Empathic Agent Technology (EAT)

Egon L. van den Broek
Department of Artificial Intelligence
Vrije Universiteit Amsterdam
De Boelelaan 1081a
1081 HV Amsterdam
The Netherlands
+31-20 5987743

Nijmegen Institute for Cognition and Information
Radboud University Nijmegen
P.O. Box 9104
6500 HE Nijmegen
The Netherlands
+31-24 3615476

egon@few.vu.nl
http://www.few.vu.nl/~egon/

ABSTRACT
A new view on empathic agents is introduced, named: Empathic Agent Technology (EAT). It incorporates a speech analysis, which provides an indication for the amount of tension present in people. It is founded on an indirect physiological measure for the amount of experienced stress, defined as the variability of the fundamental frequency of the human voice. A thorough review of literature is provided on which the EAT is founded. In addition, the complete processing line of this measure is introduced. Hence, the first generally applicable, completely automated technique is introduced that enables the development of truly empathic agents.

Categories and Subject Descriptors
I.2.0 [**Artificial Intelligence**]: General—*Cognitive simulation*; I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Speech recognition and synthesis*; I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent agents*

General Terms
Experimentation, Human Factors, Measurement, Theory

Keywords
Empathic Agent Technology, stress, fundamental frequency, pitch, indirect physiological measure

1. INTRODUCTION
The percept or capability to perceive the emotions of other people is of the utmost importance in communication. This holds for human-human communication but is also recognized as important in the field of man-machine interaction (e.g., human-agent interaction) [10]. This feature is most frequently referred to as empathy. Merriam-Webster On-Line dictionary [16] defines empathy as: "the action of understanding, being aware of, being sensitive to, and vicariously experiencing the feelings, thoughts, and experience of another of either the past or present without having the feelings, thoughts, and experience fully communicated in an objectively explicit manner; *also* : the capacity for this." Stotland defines empathy as "an observer reacting emotionally because he perceives that another is experiencing or about to experience an emotion" [26]. Wispé [37] described empathy as: "The process whereby one person feels her/himself into consciousness of another person."

The empathic capabilities vary considerably among people. However, based on the behavior of people, most of us (as 'natural agents') can naturally draw conclusions about their state of mind. If persons are happy, they will walk more straight, smile more often, and will use more intonation in their voice [3]. This basic observation was already noticed and described by Helmholtz [34] more than a century ago. He pointed out that in a joyful or vigorous frame of mind, vowel quality takes on a sharp characterization, whereas in a sad or troubled condition the quality of vowels becomes indifferent and obscure and often changes toward the final phoneme in a word like idea (Helmholtz [33], p. 115).

Non-verbal cues such as bodily and facial expressions play an important role in the assessment of emotional states [8, 9], as do verbal cues, such as the intonation patterns of a person's utterances [3, 38]. These observations, based on a complex set of interrelated factors involving personality assessment, history, and various expressions of emotions, are hard to fit into a protocol.

In the present study, we will evaluate the potential of a more recent and different approach, namely that of using language characteristics (in particular intonation patterns) to assess the state of people's minds, or more specifically, their experienced level of stress. The aim of the study lies in the determination of differences in language use by people in different mood states by means of a voice analytical measure, the fundamental frequency (F0) that is ultimately determined by the muscle tension in a speaker's body [3, 13, 14, 18, 28, 38].

ABSHL/AAMAS'05, July 26, 2005, Utrecht, The Netherlands.

We will investigate whether or not the F0 makes it possible to judge the current emotional state of the person, not by analyzing the content of the story told by the person, as is done most frequently, but by analyzing the characteristics of the speech produced by the person.

Voice analysis, which is a type of phonetic analysis, judges the psychological state of the person not on the basis of subjective measures, such as content analysis, or reported distress, but on the basis of an objective analysis of the speech signal. Contrary to heart rate measures, different sorts of scans and questionnaires (see for example [9]), this measure can be used in parallel with all kinds of activities without disturbing these. All that is needed to obtain this measure is to record the utterances. Furthermore, the analysis of the collected data does not have to take much time, because it can be completely automatized [29, 30]. The incorporation of such a measure, would enable the agent to perceive emotions and act more properly, similar to how humans do this.

