Dimensional emotion representation as a basis for speech synthesis with non-extreme emotions

Marc Schröder

DFKI GmbH, Stuhsatzenhausweg 3, 66123 Saarbrücken, Germany,
schroed@dfki.de,
Web demo: http://mary.dfki.de/emotional.html

Abstract. Past attempts to model emotions for speech synthesis have focused on extreme, "basic" emotion categories. The present paper suggests an alternative representation of emotional states, by means of emotion dimensions, and explains how this approach can contribute to making speech synthesis a useful component of affective dialogue systems.

1 Introduction: The problem

In most studies exploring the expression of emotions in speech synthesis [1, 2], a number of between three and nine discrete, extreme emotional states are modelled. However, the often implicit assumption that the expression of a few basic or primary emotion categories is most important to model, and that other emotional states can somehow be derived from that, has been questioned by Cowie [3]. He argued that systems should be able to express less intense emotions more suitable for real life applications.

There are a number of alternative choices available for representing emotions. While well established in the psychological literature, these are not very well known in the intelligent user interfaces community. Instead, representations are used that may be suboptimal for a given task, as illustrated by the frequent use of the Darwinian concept of "basic emotions" in emotional speech synthesis [2].

Instead of proposing a single "best" emotion representation, I argue that different components of an affective dialogue system may best be using different emotion representations optimised for their respective tasks. In the following, a selection of such representations is presented. Their suitability for different aspects of affective dialogue systems is discussed, and the question of mapping between these representations is raised.

The use of emotion dimensions for speech synthesis is then demonstrated in more detail.

2 Types of emotion representation

2.1 Emotion categories

The most straightforward description of emotions is the use of emotion-denoting words, or category labels. Human languages have proven to be extremely powerful in producing labels for emotional states: Lists of emotion-denoting adjectives exist that include at least 107 English [4] and 235 German [5] items. Several approaches exist for reducing these to an essential core set.

Basic emotions. Especially in the Darwinian tradition of emotion research [6], there is general agreement that some fullblown emotions are more basic than others. From this point of view, the basic emotions correspond to specific, evolutionarily shaped functions benefiting survival. They constitute highly specific, intense, multi-faceted syndromes consisting of physiological changes, action tendencies, expressive behaviour etc. In databases of spontaneous emotional speech, these basic emotions are rarely found [7].

Superordinate emotion categories. Alternatively, emotion categories have been proposed as more fundamental than others on the grounds that they *include* the others. An example may clarify the idea: [8] proposed five prototypes underlying all emotion categories: Anger, love, joy, fear, and sadness. Joy, for example, would be subdivided into pride, contentment, and zest. [9] gives a short overview of recent proposals of such lists.

Essential everyday emotion terms. A pragmatic approach is to ask for the emotion terms that play an important role in everyday life. The approach is exemplified by the work of [10], who proposed a Basic English Emotion Vocabulary. Starting from lists of emotion terms from the literature, subjects were asked to select a subset which appropriately represents the emotions relevant in everyday life. A subset of 16 emotion terms emerged.

This short list of options shows that even if one decides to model emotions in terms of categories, it is not immediately clear what categories to use. The most frequently used categories may not be the most suitable ones for a given research question or application.

2.2 Appraisal-based descriptions

In cognitive emotion theories, the central concept is *appraisal* – an evaluation of a stimulus through relatively low-level, automatic cognitive processes. The appraisal of a stimulus determines the significance of the stimulus for the individual, and triggers an emotion as an appropriate response.

Details about how and according to which criteria the perceived stimuli are evaluated and which reactions are triggered have been worked out by a number of researchers. The most notable for a speech and emotion researcher is the *component process model* developed by Scherer [11], from which Scherer has made detailed physiological predictions about the vocal changes associated with certain emotions [12], which in their large majority were verified experimentally [13].

