probably simulate the conventionalised emotional signs offered by human newsreaders rather than showing true grief at disasters or euphoria at triumphs.

Naturalistic data seems an ideal alternative to acted sources, but the reality is not straightforward. Problems of copyright and privacy are well recognised (Douglas-Cowie et al., 2003). More fundamental, but less often noted, is the fact that key tools needed to deal with naturalistic data are underdeveloped.

First, recordings of the kind seen in ?reality? TV could be obtained in reasonable quantity, but technology lacks the robust low-level analysis needed to handle them. Humans can recognise expressions of emotion from faces that are at an angle or at a distance, but the 3D modelling techniques needed to do that automatically have barely been explored (Kollias et al., 2004).

Similarly, speech analysis lacks the tools to deal with varying distances from microphones, reverberation, and noise. The traditional strategy is to expect recordings to cater to the engineering limitations, but the usual methods of doing that conflict with free emotional expression, and the long term solution is clearly to address the engineering limitations.

Second, we lack the conceptual tools to deal with the diversity of natural emotional behaviours. The traditional strategy is to look for elements common to all or most instances of an emotion (Juslin & Laukka, 2002). From the perspective of the previous sections, that filters out a huge amount of evidence, because so much evidence lies in the way emotion modulates activities that carry their own constraints. If the expression of emotion has common elements in the contexts considered above, they are at a deeper level, involving attention, planning, risk-taking, and so on. We do not have the theory at that level that would allow us to draw out the common threads in a genuinely naturalistic collection. Sophisticated learning techniques may well be relevant to developing it.

Recordings from real human?machine interactions avoid both of these problems to some extent, because they involve built-in constraints and a kind of uniformity. If short-term applications are the target, then directly relevant databases are clearly the appropriate resource?not least because of the context-dependence that has been emphasised. However, focussing on short term applications may be counterproductive in the long term, because it postpones confronting the major issues, and finding relationships which are dependent on the structure of a particular task may not clarify the general picture. To illustrate, we studied bored subjects carrying out a naming task (Cowie et al., 2003), and found reduced pausing (because speakers ran descriptions of successive items together). However, we have since found that a small change in the structure of the task results in exactly the opposite trend, increased pausing (Cowie & Schroeder, 2005). It seems that the underlying effect is not on pausing, but on attention; and the overt signs of the attentional effect depends on task details. If the field were driven by data from many very specific tasks, it would risk being overwhelmed by that kind of highly context-specific contingency. Allied to that is the risk demonstrated by Yacoub, Simske, Lin, & Burns (2003). A recogniser trained on a data containing a narrow range of emotions is likely to respond quite inappropriately if it encounters an emotion outside the training range. That raises ethical as well as practical problems (Goldie et al., 2004).

In spite of the problems, major efforts have been invested in collecting naturalistic data (Batliner et al., 2004, Batliner et al., 2004; Douglas-Cowie et al., 2003 DeVilliers et al. this issue). Key sources are listed in the Appendix to this paper.

Between acting and pure naturalism lie various emotion induction techniques. It is not difficult to create situations which induce a specific emotion such as irritation (Bachorowski, 1999) or boredom (Cowie et al., 2003). It is more challenging to induce a range of emotions in comparable conditions, but there are various established methods?verbal methods such as the Velten technique (Velten, 1968); listening to emotive music (Clark, 1983, Kenealy, 1988); looking at emotive pictures (Center for the Study of Emotion and Attention (CSEA-NIMH), 1999, Lang et al., 1999) or films (Gross & Levenson, 1995); and playing specially designed games (Scherer, Johnstone, & Bänziger, 1998). Cowie and Cornelius (2003) give a more detailed review of classical techniques and their status.

The traditional test of induction techniques is whether they evoke a state that genuinely qualifies as emotion. That is a reasonable criterion if one is looking for an inner core of phenomena that (hypothetically) accompany all emotions. However, the perspective developed here points to a different test, which is ability to induce emotion that carries through into action and interaction. Work on that problem is much more recent.

Scherer et al. (1998) carried out pioneering work with a computer game containing incidents designed to induce specific appraisals. Their effect was shown not only in transient facial responses, but also in a reading task which was part of the game. That taps strictly vocal signs of emotion in speech, though not effects at the level of phrasing, word selection, etc.

A technique developed by Auberge, Audibert, and Rilliard (2003) evokes freer expression. Students with an interest in language used a voice-driven language teaching program, and were given feedback indicating that their performance was outstandingly good or disturbingly bad. Their ongoing reactions were reflected in facial expression and vocal behaviour which included specific items to be pronounced and unscripted comments.

Our SAL technique (Cox 2004) is modelled on ?chatbot? systems. The user talks to characters whose conversation is defined by ?scripts? consisting of emotionally loaded stock phrases. The system contains different characters, each of which has a distinctive emotional outlook (in the current version, angry, happy, sad, and pragmatic), and each tries to draw the user towards its own outlook. Users do not automatically engage with SAL, but when they do, it allows them to express a range of emotions through a relatively wide range of channels, facial, vocal, discourse, and content.

One of the key difficulties for database research is that induction remains an uncertain art, seriously complicated by ethical issues. It is a telling sign that we are currently exploring ways to induce a range of genuinely emotional states that are persistent enough to affect driving behaviour (in a simulator). Transient emotion can certainly be induced by events of no enduring significance in subjects' lives, but it seems to dissipate very quickly when they are presented with a task or a challenge. The same does not hold when the inducing events or situations are part of the fabric of people's lives, but not many laboratories would manipulate those events to gather data.

2.4. Emotion and applications

Application is clearly an issue in the development of databases, but it is best addressed after the scope of the field has been indicated.

The obvious short-term application of neural nets in the domain of emotion is ?trouble shooting??detecting emotional states that may be troublesome in callers using automatic exchanges or websites, pilots, drivers, and so on (Cowie & Schroeder, 2005). Another possibility is ?affective selection??using emotion-related signs as evidence that a person might want to hear certain music, preserve certain experiences, or be shown certain types of option (Cowie & Schroeder, 2005). Neither the databases nor the computing techniques needed for these applications is particularly exciting, at least at first sight.

The longer term goal, though, is ?really natural language processing??communication between humans and machines that follows human norms in emotional terms as well as others. That means both registering and providing the kind of emotional colouring that is the norm in human interaction. In applications such as artificial companions and tutors, there is a clear incentive to do these things better than many humans do. It is to achieve targets like these that research needs databases of great scope and subtlety.

Correspondingly sophisticated computing is needed to exploit the data. The requirement in these contexts is not simply to label isolated patterns. It seems likely that sequence and timing, often crossing different modalities, are critical in the detection of emotion-related effects. Taylor's ANNA (this issue) illustrates the reasonable supposition that mechanisms modelled on the human brain will prove the best way to handle that kind of problem. The same point applies to the inverse problem, which is to generate appropriately sequenced and timed arrays of signals.

3\. Descriptive issues in databases

Emotion databases include not only recordings in various modalities, but also descriptions of the emotions involved and the signs relevant to them. This section summarises the main options in these areas.

3.1. Describing quality

Several preliminary issues should clearly be documented in a satisfying database, though they rarely are. The key examples are

* ?

Is any emotion evident? If, for instance, one is exploring a large database that involves very little emotion, one of the most useful elements a database could have is a trace showing where there is emotionality of any sort is among.

* ?

Is emotion being masked? This is linked to the Ekman concept of display rules (see Section 1).

* ?