In this paper, we will discuss whether this voice analytical measure is applicable for efficient man-machine interaction by agents, which we name: Empathic Agent Technology (EAT). The basic assumption underlying our study is that the characteristics of spoken utterances reflect the stress that accompanies internal emotions [14, 28]. When we manipulate the emotional state of a person, the voice measure should change. This link between vocal characteristics and internal emotions will now be clarified by examining the nature of stress and its relationship to emotion and voice.

In the next section, we will first provide some theoretical background, using stress as central concept. In Section 3, we provide the architecture and the means of the EAT. Section 4 discusses the relevance and the embedding of EAT in an agent framework. Next, we discuss both advantages and disadvantages of this approach, in Section 5. Last, we draw some final conclusions.

2. THEORETICAL BACKGROUND

2.1 Defining stress

The wide use of the term stress, has resulted in a multitude of different definitions as well as diverse theoretical notions concerning the underlying processes [22]. This is due to its expression on a large variety of levels, psychological and physiological.

The complex physical reaction in stress consists of an episode of temporary synchronization of all major systems of an organism (cognition, physiological regulation, motivation, motor expression, and monitoring/feeling) in response to the evaluation of an external or internal stimulus event as relevant to central concerns of the organism.

During non-emotional states the different subsystems or components operate relatively independently of each other, each dealing with its respective function in overall behavioral regulation. However, during emergency situations (i.e., emotion episodes), these systems are recruited to work in unison. Their synchronous collaboration requires the mobilization of substantial organismic resources, physiological regulatory responses to threats to the organism, in order to allow adaptation or active responses to an important event or change of the internal state.

An emotion episode begins with the onset of subsystem synchronization following a particular stimulus evaluation pattern and ends with a return to an independent functioning of the subsystems. Most theories assume that these appraisal processes will evoke one of several discrete emotions with their specific response patterning. Such response patterns disturb homeostatic regulation, which controls the different but interdependent elements or groups of elements of the human physiological system [2]. These complex processes are, however, not of central concern in this study. We will, therefore, use a more widely used definition of stress. We define stress in general terms as an unusual environmental condition that causes physiological, emotional, behavioral, or cognitive changes in an individual or as the changes or effects resulting from such environmental demands.

2.2 The voice

The voice results from a source in the larynx, filtered by the supra-laryngeal vocal tract. The periodic opening and closing of the vocal folds in the larynx produces a buzzing sound with a characteristic waveform, which determines the fundamental frequency (F0) or pitch of the voice [4].

The voice, as the face, is characterized by a unique combination of physical features related to the unique configuration of the human vocal apparatus. As with faces, inter- and intra-individual variations around this generic structure can carry useful information. Three categories of information can be distinguished: speech information, identity information, and affective information [4]. In the current research, we will focus on the latter source of information, which is directly influenced by the speaker's affective state. In the next subsection, we will discuss this relation in more detail.

2.3 Stress and the voice

On theoretical grounds [21], and supported by several pertinent findings in the literature, mood changes are expected to lead to elevated muscle tone and they will, therefore, affect characteristics of the speech of the person involved. The complexity of the relationship between mood change and voice becomes evident if one thinks about Lenneberg's [13] remark that speech is an act of a large motor complexity, requiring the control and the coordinated activation of over 100 muscles. In correspondence with this view, relations between the vocal fundamental frequency (F0) and one particular and very important muscle, the heart, are reported, see also Figure 1.

The F0 of a sustained vowel was found to be modulated over a time period equal to that of the speaker's heart cycle, illustrating its ability to express one's physiological condition. In the next section, we will describe how F0 can be used as a measure for stress.

2.4 Stress and F0

Assuming that speech can be divided into a number of functionally independent speech dimensions, the constituents of speech are: fundamental frequency (F0) or pitch, loudness or amplitude/energy, spectral characteristics, and timbre.

In the present paper, we will focus on the F0: The lowest audible frequency in a complex sound signal, such as, in our case, human speech. The F0 usually carries more energy than the other harmonics and identifies the pitch of the tone to the listener.

According to several theories (e.g., [17, 19, 23]), increased stress should result in an increase in F0, and in a decrease of F0 range and variability due to more rigidity of the phonation mechanism under high muscular tension. During recovery from stress, F0 should decrease again and its range and variability should increase (e.g., [19, 21]).