Another cognitive emotion model, detailing the presumed appraisal structure leading to the multitude of emotions, was proposed by Ortony, Clore and Collins [14]. In this so-called OCC model, emotions are seen as valenced reactions to three types of stimuli: Events, agents, and objects. The model is formulated in a way permitting its implementation in AI systems. Several conversational agent systems have adopted the model, in a so-called “affective reasoning” module [15].

2.3 Emotion dimensions

Many different approaches reported in the psychological literature have led to the proposal of dimensions underlying emotional concepts (see [1] for an overview). Through multidimensional scaling, semantic differential and other techniques, different researchers came to propose three dimensions. They are gradual in nature and represent the essential aspects of emotion concepts (how good or bad, how aroused or relaxed, how powerful or weak) rather than the fine specifications of individual emotion categories.

It is important to know that the names used for these dimensions were actually selected by the individual researchers *interpreting* their data, and did not arise from the data itself. That explains the relative arbitrarity in naming the dimensions that can be found throughout the literature (Mehrabian and Russell call them pleasure, arousal and dominance [16], Osgood et al. use the names evaluation, activity and potency [17], Cowie et al. use evaluation, activation and power [18]). In this paper, I will use the terms *evaluation* (synonymous to valence or pleasure), *activation* (used as synonymous to arousal and activity) and *power* (potency or dominance).

2.4 Suitability of a representation for a task

It is important to think carefully about the type of representation most suitable for a given task.

The planning component of a dialogue system needs to assess the situation and interpret the meaning of items in order to assess how to act and respond. This is the natural domain for appraisal-based descriptions, which can map situation appraisals to emotions.

In a dialogue, an emotional state may build up rather gradually, and may change over time as the interaction moves on. Consequently, a speech synthesis system should be able to gradually modify the voice in a series of steps towards an emotional state. In addition, it seems reasonable to assume that most human-machine dialogues will require the machine to express only mild, non-extreme emotional states. Therefore, the need to express fullblown emotions is a marginal rather than a central requirement, while the main focus should be on the system’s capability to express a large variety of emotional states of low to medium intensity. Emotion dimensions are a representation of emotional states which fulfils these requirements: They are naturally gradual, and are capable of representing low-intensity as well as high-intensity states. While they do not define the exact properties of an emotional state in the same amount of detail as a category label, they do capture the essential aspects of the emotional state.

Facial expressions, on the other hand, appear to be best generated using categorical descriptions of emotion. However, there is a need not only to vary the expressions in intensity, but also to blend them. Emotion dimensions can help in determining the similarity of emotion categories and therefore find the right "mixture" of facial attributes for intermediate states [19].

2.5 Mappings between emotion representations

It is an open research question how best to map between the different emotion representations of which only some have been listed above. As the different representations have been created for different purposes and capture different subsets of emotion-related phenomena, simple one-to-one mappings are not always possible.

Emotion categories can be *located* in emotion dimension space via rating tests [10]. The mapping from categories to dimensions is therefore a simple task, as long as the coordinates of the emotion category have been determined. The inverse, however, is not possible: As emotion dimensions only capture the most essential aspects of an emotion concept, they provide an underspecified description of an emotional state. For example, the coordinates for anger and disgust may be very close, because the two categories share the same activation/evaluation/power properties. The features distinguishing between the two categories cannot be represented using emotion dimensions, so that the corresponding region in space can only be mapped to "anger-or-disgust" rather than a specific category.

In appraisal-based descriptions, the link to and from categories is also possible. A given combination of appraisal outcomes corresponds to an emotion category.

The mapping from an appraisal-based representation to a dimensional representation becomes possible via the intermediate representation as emotion categories. This approach is not without pitfalls, however: Emotion categories are often represented by simple words, such as "anger", which are highly ambiguous. If the interpretation of the word "anger" in the context of its constituting appraisals differs from its interpretation when locating it on emotion dimensions, then the mapping becomes inaccurate. Disambiguating the emotion category, on the other hand, implies a fuller description of the emotion, which ultimately would require an all-encompassing emotion representation.

It should be kept in mind, therefore, that mappings between the currently existing emotion representations are necessarily imperfect.

