EUROPEAN ORGANISATION
FOR THE SAFETY OF AIR NAVIGATION

EUROCONTROL EXPERIMENTAL CENTRE

EVALUATION OF THE HUMAN VOICE FOR INDICATIONS OF WORKLOAD INDUCED STRESS IN THE AVIATION ENVIRONMENT

EEC Note No. 18/06

INO-1 AC STSS Project

Published: December 2006

REPORT DOCUMENTATION PAGE

Reference: EEC Note No. 18/06	**Security Classification:** Unclassified
Originator: **EEC - INO** (**INO**vative Research Area)	**Originator (Corporate Author) Name/Location:** EUROCONTROL Experimental Centre Centre de Bois des Bordes B.P.15 F - 91222 Brétigny-sur-Orge CEDEX FRANCE Telephone: +33 (0)1 69 88 75 00
Sponsor: EUROCONTROL Experimental Centre	**Sponsor (Contract Authority) Name/Location:** EUROCONTROL Experimental Centre Centre de Bois des Bordes B.P.15 F - 91222 Brétigny-sur-Orge CEDEX FRANCE Telephone: +33 (0)1 69 88 75 00 **WEB Site:** www.eurocontrol.int

TITLE:
Evaluation of the Human Voice for Indications of Workload Induced Stress in the Aviation Environment

Authors	Date	Pages	Figures	Tables	Annex	References
M. Hagmueller, E.Rank, G. Kubin (Graz University of Technology)	08/2005	81; +VI	17	7	4	100

	Project INO-1 AC STSS	Task No. Sponsor	Period 2005

Distribution Statement:
(a) Controlled by: Vu DUONG Head of INO
(b) Special Limitations: None

Descriptors (keywords):
Stress, workload, air traffic control, speech analysis, voice analysis, contact free, non-intrusive

Abstract:
This report gives a literature review of stress analysis from the human voice. While a special focus is on workload-induced stress and the special treatment of the avionic environment, a general overview of the analysis of speech under stress using a broader definition of stress is given as well. We discuss definitions of stress and workload and point out the relation between stress and human speech. We provide a summary of useful speech databases for workload induced stress analysis. A discussion on speech signal features commonly used for stress analysis is the main part of this report.
Research in the effect of workload-induced stress on speech signals has become a riveting issue recently, while it is still at an early stage. Current problems are the multitude of influence factors and the speaker dependence of the effects of stress on the speech signal.
As speech has the advantage of being a medium naturally used in the air traffic control environment it seems to be worthwhile to further explore this approach on a long-term basis.

NOTE: The methods discussed in this report are largely based on public available research literature. The ownership of the related intellectual property rights has not been determined in this literature study and should be sought directly from the authors of the publications cited in the 'references' sections.

PAGE INTENTIONALLY LEFT BLANK

Preface

Air Traffic Control (ATC) operators on duty have a large responsibility for the life of passengers and crews in the aircraft. High vigilance and performance are continually required. Human's features vary in time, with health situation and by external events. A scientific principle developed by the psychologists R.M. Yerkes and J.D. Dodson in 1908, known as the Yerkes-Dodson law demonstrates an empirical relation between arousal and performance. Thereafter human performance increases with cognitive arousal until a certain point. After this point further increasing arousal results in a decreasing performance.

In the ATC (Air Traffic Control) task good performance can be seen as the absence of unsafe actions caused by the human. Therewith the human part of the global ATC safety becomes proportional to the performance of the ATC operator. So, for an optimal safety of the human ATC task, an operator ideally needs moderate workload. Therefore it is common ATC practice to modulate the size of a control sector during the day, depending of the traffic load. With low traffic two or more sectors may be collapsed into one control unit with one controller, while with increasing traffic this unit is split again into the smaller sectors again, with one controller each. The aim is to hold the workload for the controller continuously at a moderate level to favour due to Yerkes-Dodson's law ATC safety.

Currently a supervisor of the control room decides on the so-called re-sectorisation. His/her decision is subjective. It is based on own experience, own assessment of a moderate workload and administrative constraints. Especially the assessment of a moderate workload for an individual is very difficult as it varies strongly by environmental stimuli and state of health. A real-time tool to evaluate objectively human stress indicators under a given workload, to keep the human at the optimal performance, could help to increase ATC safety.

In the medical sense workload can be seen as an external stimuli, which cause physiological stress reactions of the human body. Stress modifies the levels of specific hormones, which in turn alter heart rate, respiration, blood pressure and transpiration. So, measuring these primary or secondary stress reactions could give objective information how this individual sense this workload. This could be an objective base for supervisor's re-sectoring decisions. Off-line and under laboratory conditions such measurements have been undertaken, but they are not adapted to field studies nor to real working conditions.

Hypothetically the human voice could be used as carrier to retrieve most of human secondary stress reactions without any physical contact to the subject. Human vocal folds are directly in touch with respiration and heart activity via the blood. Consequently the voice should be slightly modified thereof and could be used as indicator of human's stress reaction.

Based on recorded ATC speech samples we tried to investigate this hypothesis. In 1998 the EUROCONTROL Experimental Centre (EEC) initiated the 'Study into the Detection of Stress in Air Traffic Control Speech'. The study was executed by the 'Speech Research' group of the British 'Defence Evaluation Research Agency (DERA) in Malvern. The study was limited to detect higher stress and doesn't show a strong correlation.

Meanwhile speech research progressed and so we revisited the issue in form of a literature study. This note reports on the study 'Stress and Speech – State of the Art Report 2005' of the 'Signal and Speech Processing Laboratory' of the Graz University of Technology.

The documented research results can be seen as promising. But the study shows as well, that especially the detection of decreasing performance under low workload conditions is requiring further basic research. Other visual analysis methods, of eyes and face, are proposed.

Horst Hering

Innovative Research Unit

TABLE OF CONTENTS

PAGE INTENTIONALLY LEFT BLANK

1. Introduction

Stress is the physiological and psychological answer of a human body to a specific workload. Stress influences human organ functions. Measuring these indications of stress is possible under laboratory conditions. Field studies under normal working conditions are not practical and some measures cannot be obtained in real-time.

In the avionic domain, optimal safety is directly related to the correct level of workload for the human operator. Underload and overload increase the human error rate. This optimal workload level is different for every human being and may vary with the daily condition. Knowledge about workload induced stress of the human operators could help to keep them at a desirable level of workload for optimal safety.

1.1. Overview

This study focuses on stress analysis from the human voice and presents a literature review of recent investigations on this topic. We start in this chapter with a short introduction of the air traffic control environment. In chapter 2 some relevant background knowledge on stress and speech signals is given. Important terms are defined and relations between human speech and stress are explained. Emotion recognition from the human voice is also mentioned, since emotions are closely related to stress. Chapter 3 provides information on speech databases which include speech under stress and also points to sources where those databases are available.