The most obvious measure to assess stress is probably the mean F0 [3, 17, 19, 23]. However, mean F0 does not covary with stress manipulation [28]. To choose the best F0-related measure for stress we must, therefore, take a closer look at the different aspects of the fundamental frequency (F0) of the human voice as a mirror of tension, in particular muscle tension. Other previously used measures involving the F0 are F0 floor and F0 variation. The F0 floor consists of the lowest F0 values in utterances (except for questions). An estimate of F0 floor is obtained by taking the frequency values at the 5% level of the F0.

Tolkmitt and Scherer [28] found that, depending on the coping style of persons, F0 floor increases or decreases under high stress. F0 floor is closer to the purely physiological contribution to phonation and may therefore be a better stress indicator than, for example, mean F0. There is, however, even more evidence for F0 variation as a major carrier of prosodic information and as the main factor responsible for the expressiveness of speech [12, 24].

The variability of the F0 signal can be computed by the standard deviation (SD) of F0 or by determining the difference between two quantiles of F0. Undoubtedly, the SD is the measure for data variability.

2.5 F0 and individual physiological characteristics

Because of its dependence on physiological and physical characteristics, F0 is also affected by a number of person oriented factors such as gender and inter-individual differences with respect to the vocal tract. Let us briefly discuss each of these in turn.

With respect to gender, F0s are usually reported to be approximately 200 Hz and 120 Hz, for female speakers and male speakers, respectively. In addition, previous research [28] found that males do not react as strongly physiologically to stress as females do. Scherer, Banse, Wallbott, and Goldbeck [24] note that future studies need to pay much more attention to potential gender differences.

Because of its physiological basis, F0 does not only depend on sex, but also on other inter-individual differences. With respect to inter-individual characteristics, a long-standing problem in emotion research in general is that different indicators of emotion verbal report, nonverbal behavior, autonomic arousal often do not inter-correlate in an emotional

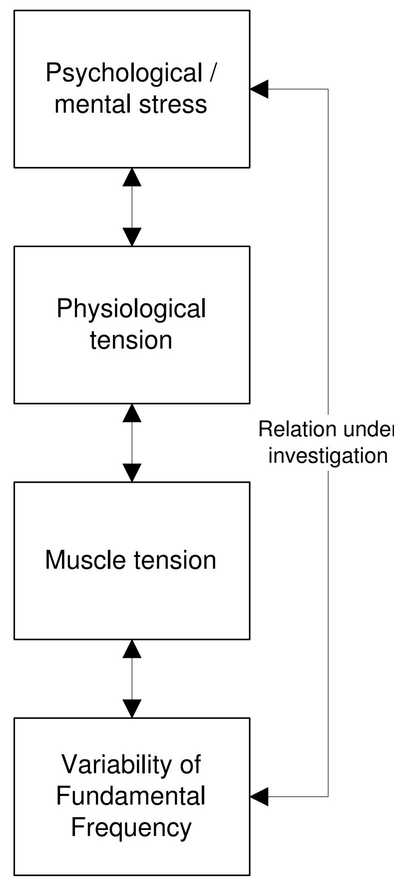

Figure 1: Schematic overview of the relation between prosody and experienced stress.

situation. On the one hand, this lack of correlation led many researchers to believe that the concept of emotion as a theoretical construct defined by empirical indicators is a fiction and has little predictive power. On the other hand, from a system-theoretic point of view, discrepancies between different control systems of homeostatic regulation can serve as an important source of information about the control processes underlying the observed behavior patterns [2]. Therefore, comparing the emotional behavior of one person in different situations is the most straight-forward approach to systematically examine inter-individual differences in emotions, using F0 as a measure. In other words, this view stresses the importance of within subject research, an approach also advocated by Lacey [11]. Derived from his notion of specificity , the concept of emotion is differentiated into the following three subcomponents:

1. Person specificity: Different people react differently to the same stressors. The physiology of the human body stays relatively constant across time, and therefore a person's unique pattern of reactions in the diverse modalities, such as speech, movement, and e.g., sweating, is consistent across time and situations.

2. Stimulus or situation specificity: Different reactions are given to different stimuli or situations, because each situation is unique by definition.

3. Emotion specificity: Cognitive versus emotional stress. Cognitive stress is defined as the information processing load placed on the human operator while performing a particular task [14]. It can be assumed that cognitive stress increases with task difficulty. Emotional stress is the psychological and physiological arousal due to emotions triggered before or during a task. In other words, a distinction is made between a high mental workload and a high emotional arousal. A deadline at work triggers another form of stress than a fight with your partner at home, without implying that one of them is not an intensive form of stress.