A first attempt into this direction was nevertheless attempted in the NECA project [20]. An affective reasoning component working with the OCC model is used for determining the appropriate emotion in a given dialogue situation, represented as a combination of emotion category and intensity. This representation is mapped onto emotion dimensions, using the intensity value to linearly interpolate between the neutral state and the coordinates of the fully developed emotional state in the dimensional space. The speech synthesis component uses the dimensional representation to generate emotional speech as described below. Finally, the animation component uses the category and intensity representation for generating facial expressions.

3 Application to speech synthesis

Emotions can be expressed in speech synthesis either by using emotional speech databases as concatenation material or by formulating explicit prosody rules on the link between emotional states and their effects on speech prosody [2]. The work presented here follows to the latter approach.

3.1 Formulation of emotional prosody rules

Emotional prosody rules were formulated on the basis of a literature review and a database analysis.

A literature review [1] brought about the following results. An unambiguous agreement exists concerning the link between the activation dimension and the most frequently measured acoustic parameters: Activation is positively correlated with mean F0, mean intensity, and, in most cases, with speech rate. Additional parameters positively correlated with activation are pitch range, "blaring" timbre, high-frequency energy, late intensity peaks, intensity increase during a "sense unit", and the slope of F0 rises between syllable maxima. Higher activation also corresponds to shorter pauses and shorter inter-pause and inter-breath stretches.

The evidence for evaluation and power is less stable. There seems to be a tendency that studies which take only a small number of acoustic parameters into account do not find any acoustic correlates of evaluation and/or power.

The limited evidence regarding the vocal correlates of power indicates that power is basically recognised from the same parameter settings as activation (high tempo, high F0, more high-frequency energy, short or few pauses, large intensity range, steep F0 slope), except that sometimes, high power is correlated with lower F0 instead of higher F0, and power is correlated with vowel duration.

There is even less evidence regarding the acoustic correlates of evaluation. Positive evaluation seems to correspond to a faster speaking rate, less high-frequency energy, low pitch and large pitch range; a "warm" voice quality; and longer vowel durations and the absense of intensity increase within a "sense unit".

In a statistical analysis of the Belfast Naturalistic Emotion Database [7], perceptual ratings of the emotion dimensions activation, evaluation and power were correlated with acoustic measures (see [1, 21] for details).The study replicated the basic patterns of correlations between emotion dimensions and acoustic variables. It was shown that the acoustic correlates of the activation dimension were highly stable, while correlates of evaluation and power were smaller in number and magnitude and showed a high variability between male and female speakers. In addition, the analysis provided numerical linear regression coefficients which were used as a starting point for the formulation of quantified emotion prosody rules.

The effects found in the literature and in the database analysis were formulated in a way suitable for implementation in a speech synthesis system.

	Prosodic parameter	Coefficients		
		Activation	Evaluation	Power
fundamental frequency	pitch	0.3	0.1	-0.1
	pitch-dynamics	0.3%		-0.3%
	range	0.4		
	range-dynamics	1.2%		0.4%
	accent-prominence	0.5%	-0.5%	
	preferred-accent-shape		$E\leq-20$: falling $-20<E\leq40$: rising $E>40$: alternating	
	accent-slope	1%	-0.5%	
	preferred-boundary-type			$P\leq0$: high $P>0$: low
tempo	rate	0.5%	0.2%	
	number-of-pauses	0.7%		
	pause-duration	-0.2%		
	vowel-duration		0.3%	0.3%
	nasal-duration		0.3%	0.3%
	liquid-duration		0.3%	0.3%
	plosive-duration	0.5%	-0.3%	
	fricative-duration	0.5%	-0.3%	
	volume	0.33%		

Table 1. Implementable emotion dimension prosody rules. Values on emotion dimensions range from -100 to 100, with 0 being the "neutral" value. The percentage values are *factors* – see text for details.

Table 1 presents the essential data required to express emotions in a speech synthesis system using emotion dimensions. The columns represent the emotion dimensions, while the rows list all the acoustic parameters for which emotion effects are modelled.