The central part of this document is chapter 4, where we present a literature review of work on analysis and classification of speech under stress. As far as possible, the focus is put on workload induced stress and the aviation environment. Since speech under stress influences voice based human-computer interaction, chapter 5 provides a short summary on implications of speech under stress, for e.g., automatic speech recognition. A summary and the evaluation of the findings with recommendations for further proceeding conclude the main part of this document (chapter 6).

In the appendix A non-speech methods for workload induced stress are briefly introduced. In appendix B a reference of the used acronyms is provided.

1.2. Air Traffic Control Environment

In this section we will briefly introduce the air traffic control environment and point out sources of stress. Additionally, we will shortly discuss some stress relevant issues at the workplace in general.

1.2.1. Main sources of stress for air traffic control

Stress sources for an air traffic control operator (ATCO) are manifold. Some directly depend on the current demand of the work (taskload), others on the working conditions and the general life situation. A list of possible sources of stress for ATCOs is provided below (from [Costa, 1995]).

Demand:

- number of aircraft under control
- peak traffic hours
- extraneous traffic
- unforeseeable events

Operating procedures:

- time pressure
- having to bend the rules
- feeling of loss of control
- fear of consequences of errors

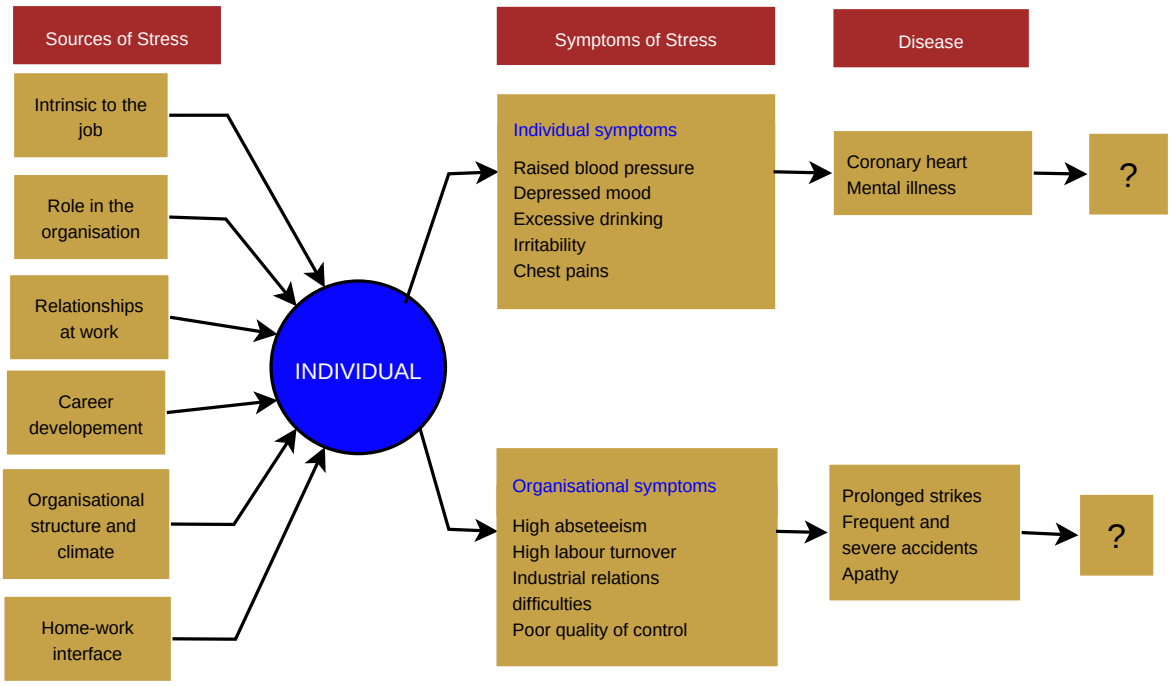


Figure 1: *Stress in the context of workplace and some long-term effects on individuals and organizations (from [Cox et al., 2000]).*

Working tools:

- limitations and reliability of equipment
- video display terminal (VDT), radio telephony (R/T) and telephone quality
- equipment layout

Working times:

- unbroken duty periods
- shift and night work

Work environment:

- lighting, optical reflections
- noise/distracters
- micro climate
- bad posture
- rest and canteen facilities

Work organization:

- role ambiguity
- relations with supervisors and colleagues
- lack of control over work process
- salary
- public opinion

Non-workplace related :

- private life conditions
- relationships
- amount of sleep
- personal capacity
- emotional conditions

1.2.2. Long-term effects of stress

In [Cox et al., 2000] a discussion on long-term effects of stress in the context of a workplace is discussed. Besides short-term consequences of stress, such as operation errors, the long-term effects of stress can have serious consequences, both on the health of the individual and also on an organization (see Fig. 1). Again workload induced stress is just one parameter in this context.

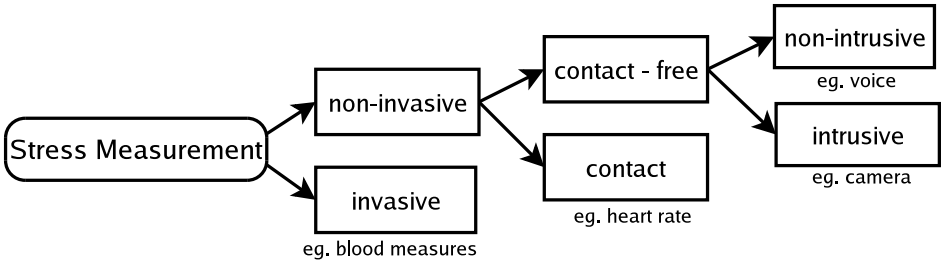


Figure 2: *Types of stress measurement and differences concerning intrusiveness.*

1.3. Types of Measurement

Concerning the measures used to serve for the analysis of a person's stress level or type, we may distinguish between different ways of how features for analysis are obtained (see fig. 2).

Probably the most accurate estimations of a subject's stress level can be obtained by measuring physiological parameters. Those measurements may imply, for example, taking blood tests or even insertion of electrodes into the brain. Many of these measures are *invasive* measures that comprise a kind of 'medical examination'.

Non-invasive measures, on the contrary, are attained by techniques in which the body is not entered by puncture or incision. Still, one can differentiate between measures obtained with *physical contact*, such as when using a wrist-band sensor for measuring the heart-rate or electrodes applied on the skin for electroencephalogram (EEG) measurements, and *contact-free* measurements where no physical contact between the subject and the measuring device is necessary.

Contact-free measurements are clearly favorable for stress analysis in real-world situations, since the application of sensors on a subjects body – or inside the body – can impair the subjects working capabilities, or even be a cause of stress by itself. However, even without physical contact, the test subject may feel intimidated by a measuring device, such as a camera positioned in front of her. Hence, for contact-free measures we may again define two sub-groups and differentiate between *intrusive* measures that interfere with the subject's privacy (such as a camera) or the operator's job performance, and *non-intrusive* measures, which would be, for example, analysis of the speech signal, in case the person under observation is using her voice and a microphone as the primary means for working anyway.