3. TECHNOLOGY

In the Introduction, the potential benefit of EAT for human-agent interaction was discussed. In Section 2, the variability of F0 (SD F0) was introduced as a general measure for stress. However, with EAT we stated to introduce a technology that can be embedded in artificial agents to show, or at least mimic, empathy. In this section, we will briefly explain the processing pipeline (see Figure 2) that operationalized the relation under investigation, as depicted in Figure 1.

The recordings can be done by a standard personal computer and a microphone. On the one hand, these recordings can be saved on the hard-disk of a computer and can incrementally update a database of (labeled) recordings. On the other hand, samples has to be processed directly to enable real time empathic reactions of agents.

Let us now provide some technical specifications. As default, a sample rate for the recordings of 44.1 kHz, mono-channel, with a resolution of 16-bit, can be used. The standard windows sound recorder or a program such as Cool Edit Pro 1.0® [27] can be used for the recording.

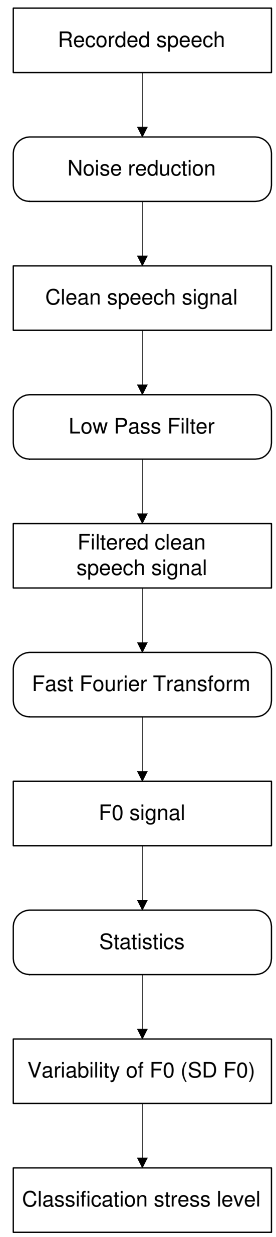

Figure 2: The processing pipeline of the speech analysis program. Rounded boxes indicate processing phases. The rectangular boxes denote the (preliminary) output of the several processing phases.

The F0 of stretches of speech can be easily determined, using scripts for the program Praat 3.9.0® [5]. The program's scripting features provide the means to determine the mean, median, standard deviation (SD), and the delta quantile of 95%-5% and of 90%-10% [29, 30], see Figure 2 and 3.

Please note that the scripting utility of Praat 3.9.0® can be used for rapid testing and prototyping. However, as most scripting languages, it impairs the processing speed. Hence, to enable true real time processing, the complete processing scheme has to be programmed in, for example, C. In the latter programming language Praat 3.9.0® itself is also built. Since it is an open source program, its components can be reused.

4. EMBEDDING

In the previous sections, the processing scheme is introduced for a speech analysis module (EAT). However, why is such a module of interest in human-agent communication? On the one hand, ubiquitous computing is promoted often referring to invisible computer systems; on the other hand, some researchers advocate direct, almost natural, human-like interaction between humans and computer devices; e.g., agents. See Shneiderman and Maes [25] for a discussion on this topic. The EAT module can be applied in both forms of communication, since for voice analysis no visual input is required nor any other direct physiological interaction is strictly needed.

4.1 Agents and Emotions

Various researchers have studied the (potential) relationships between humans and artificial agents. Reeves and Nass [20] provided a variety of evidence that assents the social and emotional involvement of people who are in interaction with various forms of media. The latter claim was sustained by Tomlinson [35] who proposed a mechanism based on emotion for agents to interact with each other and with people. In addition, Friedman, Kahn, and Hagman [7] showed that people are willing to get involved in relations with robots.

Recently, Arbib and Fellous [1] stated that multi-agent teamwork in virtual organizations is improved when agents are able to stimulate emotions since they would be more believable to humans. Moreover, Arbib and Fellous [1] show that: "emotions operate at many time scales and at many behavioral levels" and that "there is no easy separation between emotion and cognition". Hence, emotions are of the utmost importance in communication, both between humans and between artificial agents and humans.