Is the emotionality that is displayed spontaneous or contrived? This is not a simple matter of ?good? and ?bad? samples. For instance, an element of dramatisation is part of everyday communication, and sometimes the target is to achieve acceptable formal representations of emotion (see Section 2.3).

3.2. Describing emotional content

Three main types of scheme could reasonably be used to describe the emotional content of a database: categorical; continuous; and appraisal-based.

Categorical schemes are rooted in everyday language. They involve assigning terms like angry, ashamed, jealous, and so on. There is no real doubt that terms like these have an enduring place in emotion-oriented technology, because they are the means of describing emotion that people find most natural. The difficulty is how to deal with the range of descriptors that everyday language offers and the kind of emotion that appears in relatively natural settings.

Numerous teams have tried to label relatively naturalistic material using schemes based on established psychological lists of emotions. Two examples illustrate the outcomes. The Leeds-Reading database (Douglas-Cowie et al., 2003, Roach, 2000) produced fine labellings in which a single utterance might be described as moving from hate to vengeful anger. The result is a very low incidence rate of similarly labelled utterances, leading to intractable statistical problems. Craggs and Wood (2004) used a list derived from Ortony and Turner (1990)?courage, dejection, sadness, disgust, aversion, shame, anger, surprise, guilt, wonder, hate, love, happiness, desire, contempt, fear. The result was considered unacceptable because of very low inter-rater agreement.

Part of the problem is that traditional psychological lists are oriented to archetypal emotions, and those are not the states that appears in most naturalistic data. Two main types of response have developed within a categorical framework. One is to group the original categories into ?cover classes?. For example, Batliner et al used positive, pronounced, weak negative, and strong negative. Another is to develop lists oriented towards describing everyday emotionality rather than states which are considered important theoretically, but may rarely occur in a clear-cut form.

An early list of that kind (Cowie et al., 1999) was used to label the Belfast naturalistic database. More recently it has become clear that there are problems with lists that are derived without reference to context. As a result, Sweeney (2005) asked participants to rate how useful 75 emotion-related words would be to describe a recent day in their lives. For one group, the day was to be typical; for another, a good day; and for the third, a bad day. Table 1 shows the words that were most highly rated by each group (in order). The list for a typical day is similar to our older list, and to the list for a good day; but neither includes the terms needed to describe a negative day. The strategy of preselecting word lists that suit the context is a useful extension of categorical approaches, and is being applied, for instance, to labelling emotion in meetings.

Table 1. The emotion-related words rated most useful for describing three types of day (from most to least useful)

Typical day	Good day	Bad day
Amusement	Pleasure	Stress
Friendliness	Amusement	Sadness
Pleasure	Delight	Anxiety
Affection	Excitement	Frustration
Love	Satisfaction	Hurt
Politeness	Affection	Annoyance
Contentment	Content	Despair

Interest| Friendliness| Tension
Trust| Happiness| Worry
Excitement| Love| Disappointment
Calm| Interest| Powerlessness
Satisfaction| Hopefulness| Shock
A more radical alternative is to use the long-established psychological concept of emotion dimensions (Cowie & Cornelius, 2003). Statistically, most of the information contained in verbal labellings can be expressed in terms of two dimensions, activation (whether the person's emotional state makes him or her more or less likely than average to take action), and evaluation (how positive or negative he or she feels). The FEELtrace tool allows raters to track a target person's perceived emotional state in real time as they watch and/or listen to a recording of the person in question. It gives a continuously valued trace. Craggs and Wood (2004) have described a categorical rating technique based on the same underlying theory, though they substitute strength of emotion for activation. The result is actually quite close to Batliner et al's cover categories.
Dimensional description appears to be increasingly accepted as an established tool. It has to be emphasised that it is by definition an approximate tool. Two-dimensional versions, for instance, do not distinguish between fear and anger (that requires a third dimension involving power or control); and concepts like politeness and interest which figure in Table 1 lie quite outside the scheme. Its coarseness both makes it useful and limits its use.
A finer theoretical tool comes from appraisal theory, which is probably the most influential approach to emotion within psychology at present (Scherer, 1999). Appraisal theory offers a descriptive framework for emotion based on the way the person involved experiences the events, things, or people at the focus of the emotional state. This is described involving five broad headings:
 * ?
Novelty (involving suddenness, familiarity and predictability as distinct subcategories)
 * ?
Intrinsic pleasantness
 * ?
Goal significance (including the aspect of life to which the events are relevant, from the body to relationships to social order; the likelihood that there will be a significant outcome; how the events relate to expectation, and to the agent's goals; and how urgent they are)
 * ?
Coping potential (covering how much control anyone could exercise over the event, how much power the agent in particular has to control them, and how easily the agent can adapt by adjusting his or her own goals)
 * ?
Compatibility with standards (covering how far the events conform to or violate moral or ethical standards, held either by the agent personally or by society as a whole)
 * Work is under way on translating this scheme into a practical tool. One of the main motivations is that the individual items may be linked to component gestures in the expression of emotion (such as eyebrow raising, mouth movement, etc.). That opens the way to deal both with the sequencing of gestures, and with situations where some, but not all of the appraisal elements associated with full-blown emotion are present.
 * These descriptive schemes are far removed from the pattern that someone entering the field might expect to find in a database. Understanding how the descriptions involved relate to signs of emotion is also a much more interesting task.
3.3. Analyses of emotion-related signals and signs

Not many teams working on neural nets will be able to use data in the form of direct records in the audio, visual, or physiological domain. Most will need material to be analysed on at least one of two other levels, feature extraction and sign description. Feature extraction refers to the output of signal processing operations (perhaps corrected by hand). For example, existing routines can extract positions of key joints from video records. Sign description refers to the ways in which humans spontaneously interpret incoming evidence, usually involving intentionality?for instance, pointing at a car. Automatic transition from extracted features to signs is very difficult.

There are various levels at which neural network research could engage with sequences of signals, features, and signs, given appropriate data. There is still a need for general, robust solutions to many of the basic problems in feature extraction. The mapping of features to signs is an extremely challenging problem, as is the mapping of signs to descriptions of emotion. Neural nets may well also be well suited to some aspects of the inverse problem, synthesis of sign sequences that are appropriate to express a sequence of emotional states.

There is currently rapid evolution in the provision of data relevant to these tasks. The Erlangen group is in the process of making available the AIBO database, with hand corrected transcriptions of verbal content and pitch. SAL data collected under the ERMIS project includes FAPS (Balomenos et al., 2004), basic speech features (pitch and intensity contours and pause boundaries) and higher order features calculated by the ASSESS system. ASSESS features are also available for the Belfast naturalistic database. The EmoTV Corpus contains rich descriptions of signs. Programs under way use motion capture to provide machine-readable descriptions of body movements relevant to emotion (see e.g. Pollick, Hill, Calder, & Paterson 2003).

Records are often not clear about the level of analysis available, and the situation is often evolving. Hence the Appendix does not try to detail the level of analysis associated with data sources. Some additional information is available via the listing of databases on the HUMAINE website (http://emotion-research.net).