Methods to assess the subjective workload – for example questionnaires to be completed by the test subject or behavioural measures, such as performance measures – which try to relate the results of the work to the workload have both their advantages and problems, but are not covered in this review. For a review and a discussion of measures for air traffic control (ATC) workload, including task-load measures, behavioural measures and psycho-physiological measures see [Hilburn and Jorna, 2001].

2. Definitions Concerning Speech and Stress

Stress and speech are both complex processes, with different sources of influence and – concerning stress – different effects, some of them manifested in the speech signal (see Fig. 3).

Speech production is a process, which is realized at very different levels of mental and physical processing. Starting at the ideation to the formulation of actual sentences to the actual acoustic production of an utterance. This process is influenced by many factors, such as gender, age, the environment, so that every utterance is slightly different even if the same phrase is spoken by the same person twice. So besides the actual message, speech carries a lot of additional information about the speaker.

In his work on the modulation theory of speech, Traunmüller tries to classify the information transmitted in speech in four different qualities, namely *linguistic, expressive, organic*, and *perspectival* information (see Tab. 1 and [Traunmüller, 2000]). With the appropriate training human listeners can naturally extract much of the additional information (besides the 'words') when listening to other people speaking.

Stress, on the other hand, is a human state which may be caused by a multitude of reasons, including workload, emotions, or environmental influences. Stress has an effect on various parts of the human body – such as the brain, the cardiovascular system, the eyes, and among others also the voice – and thus affects the speech signal when a person is talking under stress.

Much (perhaps all) of the variability of speech carries information about the state of the speaker. This is generally utilized in social interactions between humans, however the additional information is commonly not 'understood' by machines. Human-machine interaction via speech is at present largely limited to an exchange of 'words' with precisely defined meanings. There are, however, always variations due to para-linguistic content, which is particularly the case when the speech signal is perturbed due to 'stress' on the talker. This is, on the one hand, a problem for current speech recognition systems, on the other hand, it can be useful for extracting information about the speaker – such as the state of stress – from the speech signal.

The aim of this chapter is to describe and explain a working definition of stress which forms the basis of the literature survey in the later chapters. It is roughly based on the North Atlantic Treaty Organization (NATO) project report on Speech under Stress [Hansen et al., 2000], which itself is based on the discussions at the Lisbon workshop on Speech under Stress in 1996, as summarized by [Murray et al., 1996b]. The main focus of this work is on the effects of stress on speech production by humans, in particular on the effects of workload induced stress.

Table 1: *Information in speech (from [Traunmüller, 2000])*

Quality	Information	Phenomena
Linguistic Conventional, social	Message; dialect, accent, speech style, ...	Words, speech sounds, prosodic patterns, ...
Expressive Within-speaker variation, psycho-physiological	Emotion, attitude; adaptation to environment, ...	Type of phonation, register, vocal effort, speech rate, ...
Organic Between-speaker variation, anatomical	Age, sex, pathology, ...	Larynx size, vocal tract length, ...
Perspectival Physical, spatial	Place, orientation, channel, ...	Distance, projection angles, acoustic signal attenuation, ...

Giving a rigid definition for 'stress' is a difficult problem. There is no single definition that satisfies everyone and every situation. A definition of 'stress' must also be considered in the context of a model of the human speech production system; indeed, this may be the only way to make sense of the subject. The next section considers definitions of stress in fairly general terms, while a model of speech production under stress is described in section 2.2. Section 2.3 considers a taxonomy of stressors, as related to the speech production model. In section 2.4 a link between stress analysis and investigations on emotional speech is proposed.

2.1. Definitions of Stress in Relation to Speech

Stress in relation to speech has to be carefully defined, first of all to distinguish it from the meaning of 'stress' as an emphasis given to an accented syllable, as used in linguistics. Therefore we will talk about 'speech under stress', which is less ambiguous. It implies that some form of 'pressure' is applied to the speaker, resulting in a perturbation of the speech production process, and hence of the acoustic signal. This definition necessarily implies that a 'stress free' state exists, i.e., when the pressure is absent, although it may be hard to find circumstances completely free of stress. Therefore, for practical purposes the unstressed state may be defined as the state in which the reference samples of speech are collected. In the context of workload induced stress in ATC this might be defined differently. The neutral situation would probably denote a working condition, where optimal performance is possible, i.e., a workload at which high performance can be sustained, a positive environment, and no negative influence from health or emotional conditions (see Fig. 3).

The term *'stressor'* is used to denote a stimulus that tends to produce a stress response; the actual responses produced by individuals may vary to a great degree and it may not be possible to classify all possible stimuli as either stressful or not.