4.2 Levels of Information Processing

In human-agent interaction, three levels of information processing have to be taken into account [1]. For each of these levels, emotion is a factor of importance. These three levels are:

1. a reactive system with fixed action patterns
2. a routine-based system, unconscious well-learned automatized activity
3. a reflective system of higher-order cognitive functions, including meta-cognition, consciousness, self-reflection, and full fledged emotions

But how to incorporate the EAT module in an agent framework, which uses one or multiple of these levels of information processing?

The EAT, as introduced in the previous sections, can be incorporated in the architecture of an artificial agent. For example, the synthetic character of the VICTEC empathic agent [32] can be enriched, using the EAT module. The EAT module can record speech, perceive its fundamental frequency, and classify the emotion (see also Figure 2). In the next subsection, we will discuss the incorporation of EAT in an agent such as VICTEC.

4.3 EAT within an agent

First, we will sketch the relation between an agent (e.g., the VICTEC agent) and EAT in general. Next, the implications for each of the three levels of information processing will be discussed.

The VICTEC agent perceives speech and consequently, perceives emotion, see Figure 4.The agent incorporates a database of emotions, with their expressions. In its memory, the agent stores user templates, which can consist of semantic classifications of behavior and their accompanying emotional expressions. The emotional expressions that are of importance for EAT are the voice recordings. Given a baseline of the voice of users, the relative intensity of verbal utterances can be determined, using the SD F0.

A reactive agent can utilize the extracted SD F0 as follows. Given an utterance that reveals tension, it can adapt its own tone of voice so that it sounds more nice. In the case, the voice reveals no high tension, it does not has to adapt its utterances.

An agent that incorporates some routines can utilize more complex behavior. It can, for example, construct adapted phrases in combination with an adaption of tone of voice.

A true reflective agent takes the environment into account, providing that it is able to utilize an ontology. It can react on the user's utterance in a way that is both suitable from a semantic and syntactic point of view, using an appropriate tone of voice. Moreover, it can use relevance feedback (e.g., through the voice) to adapt its ontology and consequently, adapt its behavior.

5. DISCUSSION

This study examined the applicability of the variability of the fundamental frequency of the voice (SD F0) as a measure for experienced tension. Based on existing empirical literature, it was assumed that people's stress triggers physiological tension expressed by a (more) constant muscle tension. This (more) constant muscle tension is accompanied by a lower variability in muscle tension, which will in turn lead to a lower variability of the fundamental frequency of the human voice [13, 22, 23, 28].