3.4. Describing context

Despite the importance of context, there are no generally used schemes for describing it. A recent HUMAINE report included a proposal that at least the following issues should be specified

* ?
Agent characteristics (age, gender, race)
* ?
Recording context (intrusiveness, formality, etc.)
* ?
Intended audience (kin, colleagues, public)
* ?
Overall communicative goal (to claim, to sway, to share a feeling, etc.)
* ?
Social setting (none, passive other, interactant, group)
* ?
Spatial focus (physical focus, imagined focus, none)
* ?
Physical constraint (unrestricted, posture constrained, hands constrained)
* ?
Social constraint (pressure to expressiveness, neutral, pressure to formality)

It is proposed to refine this scheme through work with the HUMAINE databases

as they develop. Parts are already implemented in the EmoTV Corpus described in the Appendix.
4\. Databases
The Appendix lists the databases that seem most likely to be of interest to the neural net community (and to research on machine learning in general). Items are included for three main reasons.
First and foremost is the use of data that is naturalistic or induced in credible ways. These are covered as fully as possible.
The second is multimodality. Multimodal databases are still rare, but the picture has changed considerably since a relatively recent review (Douglas-Cowie et al., 2003).
The third is influence. Databases which are neither multimodal nor naturalistic have nevertheless been influential for a variety of reasons, one of which may be that they are convenient to access and use. It would be remiss not to include influential sources in a systematic review.
5\. Conclusion
Ekman and Friesen's database of emotional faces remains for many people the example of what they expect an emotion database to be. The databases listed in the first part of the Appendix are very different. It seems likely that the future trend will be towards databases more like the latter than the former.
One of the main aims of this paper has been to convey the background against which that move has occurred, so that people who are inclined to expect something like the Ekman and Friesen resource understand why they are being pointed to something so different.
It is also the case that assembling databases has not traditionally been considered a high-profile or intellectually challenging area. Good quality recording and large balanced samples tend to be thought of as the basic requirements, with the human side assumed to be relatively straightforward.
A little thought shows that in the domain of emotion, that cannot be the case. The human race expends a huge proportion of its resources trying (with mixed success) to direct people out of some emotional states and into others. If it were easy to achieve the shifts, there would be no need for whole industries and cities to exist.
As a result, capturing a faithful, detailed record of human emotion as it appears in real action and interaction is a massively challenging task.
Nevertheless, the payoff is also large. At root, it is enlisting computers to co-operate in the old task of directing people away from some emotional states and into others. The lure of technologies capable of doing that is enough to keep the enterprise going in spite of the difficulties.
Acknowledgements
Preparation of this paper was supported by EC projects ERMIS (IST-2000-29319) and HUMAINE (IST-2003-507422).
Appendix.
List of key databases for computational research on emotion. The abbreviations in the ?modality? column are as follows: A=audio; V=video; G=gesture; P=physiology. Column 1 cites published sources where they are available: the corresponding references are in the general list at the end of the paper.
(Table A1)
Table A1.
Name	Modality	Elicitation method	Size	Language	Type of emotional description	Emotional content	Electronic contact
Belfast naturalistic database (Douglas-Cowie et al., 2003)	AV	Natural: 10?60 s long ?clips? taken from television chat shows, current affairs programmes

and interviews conducted by research team| 239 clips; 209 from TV recordings, 30 interview recordings. 125 subjects; 31 M, 94 F| English| Dimensional labelling/Categorical labelling| Wide range|
Geneva airport lost luggage study (Scherer and Ceschi, 1997, Scherer and Ceschi, 2000)| AV| Natural: unobtrusive videotaping of passengers at Geneva airport lost luggage counter followed up by interviews with passengers| 112 subjects| not known| Rating on emotion category scales| Anger, good humour, indifference, stress, sadness|
EmoTV corpus| AV| Natural: 4?43 s long ?clips? from French television channels| 51 clips from French TV channels; 48 different individuals;| French| Categorical labelling| Wide range| http://emotion-research.net/ws/wp5/databasesLQ.ppt
SAL (Sensitive artificial listener) database| AV| Induced: subjects talk to artificial listener and emotional states are changed by interaction with different personalities of the listener| Approx. 10 h; study 1, 20 subjects; study 2, 4 subjects x2 20 min sessions each| English| Dimensional labelling| Wide range of emotions/emotion related states but not very intense| http://emotion-research.net/ws/summerschool1/SALAS.ppt, http://emotion-research.net/ws/wp5/ellsal.ppt
SMARTKOM| AVG| Human machine in WOZ scenario: solving tasks with system| 224 speakers; 4/5 minute sessions| English, German| Categorical labelling holistic and for each channel in turn Intensity also labelled| Joy, anger, gratification, irritation, helplessness, pondering, reflecting, surprise, neutral| http://www.phonetik.uni-muenchen.de/Bas/BasMultiModaleng.html#SmartKom; http://www.phonetik.uni-muenchen.de/Publications/Steininger-02-LREC.pdf
DaFEx Database of Kinetic Facial Expressions (Batocchi & Pianesi, 2004)| AV| Acted: Instructions given to actors. Use of ?scenarios?, one for each emotion| 1008 clips (4?27 s each) coming from recordings with 8 professional actors (4 male, 4 female) Total recording time: 2 h 50 min 13 s| Italian| Categorical emotional labelling| Ekman 6 emotions plus neutral expression, 3 intensity levels, 2 acting conditions: ?Utterance? and ?Non?Utterance?.| {battocchi,pianesi@itc.it
eWiz database (Auberge et al. 2003)| AVP| Induced using a WoZ scenario simulating a language teaching exercise| 17 subjects (10 general, 7 actors) 3400 single words plus comments 5hrs speech, 15 hours of non-verbal| French| Labels chosen by subjects (category and dimensional)| Concentration, positive, negative, stressed, satisfaction, boredom, surprise, disappointment, anxious, confused| http://www.icp.inpg.fr/EMOTION
Speech motion face database| AV| Acted| Three different expression variations for each emotion and 1326 utterances: frustrated 59, happy 85, neutral 86, sad 100, angry 112| English| Pre-defined| frustrated, happy, neutral, sad, angry| http://www.cs.ucla.edu/~abingcao/Documents/Cao_TOG_small.pdf
Polzin and Waibel (Polzin & Waibel, 2000)| AV| Acted: sentence length segments taken from acted movies| Unspecified no of speakers. Segment numbers 1586 angry, 1076 sad, 2991 neutral| English| Choice from a given set of emotion tags| Anger, sadness, neutrality (other emotions, but in insufficient numbers to be used)|
Banse and Scherer (Banse & Scherer, 1996)| AV| Acted: actors were given scripted eliciting scenarios for each emotion, then asked to act out the scenario. (visual info used to verify listener judgements of emotion)| 12 subjects (6M, 6F)| German| Expert (peer) rating of authenticity; rating from 14 emotional categories| Anger (hot), anger (cold), anxiety, boredom, contempt, disgust, elation, fear (panic), happiness, interest, pride, sadness,