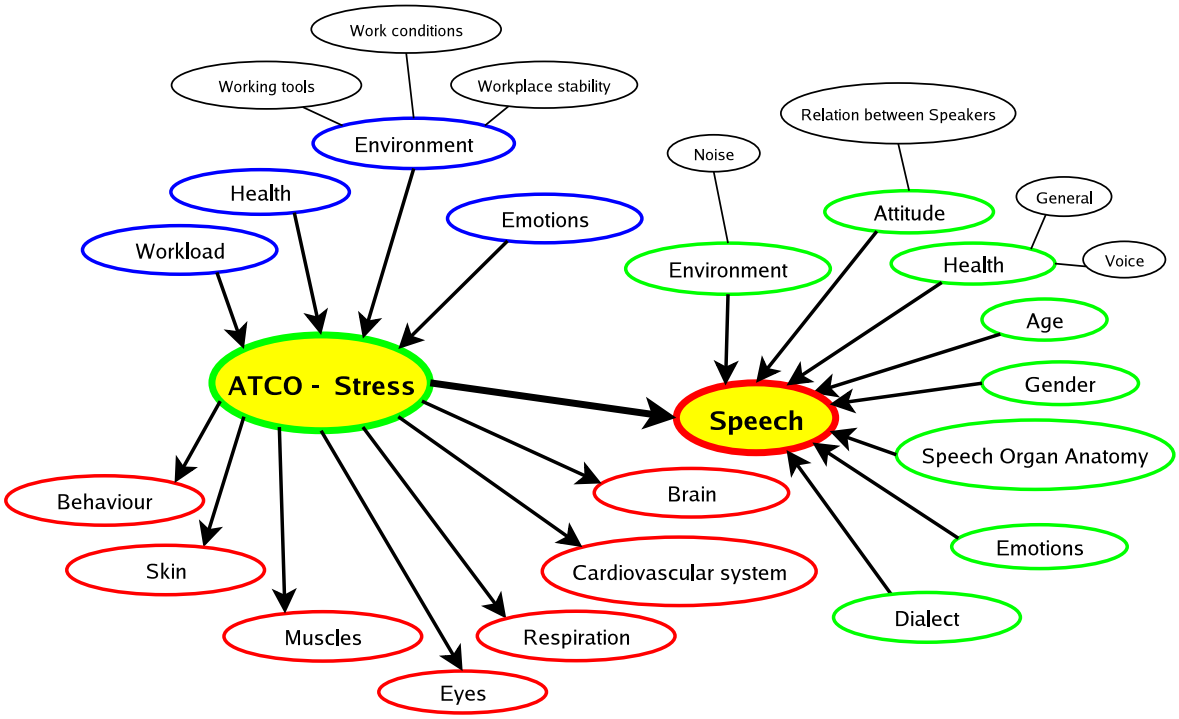


Figure 3: *Speech and stress in the context of a variety of influences.*

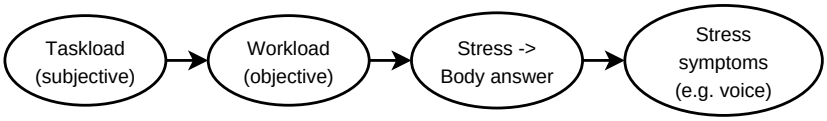


Figure 4: *Relation between taskload, workload, stress and stress symptoms*

2.1.1. Workload Stress Definitions

Since the aim of this report is to focus on the relation of workload and its impact on the human voice, we want to define some terms related to this subject. One has be careful to differentiate between taskload, workload and stress (see Fig. 4).

Taskload: An objective measure for the demand of the work. This includes the complexity of the task and the amount of work to manage (e.g., the amount of aircrafts under control).

Workload: The subject dependent level of used capacity of resources. This can be due the the complexity of the task (qualitative workload) or the amount of work (quantitative workload).

Work overload: The workload rises above a certain threshold, i.e., the subject feels that it is not up to the demands of the task. This can induce stress, which can, besides other effects, result in changes in the human speech production.

Work underload: The workload is below a certain level, where a focus on the task is not the case anymore. This can lead to inattentiveness, which can further lead to errors. Work underload can also be related to reduction of capacity due to, e.g., boredom, which then leads to increased workload [de Waard, 1996, Ch.2]. Another case of work underload would be the situation of highly repetitive work, which can lead to frustration and therefore cause stress [Byrne and Parasuraman, 1996].

For a detailed discussion on the above mentioned definitions see [de Waard, 1996, Ch.2]. Figure 5 shows the dependence of workload and performance on the taskload with the theory of decreased subjective capacity during times of low demand, which leads to higher workload. In [Athénes et al., 2002] the relation between workload an task complexity in an ATC context is discussed.

2.2. Model of Speech Production under Stress

Hansen *et al.* have proposed a very useful model of speech production under stress [Hansen et al., 2000]. It is slightly edited for better readability:

> *Speech production begins with abstract mental processes: the desire to communicate and the idea which is to be communicated. Suitable linguistic units (words or phrases) have to be chosen from memory and formed into a sentence, subject to grammatical constraints. From the abstract sequence of words, a corresponding sequence of articulatory targets must be generated, then appropriate motor programs for the targets must be activated, with modifications to take account of context and para-linguistic information. This results in patterns of nerve impulses being transmitted to the muscles which control the respiratory system and vocal tract. The final stages are purely physical: the generation of acoustic energy, the shaping of its spectral and temporal characteristics, and its radiation from the mouth and/or nostrils. These processes are summarized on the left side of Fig. 6.*

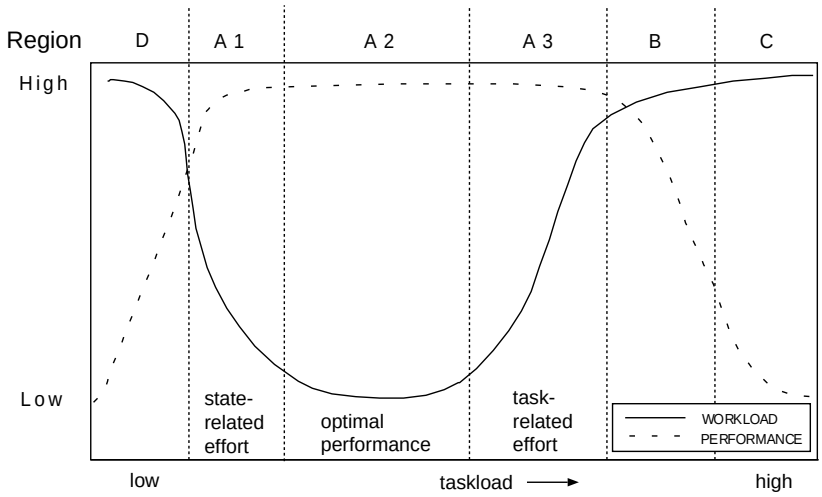


Figure 5: *Workload and performance in 6 taskload regions. In region D (D for deactivation) the operator's state is affected. In region A2 performance is optimal, the operator can easily cope with the task requirements and reach a (self-set) adequate level of performance. In the regions A1 and A3 performance remains unaffected but the operator has to exert effort to preserve an undisturbed performance level. In region B this is no longer possible and performance declines, while in region C performance is at a minimum level: the operator is overloaded (from [de Waard, 1996]).*

Although described as a sequence, these processes overlap to a considerable degree, especially in the information processing stages. It is probably quite normal for the acoustic signal to be started before the sentence has been fully formed in the higher levels of the brain. There are also many layers of feedback within the overall process, which may cause the process to be halted or re-directed at any stage if an error is found. Evidence for this is provided by the dysfluencies in normal speech. A given stressor can be considered to act primarily on a particular stage of the speech production process, and this should define its effects, within certain (possibly very wide) limits. There follows a classification of stressors based on the level in the above model at which the stressor acts (as shown on the right side of Fig. 6).