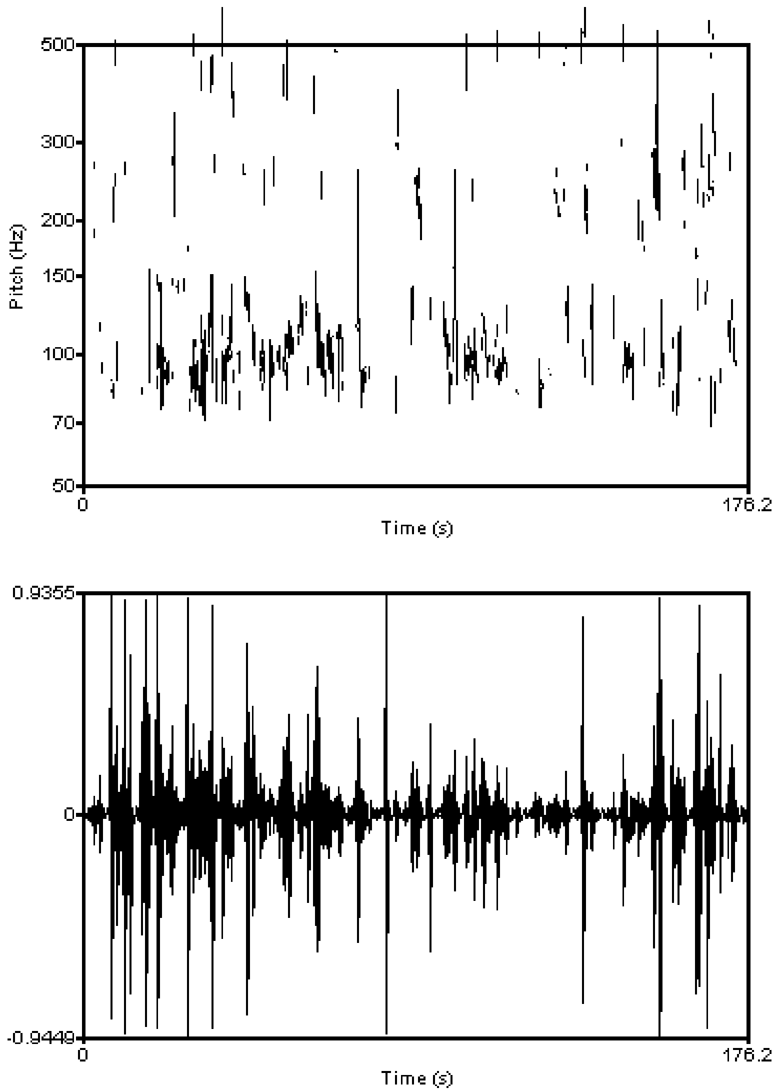

Figure 3: A recorded speech signal (bottom) and its pitch or fundamental frequency (F0) (top). See Figure 2 for the phase of processing of both visualizations.

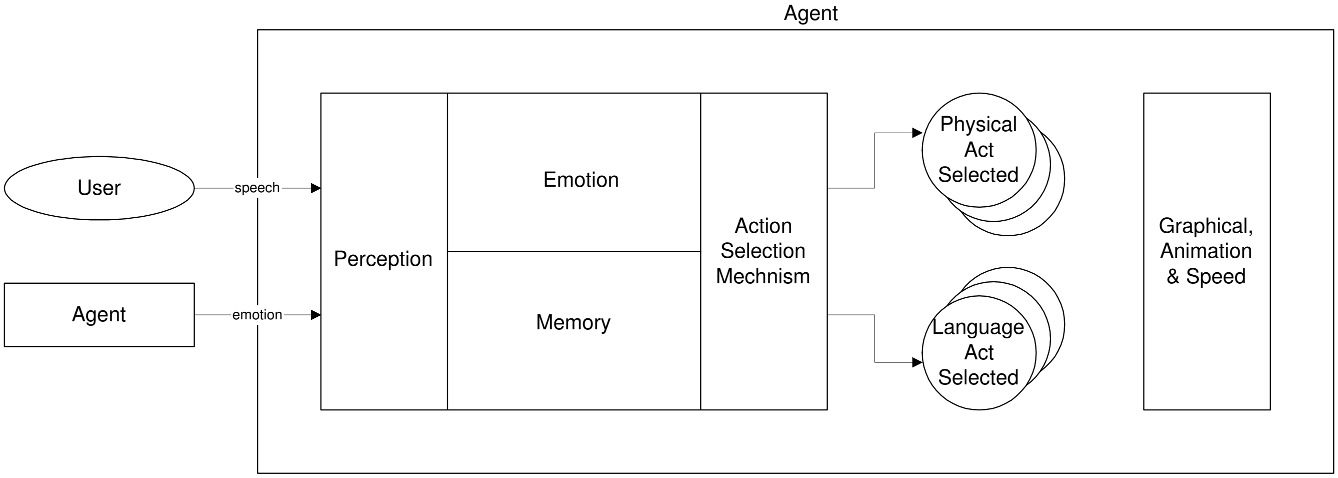

Figure 4: The overall agent architecture of the VICTEC empathic agent; adopted from [31].

A potential problem with the EAT, as introduced, is described by the existing theoretical distinction of emotional and emotive communication [15]. Emotional communication is a type of spontaneous, unintentional leakage or bursting out of emotion in speech, while emotive communication has no automatic or necessary relation to "real" inner affective states. Emotive communication is a strategic signaling of affective information in speaking to interaction partners that is widespread in interactions (see also [6]). It uses signal patterns that differ strongly from spontaneous, emotional expressions and can be both intentionally and unintentionally accessed [3]. It may be the difference between emotional and emotive expressions is reflected in the difference in variability of the F0.

Another issue is the distinction between cognitive and emotional stress. This problem is also known as the problem of emotion specificity [11]. Emotion specificity distinguishes cognitive stress, the information processing load placed on the human operator while performing a particular task [14], and emotional stress, the psychological and physiological tension due to emotions triggered before or during the task.