The numeric data fields represent the linear coefficients quantifying the effect of the given emotion dimension on the acoustic parameter, i.e. the change from the neutral default value. As an example, the value 0.5% linking Activation to `rate` means that for an activation level of +50, rate increases by +25%, while for an activation level of -30, rate decreases by -15%.

3.2 Implementation: EmoSpeak

The German text-to-speech system Mary (**M**odular **A**rchitecture for **R**esearch on speech s**Y**nthesis [22]) was used as the platform for the implementation of the emotional prosody rules specified in Table 1. This system was most suitable for the task because of the high degree of flexibility and control over the various processing steps, which arises from the use of the system-internal representation language MaryXML.

A major design feature in the technical realisation of the emotional speech synthesis system was that the acoustic effects of emotions should be specified in one single module. This module adds appropriate MaryXML annotations to the text which are then realised by the respective modules within the MARY system. As a consequence, all of the parameters are global in the sense that they will be applied to all enclosed text. This approach is considered the most transparent, as the link between emotions and their acoustic realisations is not hidden in various processing components, and the easiest to maintain and adapt, as all rules are contained in one document.

A simple emotion dimension markup language (EDML) was created, annotating text using a single `<emotion>` tag in which the positions on emotion dimensions are specified as the values of the `activation`, `evaluation` and `power` attributes. An example EDML document is the following:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<emotion activation="+30" evaluation="+70" power="+20">
Wie wunderbar!
</emotion>
```

For transforming the emotion markup language into the MARY-internal MaryXML, the emotion realisation module uses an XSLT stylesheet implementing the rules listed in Table 1. Applying that XML transformation to the above example creates the following MaryXML document:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<maryxml xmlns="http://mary.dfki.de/2002/MaryXML"
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
         version="0.3" xml:lang="de">
<voice name="de6">
<prosody accent-prominence="-20%" accent-slope="-5%"
   fricative-duration="-6%" liquid-duration="+27%"
   nasal-duration="+27%" number-of-pauses="+21%"
   pause-duration="-6%" pitch="124" pitch-dynamics="-12%"
   plosive-duration="-6%" preferred-accent-shape="alternating"
   preferred-boundary-type="low" range="37" range-dynamics="+4%"
   rate="+29%" volume="60" vowel-duration="+27%">
Wie wunderbar!
</prosody>
</voice>
</maryxml>
```

Note that the voice is set to `de6`, one of the two voice databases with three activation-related voice qualities created in the NECA project [23]. The attributes of the prosody tag, though listed here in alphabetical order rather than grouped by their meaning, correspond exactly to the rows of Table 1. As motivated above, the MaryXML document retains no information about the emotional state, but all information required for the desired acoustic realisation of that emotional state. Like any other MaryXML document, the above example can be processed by the standard MARY modules.

A graphical user interface was programmed to allow for an interactive exploration of the emotional speech synthesis system.