'Zero-order' *stressor effects are easiest to understand. They are those which have a direct physical effect on the speech production apparatus. Examples of such 'zero-order' stressors include vibration and acceleration. To a first approximation, the patterns of impulses in the motor neurons and the resulting muscle tensions will be the same as in the unstressed condition, but the responses of the articulators will change because of the external forces applied to them. There may be some modification of the neuromuscular commands as a result of auditory and proprioceptive feedback. In general, these stressors will have similar effects on all speakers, but differences in physique will affect the response, particularly under vibration when resonances in the body can magnify the effects at particular frequencies.*

'First-order' *stressors result in physiological changes to the speech production apparatus, altering the transduction of neuromuscular commands into movement of the articulators. These may be considered largely chemical effects, whether the*

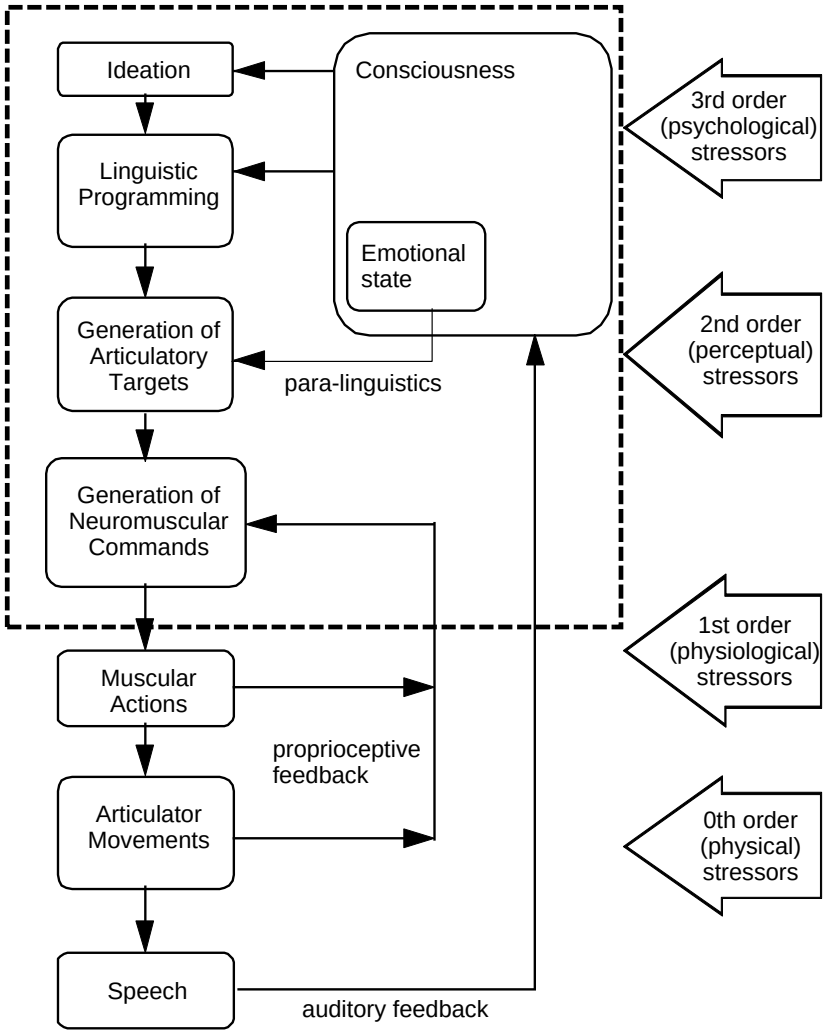


Figure 6: *Outline of speech production process and order of stress (from [Hansen et al., 2000]).*

chemical mediators originate externally as medical or narcotic drugs, or internally as a result of illness, fatigue, dehydration, etc. Factors affecting the feedback of articulator positions would also fall into this class. Differences in individual responses could be large, especially where habituation or training is involved.

'Second-order' *stressors are those which affect the conversion of the linguistic program into neuromuscular commands. This level could perhaps be described as 'perceptual' as it involves the perception of a need to change the articulatory targets, but without involving higher level emotions. The most common example is the Lombard effect, in which the stressor is noise and the response is to increase vocal effort. This is a particular case of perception of a problem with the communication channel; other similar problems may be perceived aurally or by feedback from the listener.*

'Third-order' *stressors have their effects at the highest levels of the speech production system. An external stimulus is subject to mental interpretation and eval-*

uation, possibly as a threat (as implied by the word 'stress'), but other emotional states such as happiness will also have their effect at this level. Some third-order stressors may not be external stimuli at all, but originate from within the mind. These complex mental processes may affect the original idea and the construction of an utterance, which is perhaps outside the scope of this report. There will certainly be effects on the articulatory targets and neuromuscular program expressing the emotion via para-linguistics, and possibly, through changes to physiological arousal, also at the transduction level.

The classification of stressors described above is based on the level of the speech production process at which the stressor has its primary effect. It should be borne in mind that there may also be secondary effects at other levels.

2.3. Taxonomy of Stressors

In this section we try to identify various sources which can induce human stress and try to classify them within the stress model given in the last section. Only primary effects are considered, though in reality stressors can have an impact at different levels and sometimes it is not easy to identify the level at which the primary effect occurs. Self-awareness also makes it possible that almost all stressors may have a 'third-order' effect, due to their mental perception by the human subject. For these reasons, a definitive classification of stressors is not possible and the tentative classification offered here (Tab. 2) would need to be reconsidered in the light of a particular application.

Table 2: *Stressor taxonomy (from [Hansen et al., 2000]).*

Stressor order	Description	Stressors
0	Physical	Vibration, acceleration (G-force), personal equipment, pressure breathing, breathing gas mixture
1	Physiological	Medicines, narcotics, alcohol, nicotine, fatigue, sleep deprivation, dehydration, illness, local anesthetic
2	Perceptual	Noise (Lombard effect), poor communication channel, listener has poor grasp of the language
3	Psychological	Workload, emotion, task-related anxiety, background anxiety

'Physical stressors' include vibration and acceleration, which are self-explanatory and common in cockpit environments, but not to be expected for air traffic controllers. Personal equipment includes clothing and other items worn on the body, which may exert pressure on the vocal apparatus. Examples are the oxygen mask worn by fast-jet aircrew which applies pressure to the face and restricts jaw movement, or a safety harness restricting chest movement. Positive pressure breathing (used to maintain consciousness at high G-levels or in the event of loss of cabin pressure at high altitude) has the effect of distending the vocal tract and thus changing its resonant frequencies.

'Physiological stressors' include a wide range of stimuli, from narcotic drugs to sleep deprivation, and many of these may also have second and third order effects.

'Perceptual stressors' form a much more limited set, but again may also have third order effects arising from, for example, frustration at the difficulty of communicating via a poor channel.

'Psychological stressors' involve a range of stimuli that could be considered as almost unlimited, as the input is interpreted in the light of the individual's beliefs. Workload will be a third-order stressor. The emotional state of the speaker will also be of concern (e.g. whether a controller had a personal conflict with a close person before coming to work).

Although in the literature a number of other possible sources of stress are considered, we limit our focus here specifically on workload induced stress. Nevertheless, an overview of the impact of other sources of stress on the human speech production will be given, as well. Some of the results are expected to be useful for workload induced stress observation.

Since the investigations in speech under stress are closely linked to analysis of speech under emotional arousal, we will present some considerations regarding emotional speech next.

2.4. Emotional Speech

A number of investigations considers the effect of emotional states of the speaker on the speech signal. Methods for automatic 'emotion recognition' from the speech signal often use similar features and analysis approaches as are used for stress recognition, but also attempts to synthesis of emotional speech can be of interest here.