Despite these two drawbacks, an increasing interest is present in an indirect physiological measure that can be integrated in agent systems, such as EAT. For example, in combat situations emotions play a crucial role. In training environments for combat situations, the use of agents is already under investigation [31]. Agents can play the role of colleges or of enemies. In combat situations, the emotional experiences of the people involved are intense. Empathic team members can mediate in the functioning of a group and, subsequently, improve the group performance, where agents showing no empathy at all can frustrate the team process.

6. CONCLUSION

Based on our findings, we suggest that future research should focus on the development of a theoretical framework for EAT, based on speech analysis [8, 36]. Such a framework should be flexible enough to allow an incorporation of the characteristics of each situation and individual, using several objective (physiological) measures (such as the SD F0) of psychological tension. It would make it possible to test theoretical findings and adapt the theoretical framework on the basis of new empirical results from clinical research. In the last decade, several varieties of human behavior, were already modeled, for example, by means of neural networks.

The incorporation of direct physiological measures (e.g., EMG and EEG) are often applied; however, not easy from practical, technical, and financial points of view. Furthermore, people feel uncomfortable when they are connected to all kinds of machinery. EAT, on the other hand, does not lead to discomfort and is suitable for human-agent interaction. It requires only a small endeavor in the development of agents to include EAT, which provides the agent with a powerful additional feature: empathy.

The increasing attention toward speech-driven human-agent communication illustrates the importance of EAT. The measure proposed in this paper, the variability of the fundamental frequency of the voice (SD F0), is an easy to integrate, automatized, indirect, physiological measure. We have argued that EAT is useful for measuring the (intensity of) experienced emotions in people. As such it can, for example, be a useful module for agents, providing it with empathy. With that an important step is made toward true empathic agents.

7. ACKNOWLEDGMENTS

Paul Boersma is gratefully acknowledged for the answers on all the questions fired at him. We thank Ton Dijkstra for his support at the first phase of the project and Henriette C. van Vugt for providing some useful pointers to existing research. In addition, we would like to thank the anonymous reviewers for their constructive criticism.

8. REFERENCES

[1] M. A. Arbib and J.-M. Fellous. Emotions: from brain to robot. *Trends in Cognitive Sciences*, 8(12):554–561, 2004.

[2] J. B. Asendorpf and K. R. Scherer. The discrepant repressor: Differentiation between low anxiety, high

anxiety, and repression of anxiety by autonomic-facial-verbal patterns of behavior. *Journal of Personality and Social Psychology*, 45:1334–1346, 1983.

[3] R. Banse and K. R. Scherer. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70:614–636, 1996.

[4] P. Berlin, S. Fecteau, and C. Bédard. Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, 8(3):129–135, 2004.

[5] P. P. G. Boersma and D. J. M. Weenink. Praat 3.9.0, 2001. URL: `http://www.praat.org` [Last accesed on May 17, 2005].

[6] C. Caffi and R. W. Janney. Toward a pragmatics of emotive communication. *Journal of Pragmatics*, 22:325–373, 1994.

[7] B. Friedman, J. P. H. Kahn, and J. Hagman. Hardware comparisons?: What online AIBO discussion forums reveal about the human-robotic relationship. In G. Cockton and P. Korhonen, editors, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 273–280, Ft. Lauderdale, Florida, USA, 2003. New York, NY, USA: ACM Press.

[8] S. Kaiser and K. R. Scherer. *Models of 'normal' emotions applied to facial and vocal expressions in clinical disorders*, pages 81–98. New York: Oxford University Press, 1998.

[9] S. Knapp and L. VandeCreek. *Anxiety disorders: A scientific approach for selecting the most effective treatment*. Sarasota, FL, US: Professional Resource Press/Professional Resource Exchange, Inc, 1994.

[10] E. A. Konijn and J. F. Hoorn. Some like it bad. Testing a model for perceiving and experiencing fictional characters. *Media Psychology*, 7(2):107–144, 2004.

[11] J. I. Lacey. *Somatic response patterning and stress: Some revisions of activation theory*, pages 14–42. New York: Appleton-Century-Crofts, 1967.

[12] D. Ladd, K. Silverman, F. Tolkmitt, G. Bergmann, and K. R. Scherer. Evidence for the independent function of intonation contour type, voice quality, and f0 range in signalling speaker affect. *Journal of the Acoustical Society of America*, 78:435–444, 1985.

[13] E. H. Lenneberg. *Biological foundation of language*. New York: Wiley & Sons, 1967.