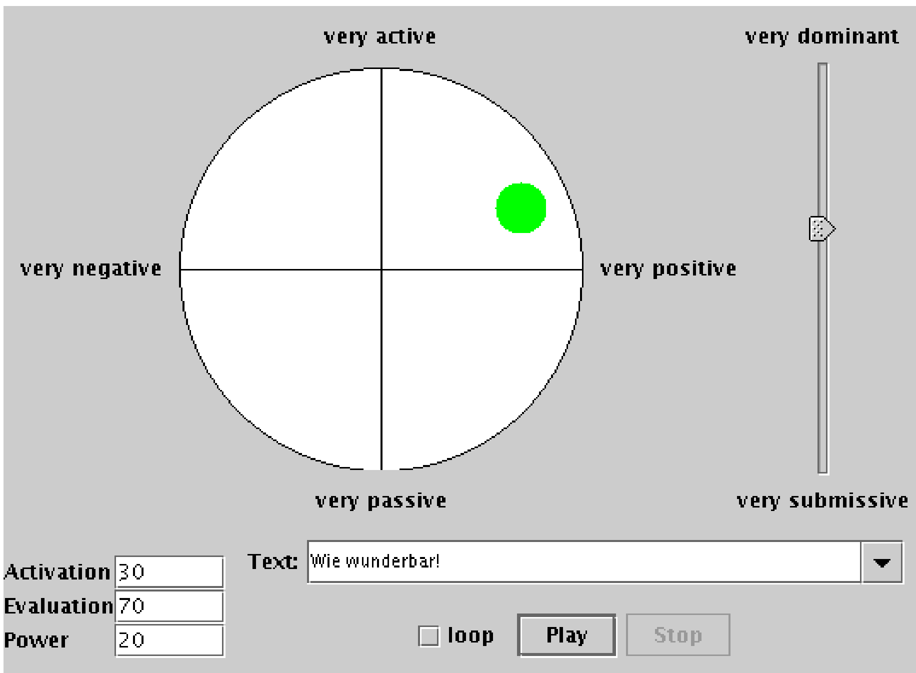


Fig. 1. The EmoSpeak interface to emotional speech synthesis. The system is available online under `http://mary.dfki.de/emotional.html`

The interface, shown in Figure 1, allows the user to type in any text, and to specify the emotional state with which the text is to be spoken. The position on the activation and evaluation dimensions is specified simultaneously, by locating a green cursor in a two-dimensional space modelled after the Feeltrace circle [24]. The third dimension, power, is set independently. Using these positions on the three emotion dimensions as well as the text to be spoken, an EDML emotion markup document is created. The emotion-to-maryxml transformation stylesheet, described above, transforms this into a MaryXML document, which is displayed for the interested user in a second window under the EmoSpeak interface (not shown). Simultaneously, the MaryXML document is sent to the

MARY server, which synthesises a corresponding audio file and sends it back to the interface. All of these transformations are carried out continuously as the user modifies the emotional state or the text. By clicking on the "Play" button, the user can hear the result.

3.3 System evaluation

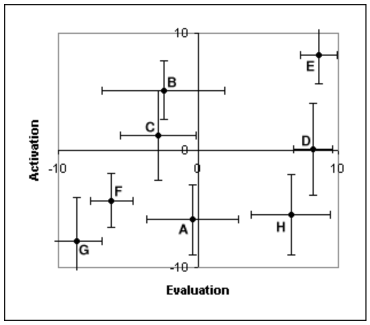
The appropriateness of the generated emotional prosody and voice quality was assessed in a perception test. Due to the multimodal nature of any emotional utterance, this appropriateness is to be thought of in terms of coherence with other channels expressing the emotion, such as verbal content, possibly visual channels, and the situational context. For applications, it is necessary for the speech prosody to "fit with" the rest of the message, in order not to put off the users. This question seems more important than the question what emotional information can be conveyed by the speech prosody alone.

Consequently, any perception test methodology aiming at the assessment of the perceived emotional state is suboptimal for the current task. This includes the identification task methodology widely used for the evaluation of synthesised emotional speech [2], as well as methods for describing the perceived emotional state in terms of emotion dimensions, such as semantic differential ratings [17] or the Feeltrace tool [24]. Instead, a preference task methodology, using coherent and contradictory multi-channel emotional messages as stimuli, was explored as a promising alternative.

In order to create reference material with a known emotional connotation, 36 textual descriptions of emotional situations were extracted from the Belfast Naturalistic Emotion Database [7], translated to German and presented in a written rating task. 15 subjects (mostly students; 4 male, 11 female) rated each situation description on the evaluation and activation dimensions. Eight situation descriptions with high inter-rater agreement (i.e., small standard deviations in activation / evaluation ratings) were retained as relatively unambiguous emotion identifiers. They were well distributed across the activation-evaluation space and included extreme as well as moderate emotional states, see Fig. 2.[1]

For each of the emotional states defined by the situation descriptions, emotional speech prosody settings were calculated, and each of the texts was synthesised with each of the prosodic settings in a factorial design, resulting in 64 stimuli. In a listening test, 20 subjects rated each stimulus according to the question: "How well does the sound of the voice fit with the content of the text?"