Human emotions are often classified into a number of 'basic emotions', e. g., anger, disgust, fear, happiness, sadness, and surprise, as compared to a 'neutral' emotional state. The notion of basic emotions goes back to Darwin, and is still controversial [Ortony and Turner, 1990, Ekman, 1992, Turner and Ortony, 1992]. However, many approaches to emotion recognition from the speech signal (and from facial expressions) and to emotional speech synthesis use a small number of emotion classes [Cahn, 1990, Amir and Ron, 1998, Huang et al., 1998, Li and Zhao, 1998, Johnstone and Scherer, 1999, Kang et al., 2000, Petrushin, 2000, Zhao et al., 2000].

Human emotions may, on the other hand, also be described in a continuous, multi-dimensional space, for example in a two-dimensional space spanned by the variables 'arousal' (passive – active) and 'valence' (negative – positive) [Lang, 1995], or in similar higher-dimensional spaces (see, e.g., [Russell, 1994, Pantic and Rothkrantz, 2003]). In such parameter spaces, the discrete emotions can be arranged around the 'neutral' emotion position. The continuous parameter space, however, allows for a fine grained description of emotional flavors, like the socially influenced level of arousal, e.g., for anger ('cold anger' vs. 'hot anger'). Investigations in emotion recognition from speech signals using this concept are for example [Ball and Breese, 1999, Pereira, 2000, Alter et al., 2003], general considerations as well as surveys on emotion modeling can be found in [Lang, 1995, Cowie et al., 2001, Cowie and Cornelius, 2003, Scherer, 2003, Pantic and Rothkrantz, 2003].

In general, a close relation of emotion recognition and stress recognition can be identified, both by the utilization of similar features (fundamental frequency, jitter, glottal vs. noise excitation, etc.) and by the possibly direct linking between the arousal dimension and stress. Hence, it seems difficult to understand the separation of research on stress and on emotion recognition (cf. [Cowie and Cornelius, 2003]). Several investigations in emotion recognition may, for example, be directly applicable to stress recognition (e.g., [Huber et al., 2000, Alter et al., 2003]).

3. Databases with Speech under Stress

One very important prerequisite for speech research is the availability of data. Collecting data can be a very time and money consuming task, so appropriate databases are an important resource for research.

There is a number of databases containing speech under stress with varying degree of usefulness for the task of ATCO workload stress monitoring. Table 3 and Table 8 in the appendix present an overview of databases of speech under stress (most information is taken from http://cslr.colorado.edu/rspl/STRESS/info.html). Whenever possible, information on the availablity of the database is given.

3.1. ATC Apprenticeship Database

This is a corpus made up of recordings of spoken and spontaneous dialogs between pilots and air traffic controllers under training [Bouraoui et al., 2003]. The languages used were English and French. The corpus consists of more than 19 hours of recording of 32 speakers taking 5946 speech turns. The corpus was not recorded with the task of stress identification in mind, but might be useful, since for trainees higher workload induced stress levels can be expected sooner than for experienced staff. For further details see Table 8 in the appendix. For further information on availability contact Philippe Truillet <philippe.tuillet@irit.fr>.

3.2. SUSC-0 Database

This database contains two main parts: FCTR and MAYDAY. The FCTR part contains speech from fighter controllers measured under stress conditions. The MAYDAY section includes transcribed speech from two real instances of radio communication with fighters, which experience situations of different possible stress level: From the initial ground aircraft system check, through preliminary discovery of engine problems during fight, followed by pilot actions to correct engine emergency, until safe resolution of the situation. Therefore,

Table 3: *Databases of speech under stress (from [Hansen, 1998]).*

Database	**Source**	**Description**	**# Spkrs**	**Sex**	**DataBase Purpose**
ATC - apprenticeship corpus	CENA, Toulouse	Speech from apprenticeship air traffic controllers	32	Male	Spoken dialog analysis
SUSAS (Speech under simulated and actual stress)	RSPL-Duke University (USA), LDC	Stressed Speech from fairground rides and helicopters, multi-style speech	20	Both	Simulated stress: analysis, modeling, stress classification, recognition, and synthesis
SUSC-0	TNO/IZF (The Netherlands)	Speech under stress conditions	9	Male	Analysis of stressed speech and testing systems
DERA Number Plate Database	Defense Research and Evaluation Agency (DERA), Speech Research Unit, Malvern, England	A Corpus of Read Car Number Plates Recorded Under Conditions of Cognitive Stress	15	Both	Investigation for the effects of cognitive stress on speech recognition performance and speaker coping strategies.
DCIEM Map Task Corpus	Linguistic Data Consortium	Map Task Corpus Dialogue under sleep deprivation and drug treatment	35	Both	Study effects of sleep deprivation and counter drugs

this database can be used to assess speech signals produced under various degrees of stress. There are nine male speakers. All speech signals were recorded from actual aircraft pilot-controller communications with fairly noisy background. See also Table 8 in the appendix and [Hansen et al., 2000].

3.3. SUSAS Database

The speech under simulated and actual stress (SUSAS) database [Hansen and Bou-Ghazale, 1997, Hansen et al., 2000] was established to perform research on analysis and recognition of speech under stress in different domains. Structure:

- 32 speakers (13 female, 9 male), 22–76 years of age
- 16.000 utterances
- five stress domains:
  - talking styles (slow, fast, soft, loud, angry, clear, question)
  - single tracking task or speech produced in noise (Lombard effect)
  - dual tracking (compensatory and acquisition) computer response task
  - actual subject motion-fear task (roller-coaster, Apache helicopter; G-force, Lombard effect, noise, fear)
  - psychiatric analysis data (speech under depression, fear anxiety)
- highly confusable vocabulary set of 35 aircraft communication words

The tracking task is the domain which is the most appropriate for workload induced stress research. The database can be obtained from the Linguistic Data Consortium (LDC). Information on the database can be found at: http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC99S78

3.4. DCIEM Map Task Corpus

This corpus [Bard et al., 1996] contains recordings of people under sleep deprivation and the effect of drugs used to combat the effect of sleep loss. The recorded speech is from a task, where pairs of talkers collaborated to reproduce on one partner's schematic map a route preprinted on the other's (as used in the HCRC Map Task Corpus [Anderson et al., 1991]).

The corpus consists of recordings of 35 test persons (33 male/2 female) performing 216 dialogs. Included are recordings before the sleep deprivation, with sleep deprivation of increasing intensity with and without drug treatment (a part of the group only received placebo drugs). The corpus is collected with a military background in mind; it was compiled in the framework of a NASA study on automatic speech processing. The database is available from the LDC. Information on the database can be found at: http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC96S38.