shame|
AIBO ROBOT DATABASE(Erlangen database) (Batliner et al., 2004a)| AV/A| Task directions to robot| 51 German children, 51393 words| German| Categorical labelling of emotion by several labellers| Joyful, surprised, emphatic, helpless, touchy (irritated), angry, motherese, bored, reprimanding, neutral| http://pfstar.itc.it/public/publications/lrec04.pdf
Amir et al. (Amir, Ron & Laor, 2000)| AP| Induced: subjects asked to recall personal experiences involving each of the emotional states| 140 subjects; 60 Hebrew speakers, 1 Russian speaker| Hebrew, Russian| Categorical labelling| Anger, disgust, fear, joy, neutrality, sadness| http://www.qub.ac.uk/en/isca/proceedings/pdfs/amir.pdf
Chung (Chung, 2000)| AV| Natural: television interviews in which speakers talk on a range of topics including sad and joyful moments in their lives| 77 subjects; 61 Korean speakers, 6 Americans| English, Korean| Dimensional labelling/ Categorical labelling| Joy, neutrality, sadness (distress)|
ORESTEIA database Balomenos, Cowie, Malinescos, McMahon & Kasderidis (2003)| P| Induced: subjects encounter various problems while driving (deliberately positioned obstructions, dangers, annoyances ?on the road?| 29 subjects, 90 min sessions per subject| English| Performance in tasks used as index of subject state| Stress, irritation, shock|
Kim, Bang and Kim| P| Induced?multimodal approach to evoke specific targeted emotional statuses| Databse 1=125 subjects (children 5?8 yrs old), Database 2=50 subjects (children 5-6yrs old)| n/a| Self report by subjects using verbal rating| sadness, anger, stress, surprise| http://www.iee.org/Publish/Journals/ProfJourn/MBEC/20043886.pdf
The HUT Facial Expression Database| AV| Acted: static pictures and short video sequences| 2 actors| n/a| Pre-defined| Anger, disgust, fear, surprise happiness, sadness| http://www.lce.hut.fi/research/cogntech/emo.shtml
Reading-Leeds database (Greasley et al., 1995; Greasley et al., 2000; Roach et al., 1998; Stibbard, 2001)| A| Natural: unscripted interviews on radio/television in which speakers are asked by interviewers to relive emotionally intense experiences| 5 hours| English| emotion labels| Range of full blown emotions|
Campbell CREST database, (ongoing) (Douglas-Cowie et al., 2003)| A| Natural: volunteers record their domestic and social spoken interactions for extended periods throughout the day| Target?1000 hrs over 5 years at 2004 500+ hours| English, Japanese, Chinese| 3 levels of label: speaker state, speaking style, voice quality| Wide range of emotional states and emotion-related attitudes| http://feast.his.atr.jp/data/, http://www.mpi.nl/lrec/papers/lrec-pap-06-nick-speech.pdf
Capital Bank Service and Stock Exchange Customer Service (used by Devillers & Vasilescu, 2004)| A| Natural: call center human-human interactions| Stock Exchange: 100 dialogues, 5000 speaker turns; Capital Bank: 250 dialogues, 5000 turns of word extracts| French| 2 annotators using categorical labels| Mainly negative - fear, anger, stress| http://www.isca-speech.org/archive/sp2004/sp04_205.html
France et al. (France, Shiava, Silverman, Silverman & Wilkes, 2000)| A| Natural: therapy sessions and phone conversations. Post therapy evaluation sessions were also used to elicit speech for control subjects.| 115 subjects 48 females, 67 males, controls (therapists), 38, patients 77| English| DSM-IV diagnosis of depression| Dysthemic disorder; major depression (single episode and recurrent, no psychotic features)|
SYMPAFLY (as used by Batliner et al., 2004b) (Batliner et al., 2004, Batliner et al., 2004)| A| Human machine dialogue system| 110 dialogues, 29200 words

(i.e. tokens, not vocabulary)| German| Word-based emotional user states| Joyful, neutral, emphatic, surprised, ironic, helpless, touchy, angry, panic|
Belfast Boredom database (Cowie, et al 2003)| A| Induced: naming objects on computer screen| 12 subjects| English| Self rating of boredom/interest, rate at which trials completed, task error rate| Boredom| Belfast Boredom database (Cowie, et al. 2003)
DARPA Communicator corpus (as used by Ang et al., 2002) (Ang, et al, 2002)| A| Human machine dialogue system (Extracts from recordings of simulated interactions with a call centre)| 13187 utterances, average length about 2.75 words, 1750 considered emotional| English| Categorical emotional labeling by 5 students from 1 of 7 possible labels| Frustration, annoyance, neutral, tired, amused, other, not applicable| http://www.speech.cs.cmu.edu/Communicator/
Fernandez and Picard (Fernandez & Picard, 2003)| A| Induced: subjects give verbal responses to maths problems in simulated driving context| Data reported from 4 subjects| English| Performance in tasks used as indicator of subject state| Stress|
Lee et al. (Lee & Narayanan, 2003)| A| Dialogue with automated system| 1187 calls, Total 7200 utterances| English| Tagged: negative and non negative| Negative, non-negative| http://sail.usc.edu/publications/euro_fuzzy.pdf
Yacoub et al| A| Acted| 2433 utterances from 8 actors| English| Emotional categorical labelling| hot anger, cold anger, happy, sadness, disgust, panic, anxiety, despair, interest, shame, pride, boredom, contempt elation, neutral| http://www.isca-speech.org/archive/eurospeech_2003/e03_0729.html
Tolkmitt and Scherer (Tolkmitt & Scherer, 1986)| A| Induced: Cognitive stress induced through slides containing logical problems; emotional stress induced through slides of human bodies showing skin disease/accident injuries| 60 subjects (33 male, 27 female)| German| Low versus high stress| Stress (both cognitive and emotional)|
Iriondo et al. (Iriondo et al., 2000)| A| Contextualised acting: subjects asked to read passages written with appropriate emotional content| 8 subjects reading paragraph length passages (20?40 mm s each)| Spanish| Emotional categorical labelling| Desire, disgust, fury, fear, joy, surprise, sadness|
Mozziconacci (Mozziconacci 1998)| A| Contextualised acting: actors asked to read semantically neutral sentences in range of emotions, but practised on emotionally loaded sentences beforehand to get in the right mood| 3 subjects reading 8 semantically neutral sentences (each repeated 3 times)| Dutch| Emotional categorical labelling| Anger, boredom, fear, disgust, guilt, happiness, haughtiness, indignation, joy, rage, sadness, worry, neutrality|
McGilloway Database (McGilloway, 1997)| A| Contextualised acting: subjects asked to read passages written in appropriate emotional tone and content for each emotional state| 40 subjects reading 5 passages each| English| Emotional categorical labelling| Anger, fear, happiness, sadness, neutrality|
Belfast structured Database (McGilloway et al., 2000)| A| Contextualised acting: subjects read 10 McGilloway?style passages AND 10 other passages?scripted versions of naturally occurring emotion in the Belfast Naturalistic Database| 50 subjects reading 20 passages| English| Emotional categorical labelling| Anger, fear, happiness, sadness, neutrality|
Danish Emotional Speech Database| A| Scripted (material not emotionally coloured)| 4 subjects read 2 words, 9 sentences and 2 passages in range of emotions| Danish| Emotional categorical labelling| Anger, happiness sadness, surprise, neutrality| http://www.isca-speech.org/archive/eurospeech_1997/e97_1695.html
Groningen ELRA corpus number S0020| A| Acted| 238 subjects reading 2 short texts| Dutch| Unclear| Database only partially oriented to emotion|