[1] The situation descriptions are summarised as follows. A: matter-of-fact description of a painting's history; B: he complains to his ex-wife about wanting to speak to the common son more often; C: he accuses his sister of having lied to him; D: he remembers how he first saw his future wife; E: he is very relieved about not being thrown out of his flat today; F: he talks about vain attempts of a psychiatrist to help his wife; G: he is convinced everybody hates him so he stopped even trying to speak to people; H: a hiker describes the enjoyable peace in the mountains.



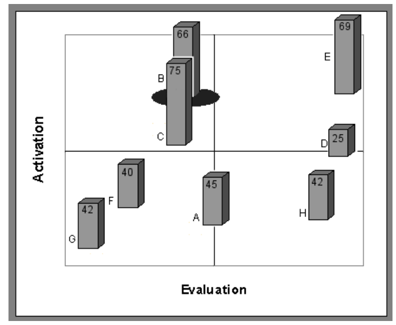


Fig. 2. Co-ordinates of selected situation descriptions in activation-evaluation space. Error bars show one standard deviation in each direction.

Fig. 3. Example of a result plot: Combination of prosody B with all texts. Pillar height represents goodness-of-fit ratings.

The results confirmed the hypothesis that the prosodic configurations succeed best at conveying the activation dimension. Moreover, the appropriateness of a prosodic configuration for a given emotional state was shown to depend on the *degree* of similiarity between the emotional state intended to be expressed by the prosody and that in the textual situation description: There is a highly significant negative correlation between the difference in activation between text and prosody on the one hand and the goodness-of-fit ratings on the other (partial correlation controlling for difference in evaluation: $r = -.331$, one-tailed $p < .001$). In other words, the more different the activation expressed in the text was from that expressed through the prosody, the lower the ratings for their combination (the examplary result plot in Fig. 3 also illustrates this).

In agreement with previous findings for human speech, the evaluation dimension was found to be more difficult to convey through the prosody. Only a very small partial correlation ($r = -.079$, one-tailed $p = .002$) was found between the distance between text and prosody on the evaluation dimension and the goodness-of-fit ratings.

In summary, the speech synthesis system succeeded in expressing the activation dimension (the speaker "arousal"), but not the evaluation dimension. See [1] for a more detailed account of the experiment.

4 Conclusion

This paper has raised the issue of emotion representation in the context of affective dialogue systems. Selected available representations were described and considered in relation to their suitability for various processing components. The possibilities and limitations of mapping between these representations were also discussed.

In more depth, the paper has investigated the principled use of emotion dimensions in the context of speech synthesis, this formalism being well suited for the description of gradual, non-extreme emotional states.

A set of rules for mapping emotion dimensions onto prosodic settings was compiled from the literature and from a database analysis, and implemented in a German text-to-speech system. A perception test confirmed that the system succeeded in expressing the speaker activation or arousal. In line with the literature on human speech, the expression of the evaluation dimension did not succeed. It is an open question whether this is a principal restriction of the speech channel or whether modelling of additional prosodic parameters, as those linked to smiling [25], would improve perception.

Acknowledgements

This research is supported by the EC Projects NECA (IST-2000-28580) and HUMAINE (IST-507422).