3.5. DERA Numberplate Database

This database [Mellor and Graham, 1995] provides a corpus of spoken car number plates, recorded at the Defence Evaluation Research Agency (DERA) Speech Research Unit in 1992. The corpus comprises recordings of read car number plates dictated in the International Civil Aviation Organization (ICAO) alphabet form plus digits. 16 speakers were recorded, eleven male and four female (plus a trial recording with a male) covering a range of ages and featuring

a number of non-extreme British regional accents. Recordings were carried out under two levels of cognitive stress, achieved by varying the presentation rate of the number plates. The subjects' perceived workload stress during the dictation sessions was recorded to provide data for a corpus of speech under limited stress conditions, though, no information is given how the subjective ratings were obtained. Two stress levels were provided by displaying the data at high and low rates. The fast data display rate was aimed at requiring a dictation performance above the speakers ability, and so lead to performance failure and hence stress. See also Table 8 for further details. Aurix Limited can make this database commercially available when needed. For further information contact Martin Abbott <m.abbott@aurix.com>.

4. Stress Observation in Human Speech

Using speech for stress observation in the human voice has been of research interest for quite a while. Unfortunately most work only deals with a very limited set of test speakers, some even only considering a single speaker. Another problem is the broad definition of stress: From the taxonomy of stressors (sect. 2.3) we see that speech under stress can be everything, from speech in loud background noise, to effects due to physical exertion, emotions, external physical influences, and workload. Results are therefore difficult to compare and not necessarily relevant for the observation of effects of workload induced stress in the human voice. Nevertheless, several features have turned out to be of special interest for analysis of speech under stress. The most prominent feature reported in the literature is the fundamental frequency (F_0).

This section will first review human classification of speech under stress, then it covers general literature on stress analysis from the human voice, rather providing speaking style classification than stress classification. Then, we will focus on workload induced stress and results of investigations, which specifically deal with this subject.

4.1. Human Stress Classification

Human classification ability of speech under stress – which represents a baseline for automatic stress classification – has been investigated in [Bolia and Slyh, 2003]. For the experiments the SUSAS database was used. A group of three males and four females acted as the test listeners. Unfortunately, the classification used comprises rather 'speaking styles' than stress types, given the 8 available classes: *angry, clear, fast, loud, question, slow, soft, neutral*. No workload related stressor was under examination. The experiments were carried out for three different American accents (New York, Boston, General American). Though the results are above chance level, the identification rate is rather low (See Fig. 7 for results for Bostonian accent).

The low identification results may be due to the length of the speech tokens. Only mono- or disyllabic words were used for this experiment.

Another experiment was carried out to identify the acoustic correlates of perceived emotional stress [Protopapas and Lieberman, 1997]. It was analyzed how jitter and other F_0 related features correlate with human stress perception. Perception tests were performed on manipulated stressed and unstressed speech samples. For jitter, no correlation with perceived

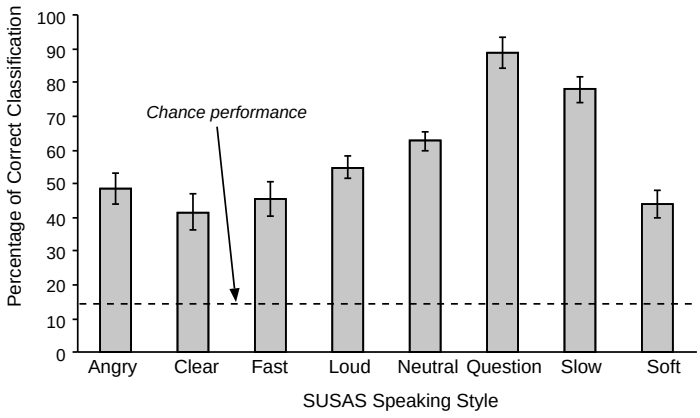


Figure 7: *Mean percentage of correct style classifications for Bostonian accent (from [Bolia and Slyh, 2003]).*

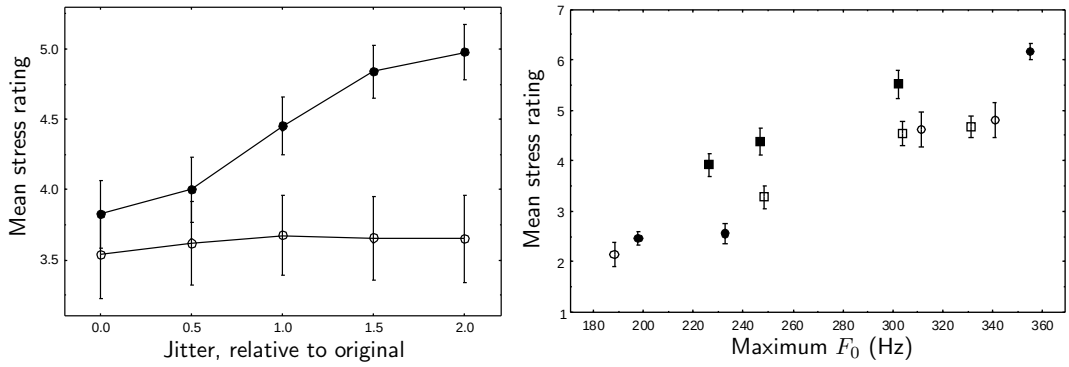


Figure 8: *Left: Mean ratings of speakers' stress (○) and voice hoarseness (●), averaged across subjects and utterances, as a function of relative amount of jitter. Right: Stress ratings ascribed to variants U1(○), U2(□), S1(●), and S2(■) as a function of maximum F0. The rating scale was 1 to 7. Error bars show standard deviation (from [Protopapas and Lieberman, 1995]).*

stress was found (Fig. 8, left), in contrast, hoarseness and jitter are clearly correlated. As well, perceptual experiments with manipulations of the F_0 contour of stressed (S1, S2) and unstressed (U1,U2) test utterances showed that temporal properties (the prosody was time-inverted) have no influence on stress perception. Maximum and mean F_0 show a significant correlation with perceived stress (Fig. 8, right), note all speakers were male.

4.2. Features for Stress Analysis

There are some stress-related features, which are present in most of the published work, these include pitch, duration, intensity, glottal source properties, and vocal tract spectrum. The group around J.H.L. Hansen has introduced the Teager energy operator for stress analysis and published several papers on it. We present an overview of the different features used, and classification results in the following.

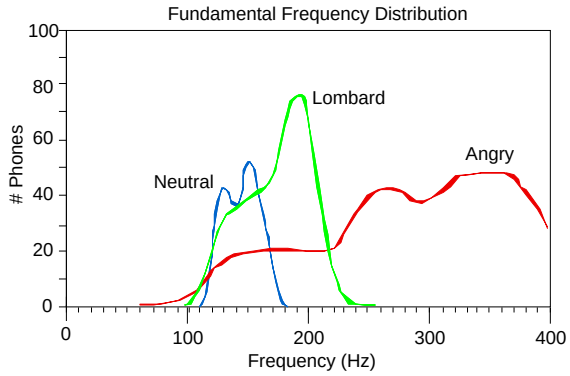
4.2.1. Pitch

Pitch (or F_0) is a widely studied feature for analysis and classification of speech under stress. Pitch features include:

- Pitch contour
- Jitter (short-term pitch variations)
- Pitch range
- Pitch variance
- Maximum Pitch

In a subjective listening test, the mean and maximum pitch were found to be a significant cue for human perception of emotional stress, see [Protopapas and Lieberman, 1997] and section 4.1.