http://www.elda.org/catalogue/en/speech/S0020.html
Berlin database (Kienast & Sendlmeier, 2000)| A| Acted: Scripted (material selected to be semantically neutral)| 10 subjects (5 M, 5 F) reading 10 sentences each| German| Emotional categorical labelling| Anger- hot, boredom, disgust, fear- panic, happiness, sadness-sorrow, neutrality| http://www.expressive-speech.net/
Leinonen and Hiltunen (Leinonen & Hiltunen, 1997)| A| Acted| 483 samples selected down to 120 (one word utterances)| Finnish| Emotional categorical labelling| Angry, frightened, commanding, pleading content, (naming), astonished, scornful, admiring, sad|
Nakatsu et al. (Nakatsu, Tosa & Nicholson 1999)| A| Utterances acted in various emotions then imitated by speakers| 100 words, uttered (by 100 speakers) once for each emotion, totalling 800 words| Japanese| Emotional categorical labelling| Neutrality, anger, sadness, fear, disgust, playfulness, surprise, happiness|
Nogueiras et al 2001| A| Acted| 6 sentences, paragraphs, words, numbers by each of 2 Spanish, French Slovenian, English speakers| Spanish, French, Slovenian, English| Emotional categorical labelling| Anger, disgust, fear, joy, sadness, surprise, neutrality| http://www.isca-speech.org/archive/eurospeech_2001/e01_2679.html
Media Team corpus in Finnish| A| Acted| 14 actors (8m, 6f) Finnish passage of aprox 100 words read in 6 emotions plus neutral| Finnish| Emotional categorical labelling| Happiness/joy, sadness, fear, anger, boredom, disgust, neutral| http://www.mediateam.oulu.fi/publications/pdf/438.pdf, http://www.mediateam.oulu.fi/projects/mpappce/?lang=en
ORATOR| A| Acted (actors asked to interpret same 8 sentence monologue in different social situations)| 117 samples produced by professional actors, 14 non-actors produced 28 samples| German| Not specified| Not specified| http://mplab.ucsd.edu/databases/databases.html
RUFACS| V| Naturalistic/Induced| 100 subjects, 2.5 minutes each. Aiming for 400-800 mins| n/a| unclear| not exactly specified, but spontaneous facial expressions| http:mplab.ucsd.edu/databases/databases.html
The AR face database| V| Posed| 154 subjects (82 male, 74 female) 26 pictures per person| n/a| Pre-defined| Smile, anger, scream neutral| http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html
CVL face database| V| Posed| 114 subjects (108 male, 6 female) 7 pictures per person| n/a| Pre-defined| Smile| http://lrv.fri.uni-lj.si/facedb.html
The Japanese Female Facial Expression (JAFFE) Database| V| Posed| 10 subjects| n/a| Rated by 60 Japanese raters using 6 emotion adjectives| Sadness, happiness, surprise, anger, disgust, fear, neutral| http://www.mis.atr.co.jp/~mlyons/jaffe.html
CMU PIE Database (CMU Pose, Illumination, and Expression (PIE) database)| V| Posed| 68 subjects| n/a| Pre-defined| Neutral, smile, blinking and talking| http://www.ri.cmu.edu/projects/project_418.html
The Yale Face Database A| V| Posed| 15 subjects| n/a| Pre-defined| Sad, sleepy, surprised| http://cvc.yale.edu/projects/yalefaces/yalefaces.html
CMU Facial Expression Database (Cohn-Kanade)| V| Posed| 210 subjects| n/a| Pre-defined| Six of the displays were based on descriptions of prototypic emotions (i.e., joy, surprise, anger, fear, disgust, and sadness)| http://vasc.ri.cmu.edu/idb/html/face/facial_expression/index.html
ATandT The Database of Faces (formerly called ORL database)| V| Posed| 40 subjects, 10 pictures per person| n/a| Pre-defined| Smiling| http://www.uk.research.att.com/facedatabase.html
The Psychological Image Collection at Stirling| V| Posed| Aberdeen: 116

subjects, Nottingham scans: 100, Nott-faces-original: 100, Stirling faces:36| n/a| Pre-defined| Smile, surprise, disgust| http://pics.psych.stir.ac.uk/
Special issue articlesRecommended articles
References

1. Amir et al., 2000
N. Amir, S. Ron, N. Laor
Analysis of an emotional speech corpus in Hebrew based on objective criteria
E. Douglas-Cowie, R. Cowie, M. Schroeder (Eds.), Proceedings of the ISCA workshop on speech and emotion, belfast 2000 (2000), pp. 29-33
View in ScopusGoogle Scholar
2. Ang et al., 2002
J. Ang, R. Dhillon, A. Krupski, E. Shriberg, A. Stolcke
Prosody-based automatic detection of annoyance and frustration in human?computer dialog
Proceedings ICSLP, Denver, Colorado, Sept. 2002 (2002)
Google Scholar
3. Auberge et al., 2003
V. Auberge, N. Audibert, A. Rilliard
Why and how to control the authentic emotional speech corpora
Eurospeech 2003, Geneva, Switzerland (2003)
pp. 185?188
Google Scholar
4. Bachorowski, 1999
J.A. Bachorowski
Vocal expression and perception of emotion
Current Directions in Psychol Science, 8 (2) (1999), pp. 53-57
CrossrefView in ScopusGoogle Scholar
5. Balomenos et al., 2005
Balomenos, T., Cowie, R., Malinescos, G., McMahon, E., Kasderidis, & S. (2003). Data Description, Sampling Methods, Experimental Design. Deliverable ND1.5: Retrieved 19 March, 2005 from http://manolito.image.ece.ntua.gr/oresteia/.
Google Scholar
6. Balomenos et al., 2004
T. Balomenos, A. Raouzaiou, S. Ioannou, A. Drosopoulos, K. Karpouzis, S. Kollias
Emotion analysis in man-machine interaction systems
Workshop on multimodal interaction and related machine learning algorithms. (MLMI04) 2004 (2004)
Google Scholar
7. Banse and Scherer, 1996
R. Banse, K.R. Scherer
Acoustic profiles in vocal emotion expression
Journal of Personality and Social Psychology, 70 (3) (1996), pp. 614-636
View in ScopusGoogle Scholar
8. Batliner et al., 2003
A. Batliner, K. Fischer, R. Huber, J. Spilker, E. Nöth
How to find trouble in communication
Speech Communication, 40 (2003), pp. 117-143
View PDFView articleView in ScopusGoogle Scholar
9. Batliner et al., 2004
A. Batliner, C. Hacker, S. Steidl, E. Noth, S. D'Arcy, M. Russell, _et al._
You stupid tin box?children interacting with the AIBO robot: A cross-linguistic emotional speech corpus

Proceedings LREC 2004 (2004)
Google Scholar
10. Batliner et al., 2004
A. Batliner, C. Hacker, S. Steidl, E. Noth, J. Haas
From emotion to interaction: Lessons learned from real human?machine dialogues
Proceedings of Affective Dialogue Systems: Tutorial and Research Workshop, ADS 2004, June 14?16, Springer-Verlag, Kloster Irsee, Germany (2004)
pp. 1?12
Google Scholar
11. Batocchi and Pianesi, 2004
Batocchi, A., & Pianesi, F. (2004). PianesiDaFEx: Database of kinetic facial expressionsproceedings sli-gscp workshop comunicazione parlata e manifestazione delle emozioni, padova (Italy) 30 Nov.-1, Dec. 2004.
Google Scholar
12. Berger and Bradac, 1982
C.R. Berger, J.J. Bradac
Language and social knowledge: Uncertainty in interpersonal relationships
Edward Arnold, London (1982)
Google Scholar
13. Cacioppo et al., 1993
J.T. Cacioppo, D.J. Klein, G.G. Berntson, E. Hatfield
The psychophysiology of emotion
M. Lewis, J. Haviland (Eds.), Handbook of emotions, Guilford Press, NewYork (1993), pp. 119-142
Google Scholar
14. Campbell, 2002a
N. Campbell
What type of inputs will we need for expressive speech synthesis?
Proceedings of 2002 IEEE workshop on speech synthesis (2002)
pp. 95?98
Google Scholar
15. Campbell, 2002b
N. Campbell
Recording and storing of speech data
Proceedings LREC (2002), p. 2002
Google Scholar
16. Campbell, 2004
N. Campbell
Specifying affect and emotion for expressive speech synthesis
A. Gelbukh (Ed.), Computational linguistics and intelligent text processing: Fifth international conference, CICLing 2004 Seoul, Korea, February 15?21 (2004), pp. 395-406
CrossrefView in ScopusGoogle Scholar
17. Center for the Study of Emotion and Attention (CSEA-NIMH), 1999
Center for the Study of Emotion and Attention (CSEA-NIMH)
Center for the study of emotion and attention (CSEA-NIMH) The International affective picture system: Digitized photographs
The center for research in psychophysiology, University of Florida, Gainesville, FL (1999)
Google Scholar
18. Chung, 2000
Chung, S. -J. (2000). _L'expression et la perception de l.émotion extraite de la parole spontanée: évidences du coréen et de l'anglais_. Unpublished doctoral dissertation, Université de la Sorbonne Nouvelle, Paris III.