[14] S. E. Lively, D. B. Pisoni, W. van Summers, and R. H. Bernacki. Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *Journal of the Acoustical Society of America*, 93:2962–2973, 1993.

[15] A. Marty. *Untersuchungen zur allgemeinen grundlegung der grammatik und sprachphilosophie.* Halle/Saale, Germany: Niemeyer, 1908. Translation of title: Investigations on the general foundations of grammar and the philosophy of language.

[16] Merriam-Webster, Incorporated. Merriam-Webster Online. URL: `http://www.m-w.com/`, [Last accessed on May 18, 2005].

[17] I. R. Murray and J. L. Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, 93:1097–1108, 1993.

[18] A. Nilsonne, J. Sundberg, S. Ternström, and A. Askenfelt. Measuring the rate of change of voice fundamental frequency in fluent speech during mental depression. *Journal of the Acoustical Society of America*, 83:716–728, 1988.

[19] J. Pittam and K. R. Scherer. *Vocal expression and communication of emotion*, pages 185–197. New York: Guilford, 1993.

[20] B. Reeves and C. Nass. *The media equation*. Cambridge: Cambridge University Press, 1996.

[21] K. R. Scherer. *Personality markers in speech*, pages 147–209. Cambridge, UK: Cambridge University Press, 1979.

[22] K. R. Scherer. *Vocal indicators of stress*, pages 171–187. New York: Grune & Stratton, Inc, 1981.

[23] K. R. Scherer. Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99:143–165, 1986.

[24] K. R. Scherer, R. Banse, H. G. Wallbott, and T. Goldbeck. Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, 15:123–148, 1991.

[25] B. Shneiderman and P. Maes. Direct manipulation vs. interface agents. *Interactions*, 4(6):42–61, 1997.

[26] E. Stotland, K. Mathews, S. Sherman, R. Hanson, and B. Richardson. *Empathy, fantasy, and helping*. Beverly Hills: Sage Publications, 1978.

[27] Syntrillium Software. Cool Edit Pro 1.0 (now Adobe Audition), 1998. URL: `http://www.cooledit.com` [Last accesed on May 17, 2005].

[28] F. J. Tolkmitt and K. R. Scherer. Effect of experimentally induced stress on vocal parameters. *Journal of Experimental Psychology: Human Perception and Performance*, 12:302–313, 1986.

[29] E. L. van den Broek. Diagnosis of emotional state using pitch analysis and sud: An explorative feasibility study. M.Sc. Thesis, Department of Artificial Intelligence, Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, June 2001.

[30] E. L. van den Broek. *Emotional Prosody Measurement (EPM): A voice-based evaluation method for psychological therapy effectiveness*, volume 103 of *Studies in Health Technology and Informatics*, pages 118–125. IOS Press - Amsterdam, The Netherlands, 2004.

[31] W. A. van Doesburg, A. Heuvelink, and E. L. van den Broek. TACOP: A cognitive agent for a naval training simulation environment. In M. Pechoucek, D. Steiner, and S. Thompson, editors, *Proceedings of the Industry Track of the Fourth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-05)*, Utrecht - The Netherlands, [in press].

[32] VICTEC. Virtual ICT with Empathic Characters (VICTEC), URL: `http://www.VICTEC.org` [Last accessed on May 15, 2005].

[33] H. von Helmholtz. *Handbuch der physiologischen Optik*, volume 2. Leopold Voss: Leipzig, Germany, 1867. 286–287.

[34] H. von Helmholtz. *Die lerne van den Tonempfindungen als physiologische grundlage fur die theorie der musik*. Braunschweig: Friedrich Vieweg und sohn, 5 edition, 1896. Original work published in 1864.

[35] J. W. M. Tomlinson. *Synthetic Social Relationships*. PhD thesis, Program in Media Arts and Sciences, Program in Media Arts and Sciences, Massachusetts Institute of Technology, 2002.

[36] T. Wehrle and K. R. Scherer. *Towards computational modeling of appraisal theories*, pages 350–365. New York: Oxford University Press, 2001.

[37] L. Wispé. *History of the concept of empathy*, pages 17–35. New York: Cambridge University Press, 1987.

[38] B. Zei Pollerman. Measurement of emotional involvement in spontaneous speaking behavior. In *Proceedings of Measuring Behavior 2000 3rd International Conference on Methods and Techniques in Behavioral Research*, pages 357–358, Nijmegen, The Netherlands, 2000.