References

1. Schröder, M.: Speech and emotion research: An overview of research frameworks and a dimensional approach to emotional speech synthesis. PhD thesis, Institute of Phonetics, Saarland University (to appear)
2. Schröder, M.: Emotional speech synthesis: A review. In: Proceedings of Eurospeech 2001. Volume 1., Aalborg, Denmark (2001) 561–564
3. Cowie, R.: Describing the emotional states expressed in speech. In: Proceedings of the ISCA Workshop on Speech and Emotion, Northern Ireland (2000) 11–18
4. Whissell, C.M.: The dictionary of affect in language. In Plutchik, R., Kellerman, H., eds.: Emotion: Theory, Research, and Experience. Volume 4: The Measurement of Emotions. Academic Press, New York (1989) 113–131
5. Scherer, K.R.: Emotion as a multicomponent process: A model and some cross-cultural data. Review of Personality and Social Psychology **5** (1984) 37–63
6. Cornelius, R.R.: The Science of Emotion. Research and Tradition in the Psychology of Emotion. Prentice-Hall, Upper Saddle River, NJ (1996)
7. Douglas-Cowie, E., Campbell, N., Cowie, R., Roach, P.: Emotional speech: Towards a new generation of databases. Speech Communication Special Issue Speech and Emotion **40** (2003) 33–60
8. Shaver, P., Schwartz, J., Kirson, D., O'Connor, C.: Emotion knowledge: Further exploration of a prototype approach. Journal of Personality and Social Psychology **52** (1987) 1061–1086
9. Cowie, R., Cornelius, R.R.: Describing the emotional states that are expressed in speech. Speech Communication Special Issue on Speech and Emotion **40** (2003) 5–32
10. Cowie, R., Douglas-Cowie, E., Appolloni, B., Taylor, J., Romano, A., Fellenz, W.: What a neural net needs to know about emotion words. In Mastorakis, N., ed.: Computational Intelligence and Applications. World Scientific & Engineering Society Press (1999) 109–114
11. Scherer, K.R.: On the nature and function of emotion: A component process approach. In Scherer, K.R., Ekman, P., eds.: Approaches to emotion. Erlbaum, Hillsdale, NJ (1984) 293–317
12. Scherer, K.R.: Vocal affect expression: A review and a model for future research. Psychological Bulletin **99** (1986) 143–165

13. Banse, R., Scherer, K.R.: Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology **70** (1996) 614–636
14. Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotion. Cambridge University Press, Cambridge, UK (1988)
15. André, E., Klesen, M., Gebhard, P., Allen, S., Rist, T.: Integrating models of personality and emotions into lifelike characters. In: Proceedings of the Workshop on Affect in Interactions – Towards a new Generation of Interfaces, Siena, Italy (1999) 136–149
16. Mehrabian, A., Russell, J.A.: An Approach to Environmental Psychology. MIT Press, Cambridge, MA, USA; London, UK (1974)
17. Osgood, C.E., Suci, G.J., Tannenbaum, P.H.: The measurement of meaning. University of Illinois Press, Urbana, USA (1957)
18. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.: Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine **18** (2001) 32–80
19. Tsapatsoulis, N., Raouzaiou, A., Kollias, S., Cowie, R., Douglas-Cowie, E.: Emotion recognition and synthesis based on MPEG-4 FAPs. In Pandzic, I.S., Forchheimer, R., eds.: MPEG-4 Facial Animation - The standard, implementations, applications. John Wiley & Sons, Hillsdale, NJ, USA (2002)
20. Krenn, B., Pirker, H., Grice, M., Piwek, P., van Deemter, K., Schröder, M., Klesen, M., Gstrein, E.: Generation of multimodal dialogue for net environments. In: Proceedings of Konvens, Saarbrücken, Germany (2002)
21. Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., Gielen, S.: Acoustic correlates of emotion dimensions in view of speech synthesis. In: Proceedings of Eurospeech 2001. Volume 1., Aalborg, Denmark (2001) 87–90
22. Schröder, M., Trouvain, J.: The German text-to-speech synthesis system MARY: A tool for research, development and teaching. International Journal of Speech Technology **6** (2003) 365–377
23. Schröder, M., Grice, M.: Expressing vocal effort in concatenative synthesis. In: Proceedings of the 15th International Conference of Phonetic Sciences, Barcelona, Spain (2003) to appear.
24. Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., Schröder, M.: 'FEELTRACE': An instrument for recording perceived emotion in real time. In: Proceedings of the ISCA Workshop on Speech and Emotion, Northern Ireland (2000) 19–24
25. Tartter, V.C.: Happy talk: Perceptual and acoustic effects of smiling on speech. Perception & Psychophysics **27** (1980) 24–27