Retrieved March 19, 2005, from http://www.geocities.com/soojinchung/.
Google Scholar
19. Clark, 1983
D.M. Clark
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 (1983), pp. 27-49
View PDFView articleView in ScopusGoogle Scholar
20. Cornelius, 1966
R. Cornelius
The science of emotion: Research and tradition in the psychology of emotion
Prentice-Hall, Upper Saddle River, NJ (1966)
Google Scholar
21. Cowie, 2005
R. Cowie
What are people doing when they assign everyday emotion terms?
Psychological Inquiry, 16 (1) (2005), pp. 11-48
View in ScopusGoogle Scholar
22. Cowie and Cornelius, 2003
R. Cowie, R. Cornelius
Describing the Emotional States that are Expressed in Speech
Speech Communication, 40 (2003), pp. 5-32
View PDFView articleView in ScopusGoogle Scholar
23. Cowie et al., 1999
R. Cowie, E. Douglas-Cowie, B. Apolloni, J. Taylor, A. Romano, W. Fellenz
What a neural net needs to know about emotion words
N. Mastorakis (Ed.), Computational intelligence and applications, World Scientific Engineering Society (1999), pp. 109-114
View in ScopusGoogle Scholar
24. Cowie et al., 2003
R. Cowie, A. McGuiggan, E. McMahon, E. Douglas-Cowie
Speech in the process of becoming bored
Proceedings of 15th international congress of phonetic sciences, Barcelona (2003)
Google Scholar
25. Cowie and Schroeder, 2005
R. Cowie, M. Schroeder
Piecing together the emotion jigsaw
S. Bengio, H. Bourlard (Eds.), Machine learning for multimodal interaction: Lecture notes in computer science, Vol. 3361, Springer Verlag, Heidelberg (2005), pp. 305-317
CrossrefView in ScopusGoogle Scholar
26. Cox, 2004
Cox, C. (2004). SALAS (Sensitive Artificial Listener induction techniques). Paper presented to HUMAINE Network of Excellence Summer School, 2004. Retrieved 23 March 2005 from http://emotion-research.net/ws/summerschool1/SALAS.ppt#259.
Google Scholar
27. Craggs and Wood, 2004
R. Craggs, M. Wood
A 2 dimensional annotation scheme for emotion in dialogue
AAAI spring symposium on exploring attitude and affect in text: Theories and applications, AAAI Press, Stanford University (2004)
pp. 44?49

Google Scholar
28. Davidson, 1999
R.J. Davidson
Neuropsychological perspectives on affective styles and their cognitive consequences
T. Dalgleish, M. Power (Eds.), Handbook of cognition and emotion, Wiley, New York (1999), pp. 103-123
CrossrefGoogle Scholar
29. Descartes, 1911?1912
R. Descartes
The passions of the soul (1649)
E.L. Haldane, G.R. Ross (Eds.), The philosophical works of descartes, Cambridge University Press, Cambridge (1911?1912)
30. Devillers and Vasilescu, 2004
L. Devillers, I. Vasilescu
Reliability of lexical and prosodic cues in two real-life spoken dialog corpora
Proceedings LREC 2004 (2004)
Google Scholar
31. Douglas-Cowie et al., 2003
E. Douglas-Cowie, N. Campbell, Cowie, P. Roach
Emotional speech: Towards a new generation of databases
Speech Communication, 40 (1?2) (2003), pp. 33-60
View PDFView articleView in ScopusGoogle Scholar
32. Ekman, 1999
P. Ekman
Basic emotions
T. Dalgleish, M.J. Power (Eds.), Handbook of cognition and emotion, Wiley, New York (1999), pp. 301-320
CrossrefGoogle Scholar
33. Ekman and Friesen, 1969
P. Ekman, W. Friesen
The repertoire of non verbal behavior: Categories, origins, usage and coding
Semiotica, 1 (1969), pp. 49-98
CrossrefGoogle Scholar
34. Ekman and Friesen, 1975a
P. Ekman, W. Friesen
Pictures of facial affect
Consulting Psychologists' Press, Palo Alto, CA (1975)
Google Scholar
35. Ekman and Friesen, 1975b
P. Ekman, W. Friesen
Unmasking the face: A guide to recognizing emotions from facial clues
Prentice-Hall, Inc., Englewood Cliffs, NJ (1975)
Google Scholar
36. Ekman et al., 1987
P. Ekman, W.V. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, _et al._
Universals and cultural differences in the judgments of facial expressions of emotion
Journal of Personality and Social Psychology, 53 (4) (1987), pp. 712-717
View in ScopusGoogle Scholar
37. Fernandez and Picard, 2003
R. Fernandez, R. Picard

Modeling drivers' speech under stress
Speech Communication, 40 (2003), pp. 145-159
View PDFView articleView in ScopusGoogle Scholar
38. France et al., 2000
D. France, R. Shiavi, S. Silverman, M. Silverman, D. Wilkes
Acoustical properties of speech as indicators of depression and suicidal risk
IEEE Transactions on Biomedical Engineering, 47 (7) (2000)
Google Scholar
39. Frijda, 1986
N.H. Frijda
The emotions
Cambridge University Press, Cambridge, UK (1986)
Google Scholar
40. Goldie et al., 2004
Goldie, P., Doring, S., & WPIO members. (2004). Report on ethical prospects and pitfalls for emotion-oriented systems. Report for EC funded IST Network of Excellence 507422 (HUMAINE). Retrieved 22nd March 2005 from http://emotion-research.net/deliverables/D10a.pdf.
Google Scholar
41. Greasley et al., 1995
P. Greasley, J. Setter, M. Waterman, C. Sherrard, P. Roach, S. Arnfield, _et al._
Representation of prosodic and emotional features in a spoken language database
Proceedings XIIIth international congress of phonetic sciences, Stockholm, Vol. 1 (1995)
pp. 242?245
Google Scholar
42. Greasley et al., 2000
P. Greasley, C. Sherrard, M. Waterman
Emotion in language and speech: Methodological issues in naturalistic approaches
Language and Speech, 43 (2000), pp. 355-375
CrossrefView in ScopusGoogle Scholar
43. Gross and Levenson, 1995
J.J. Gross, R.W. Levenson
Emotion elicitation using films
Cognition and Emotion, 9 (1995), pp. 87-108
CrossrefView in ScopusGoogle Scholar
44. Hadfield and Marks, 2000
P. Hadfield, P. Marks
This is your captain dozing
New Scientist, 1682267 (2000), p. 21
View in ScopusGoogle Scholar
45. Iriondo, 2000
I. Iriondo, _et al._
Validation of an acoustical modelling of emotional expression in spanish using speech synthesis techniques
Proceedings of the ISCA workshop on speech and emotion, Belfast 2000 (2000)
Google Scholar
46. James, 1884
W. James
What is emotion?
Mind, 9 (1884), pp. 188-205

CrossrefGoogle Scholar
47. Juslin and Laukka, 2002
P. Juslin, P. Laukka
Communication of emotions in vocal expression and music performance
Psychological Bulletin, 129 (5) (2002), pp. 770-814
Google Scholar
48. Kenealy, 1988
P. Kenealy
Validation of a music mood induction procedure: Some preliminary findings
Cognition and Emotion, 2 (1988), pp. 41-48
CrossrefGoogle Scholar
49. Kienast and Sendlmeier, 2000
M. Kienast, W.F. Sendlmeier
Acoustical analysis of spectral and temporal changes in emotional speech
R. Cowie, E. Douglas-Cowie, M. Schroeder (Eds.), Speech and emotion: Proceedings of the ISCA workshop. Newcastle, Co. Down, September 2000 (2000), pp. 92-97
Google Scholar
50. Kollias et al., 2004
Kollias, S., Karpouzis, K., Ioannpou, S., & Raouzaiou, A. (2004). The facial feature extraction subsystem. Report (Deliverable D10) for EC funded fifth framework IST project IST-2000?29319 ERMIS (Emotionally Rich Man-machine Intelligent System). See http://manolito.image.ece.ntua.gr/ermis/.
Google Scholar
51. Kollias et al., 2005
Kollias, S. et al., 2005. Presentation to the final review of the ERMIS project (EC funded Fifth Framework IST project IST-2000-29319), Athens, March 2005.
Google Scholar
52. Lang et al., 1999
P.J. Lang, M.M. Bradley, B.N. Cuthbert
International affective picture system IAPS: Technical manual and affective ratings
The center for research in psychophysiology, University of Florida, Gainesville, FL (1999)
Google Scholar
53. Laukka, 2005
P. Laukka
Verbal expressions of emotion: Discrete emotions and dimensional accounts
Acta Universitatis Uppsalensis, Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences (2004), p. 141
Google Scholar
54. LeDoux, 1987
J.E. LeDoux
Emotion
V.B. Mountcastle (Ed.), Handbook of physiology, Section I: The nervous system, Higher functions of the brain, Vol. V, American Physiological Society, Bethesda, MD (1987)
Google Scholar
55. Lee and Narayanan, 2003
C. Lee, S. Narayanan
Emotion recognition using a data-driven fuzzy inference system
Proceedings eurospeech 2003, Geneva (2003)
Google Scholar

56. Lee and Narayanan, 2005
C.M. Lee, S. Narayanan
Towards detecting emotions in spoken dialogs
IEEE transactions on speech and audio processing, 13 (2) (2005), pp. 293-303
View in ScopusGoogle Scholar
57. Leinonen and Hiltunen, 1997
L. Leinonen, T. Hiltunen
Expression of emotional-motivational connotations with a one-word utterance
Journal of the Acoustical Society of America, 102 (3) (1997), pp. 1853-1863
View in ScopusGoogle Scholar
58. Lykken, 1984
D.T. Lykken
Polygraphic interrogation
Nature, 387 (1984), pp. 681-684
CrossrefView in ScopusGoogle Scholar
59. McGilloway, 1997
McGilloway, S. (1997). _Negative symptoms and speech parameters in schizophrenia_. Unpublished doctoral thesis, Queen's University, Belfast.
Google Scholar
60. McGilloway et al., 2000
S. McGilloway, R. Cowie, E. Douglas-Cowie, S. Gielen, M. Westerdijk, S. Stroeve
Automatic recognition of emotion from voice: A rough benchmark
R. Cowie, E. Douglas-Cowie, M. Schroeder (Eds.), Speech and emotion: Proceedings of the ISCA workshop. Newcastle, Co. Down, September 2000 (2000), pp. 207-212
Google Scholar
61. McMahon et al., 2003
E. McMahon, R. Cowie, S. Kasderidis, J. Taylor, S. Kollias
What chance that a DC could recognise hazardous mental states from sensor outputs?
Proc, DC Tales conference, Sanotrini, June 2003 (2003)
Google Scholar
62. Mervis et al., 1976
C.B. Mervis, J. Catlin, E. Rosch
Relationships among goodness-of-example, category norms, and word frequency
Bulletin of the Psychonomic Society, 7 (1976), pp. 268-284
Google Scholar
63. Mozziconacci, 1998
Mozziconacci, S. (1998). _Speech variability and emotion: Production and perception_. Unpublished doctoral thesis. Technical University Eindhoven, Eindhoven.
Google Scholar
64. Nakatsu et al., 1999
R. Nakatsu, N. Tosa, J. Nicholson
Emotion recognition and its application to computer agents with spontaneous interactive capabilities
Proceedings IEEE workshop on multimedia signal processing (1999)
pp. 439?444
Google Scholar
65. Nogueiras et al., 2001
A. Nogueiras, A. Moreno, A. Bonafonte, J. Mariño
Speech emotion recognition using hidden markov models
Proceedings Eurospeech 2001, Aalborg, Denmark (2001)

Google Scholar
66. Ortony and Turner, 1990
A. Ortony, T.J. Turner
Whats basic about basic emotions
Psychological Review, 97 (3) (1990), pp. 315-331
View in ScopusGoogle Scholar
67. Picard et al., 2001
R.W. Picard, E. Vyzas, J. Healey
Toward machine emotional intelligence: Analysis of affective physiological state
IEEE Transactions On Pattern Analysis and Machine Intelligence, 23 (2001), pp. 1175-1191
View in ScopusGoogle Scholar
68. Place, 1996
U.T. Place
Folk psychology from the standpoint of conceptual analysis
W. O'Donohue, R.F. Kitchener (Eds.), The philosophy of psychology, Sage, London (1996), pp. 264-270
Google Scholar
69. Pollick et al., 2003
F.E. Pollick, H. Hill, A. Calder, H. Paterson
Recognising facial expression from spatially and temporally modified movements
Perception, 32 (2003), pp. 813-826
View in ScopusGoogle Scholar
70. Polzin and Waibel, 2000
T.S. Polzin, A. Waibel
Emotion-sensitive human?computer interfaces
E. Douglas-Cowie, R. Cowie, M. Schroeder (Eds.), Proceedings of the ISCA workshop on speech and emotion, Belfast 2000 (2000), pp. 201-206
pp. 201?206
Google Scholar
71. Roach, 2000
P. Roach
Techniques for the phonetic description of emotional speech
E. Douglas-Cowie, R. Cowie, M. Schroeder (Eds.), Proceedings of the ISCA workshop on speech and emotion, Belfast 2000 (2000)
Google Scholar
72. Roach et al., 1998
P. Roach, R. Stibbard, J. Osborne, S. Arnfield, J. Setter
Transcription of prosodic and paralinguistic features of emotional speech
Journal of the International Phonetic Association, 28 (1998), pp. 83-94
View in ScopusGoogle Scholar
73. Russell and Barrett-Feldman, 1999
J. Russell, L. Barrett-Feldman
Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant
Journal of Personality and Social Psychology, 5 (1999), pp. 37-63
Google Scholar
74. Sabini and Silver, 2005
Sabini, Silver
Why emotion names and experiences don't neatly pair
Psychological Inquiry, 16 (1) (2005), pp. 1-10
CrossrefView in ScopusGoogle Scholar
75. Scherer, 1999

K.R. Scherer
Appraisal theory
T. Dalgleish, M.J. Power (Eds.), Handbook of cognition and emotion, Wiley, New York (1999), pp. 637-663
CrossrefGoogle Scholar
76. Scherer and Ceschi, 1997
K.R. Scherer, G. Ceschi
Lost luggage emotion: A field study of emotion?antecedent appraisal
Motivation and Emotion, 21 (1997), pp. 211-235
View in ScopusGoogle Scholar
77. Scherer and Ceschi, 2000
K.R. Scherer, G. Ceschi
Studying affective communication in the airport: The case of lost baggage claims
Personality and Social Psychology Bulletin, 26 (3) (2000), pp. 327-339
CrossrefView in ScopusGoogle Scholar
78. Scherer et al., 1998
K.R. Scherer, T. Johnstone, T. Bänziger
Automatic verification of emotionally stressed speakers: The problem of individual differences
Proceedings of the international workshop on speech and computer, St Petersburg (1998)
Google Scholar
79. Scherer et al., 2004
Scherer, K. et al. (2004). Preliminary plans for exemplars: Theory. Report for EC funded IST Network of Excellence 507422 (HUMAINE). Retrieved 22nd March 2005 from http://emotion-research.net/deliverables/D3c.pdf.
Google Scholar
80. Schneider, 1995
W.X. Schneider
VAM: A neuro-cognitive model for visual control of segmentation, object recognition, and space-based motor action
Visual Cognition, 2 (1995), pp. 331-376
CrossrefView in ScopusGoogle Scholar
81. Schroeder, 2003
M. Schroeder
Experimental study of affect bursts
Speech Communication, 40 (1?2) (2003), pp. 99-116
Google Scholar
82. Stibbard, 2001
Stibbard, R. (2001). _Vocal expression of emotions in non-laboratory speech_. Unpublished doctoral dissertation, University of Reading, UK.
Google Scholar
83. Sweeney, 2005
Sweeney, F. (2005). _The words people use to describe their emotional lives BSc thesis_ : School of Psychology, Queen's University Belfast.
Google Scholar
84. Tolkmitt and Scherer, 1986
F. Tolkmitt, K.R. Scherer
Effect of experimentally induced stress on vocal parameters
Exp Psychol Hum Percept Perform, 12 (3) (1986), pp. 302-313
View in ScopusGoogle Scholar
85. Tsukahara and Ward, 2001
Tsukahara, W., & Ward, N. (2001). Responding to subtle, fleeting changes in

the user's internal state. CHI'01, ACM, 2001, pp. 77?84.
Google Scholar
86. Velten, 1968
E. Velten
A laboratory task for induction of mood states
Behavior Research and Therapy, 6 (1968), pp. 473-482
View PDFView articleView in ScopusGoogle Scholar
87. Ward, 2003
N. Ward
On the expressive competencies needed for responsive systems
CHI 2003 workshop on subtle expressivity for characters and robots, April 7
2003, Fort Lauderdale, FL (2003)
Google Scholar
88. Yacoub et al., 2003
S. Yacoub, S. Simske, X. Lin, J. Burns
Recognition of emotions in interactive voice response systems
Proceedings of the Eurospeech 2003, Geneva (2003)
Google Scholar
Cited by (95)
* ### Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge
2011, Speech Communication
Citation Excerpt :
Note that with a broader coverage of representative scenarios and emotion-related states that we will hopefully encounter in some future, the distinction between a few universal emotions and scenario-specific states will give room for a better conceptualisation of ?representativity? and ?genericity?.) Here, we will only give a short typology of emotional databases and refer to some representative examples; catalogues can be found, for instance, in (Ververidis and Kotropoulos, 2003; Cowie et al., 2005; Schuller et al., 2009e). Still widely used ? because freely available ? acted databases are The Danish Emotional Speech Database (DES) (Engberg et al., 1997), and The Berlin Emotional Speech Database (BES) (Burkhardt et al., 2005), as well as The Speech Under Simulated and Actual Stress (SUSAS) Database (Hansen and Bou-Ghazale, 1997).
Show abstract
More than a decade has passed since research on automatic recognition of emotion from speech has become a new field of research in line with its ?big brothers? speech and speaker recognition. This article attempts to provide a short overview on where we are today, how we got there and what this can reveal us on where to go next and how we could arrive there. In a first part, we address the basic phenomenon reflecting the last fifteen years, commenting on databases, modelling and annotation, the unit of analysis and prototypicality. We then shift to automatic processing including discussions on features, classification, robustness, evaluation, and implementation and system integration. From there we go to the first comparative challenge on emotion recognition from speech ? the INTERSPEECH 2009 Emotion Challenge, organised by (part of) the authors, including the description of the Challenge?s database, Sub-Challenges, participants and their approaches, the winners, and the fusion of results to the actual learnt lessons before we finally address the ever-lasting problems and future promising attempts.
* ### Speech Emotion Recognition Using Deep Learning Techniques: A Review
2019, IEEE Access
* ### CREMA-D: Crowd-sourced emotional multimodal actors dataset

2014, IEEE Transactions on Affective Computing
 * ### The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent
2012, IEEE Transactions on Affective Computing
 * ### A survey of affect recognition methods: Audio, visual, and spontaneous expressions
2009, IEEE Transactions on Pattern Analysis and Machine Intelligence
 * ### IEMOCAP: Interactive emotional dyadic motion capture database
2008, Language Resources and Evaluation
View all citing articles on Scopus
View Abstract

Part of special issue
Emotion and Brain
Edited by
J. G. Taylor, K. Scherer, R. Cowie
Download full issue
Other articles from this issue
 * ### A systems approach to appraisal mechanisms in emotion
May 2005
David Sander, ?, Klaus R. Scherer
View PDF
 * ### The interaction of attention and emotion
May 2005
John G. Taylor, Nickolaos F. Fragopanagos
View PDF
 * ### Emotion recognition in human?computer interaction
May 2005
N. Fragopanagos, J.G. Taylor
View PDF
View more articles
Recommended articles
No articles found.

of Google Ad-Tech Vendors.
You may choose not to allow some types of cookies. However, blocking some types may impact your experience of our site and the services we are able to offer. See the different category headings below to find out more or change your settings.
Allow all
Manage Consent Preferences
Strictly Necessary Cookies
Always active
These cookies are necessary for the website to function and cannot be switched off in our systems. They are usually only set in response to actions made by you which amount to a request for services, such as setting your privacy preferences, logging in or filling in forms. You can set your browser to block or alert you about these cookies, but some parts of the site will not then work. These cookies do not store any personally identifiable information.
Cookie Details List?
Functional Cookies
Functional Cookies
These cookies enable the website to provide enhanced functionality and personalisation. They may be set by us or by third party providers whose services we have added to our pages. If you do not allow these cookies then some or all of these services may not function properly.
Cookie Details List?
Performance Cookies
Performance Cookies
These cookies allow us to count visits and traffic sources so we can measure and improve the performance of our site. They help us to know which pages are the most and least popular and see how visitors move around the site.
Cookie Details List?
Targeting Cookies
Targeting Cookies
These cookies may be set through our site by our advertising partners. They may be used by those companies to build a profile of your interests and show you relevant adverts on other sites. If you do not allow these cookies, you will experience less targeted advertising.
Cookie Details List?
Back Button
Cookie List
Search Icon
Filter Icon
Clear
checkbox label label
Apply Cancel
Consent Leg.Interest
checkbox label label
checkbox label label
checkbox label label
Confirm my choices