Hansen reported that mean pitch values and pitch variance can be used as a significant indicator for different types of speech under stress, see Fig. 9, [Hansen, 1996]. In other studies by this group pitch also appeared to be a very good feature for classification of speech under stress [Zhou et al., 2001].



	Neutral	Lombard	Angry
Mean (Hz)	145	160	253
Std Dev	15	24	95

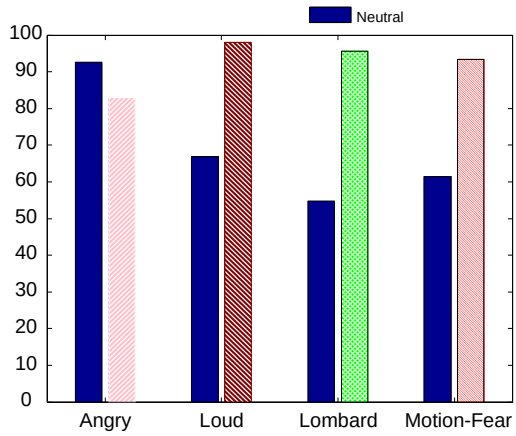


Figure 9: *Left: Sample F_0 distribution variation for speech under neutral, Lombard, and angry stress conditions (from [Hansen et al., 2000]). Right: Text-independent pairwise stress classification results using Pitch (from [Zhou et al., 2001]). The height of the bars indicate the percentage of correct classification of neutral and stressed speech for different domains. Both use the SUSAS database.*

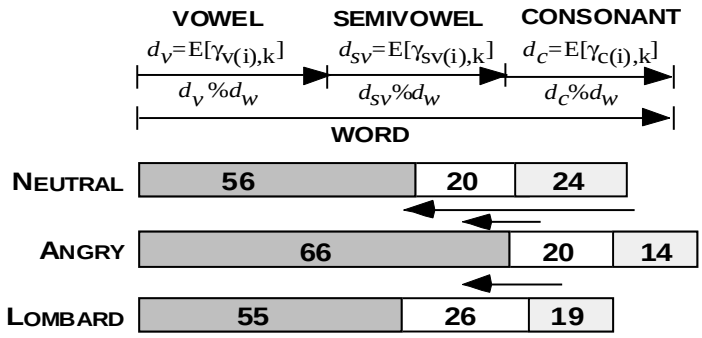


Figure 10: *Duration variation for neutral, angry and Lombard speech for different phoneme types (from [Hansen et al., 2000]).*

4.2.2. Duration

[Womack and Hansen, 1996] included phoneme duration in their target driven feature approach to stress classification. For the feature analysis, they distinguished between different phone classes and made a ranking of separability for each feature and phone class. See page 19 for further information on this work and Fig. 13 for results, which do not include the duration feature, though.

[Hansen, 1996] investigated duration and word rate as features for classification of speech under stress. Significant changes are reported for different stress classes. In Fig. 10, the duration variation for *neutral, angry* and *Lombard* speech is compared. Unfortunately, actual classification results based on duration are not reported.

4.2.3. Energy

Teager energy operator (TEO): Hansen and his group used different features based on the TEO [Cairns and Hansen, 1994, Hansen et al., 2000, Zhou et al., 2001].

The scope of those studies is very wide, since stress includes speech in noisy conditions, speech under influence of Lombard effect, anger, fear, G-force movement (fighter pilot), work-

load. They investigated features based on the TEO, which is a nonlinear operator. It is based on the observation of the energy stored in a physical oscillation process. Kaiser introduced a simple form of the operator as

$$\Psi\left[x(t)\right] = \left(\frac{d}{dt}x(t)\right)^2 - x(t)\left(\frac{d^2}{dt^2}x(t)\right),$$

or

$$\Psi\left[x(t)\right] = \dot{x}^2 - x\ddot{x}$$

where $\Psi[\cdot]$ is the Teager energy operator and $x(t)$ is a continuous speech signal. He also derived a discrete-time form of the operator:

$$\Psi\{x[n]\} = x^2[n] - x[n+1]x[n-1],$$

where $x[n]$ is a sampled speech signal [Kaiser, 1993]. He found the TEO to be useful for amplitude modulation (AM)-frequency modulation (FM) energy tracking. The application for speech is based on a model, where the speech signal is seen as the sum of AM-FM signals of which each is centered at a formant frequency. [Gopalan, 2001] uses the TEO to track F_0 and formats and finds some relations to

The TEO is usually applied to band-pass filtered speech signals, since it reflects the energy of the nonlinear flow within the vocal tract for a single resonance frequency.

Zhou *et al.* used the following TEO features [Zhou et al., 2001]:

- Variations of FM Component (TEO-FM-Var)
- Normalized TEO Autocorrelation Envelope (TEO-Auto-Env)
- Critical Band Based TEO Autocorrelation Envelope (TEO-CB-Auto-Env)

For evaluation the SUSAS database [Hansen and Bou-Ghazale, 1997] was used. Since the TEO is better applicable to voiced sound, only high-energy voiced sections of the speech utterances (16 bit, 8 kHz) were used for the evaluation.

The classification task was done using a five-state, continuous-distribution hidden Markov model (HMM)-based classifier with two Gaussian mixtures each.

Three different evaluation procedures were performed including the traditional pitch and mel frequency cepstral coefficient (MFCC) features for comparison. First, a text-dependent pairwise classification was used to preselect features, then a text-independent pairwise classification and a text-independent multi-style stress classification was performed. The TEO-FM-Var and TEO-Auto-Env features are said to be not effective for stress classification due to their dependence on pitch estimation accuracy. The TEO-CB-Auto-Env feature was found to be the best feature for stress classification in terms of accuracy and reliability (see Fig. 11). The MFCC feature performs well for text-dependent stress classification, but degrades for text-independent classification. Pitch is reported to be useful, but lacks consistency and reliability, again because it relies on accurate pitch determination (in this study manual corrections were performed).

The results for the stress detection using the TEO-CB-Auto-Env feature are summarized in table 4. The results show that stress/neutral classification is quite reliable, tough specially for the multi-style scenario the results are not very accurate. Another question is, whether a workload induced stress manifests itself in the voice so that it can be measured with this kind of parameters. Another open issue is the determination of a certain threshold of work overload and underload and whether speech can be used reliably to classify a violation of those workload limits.

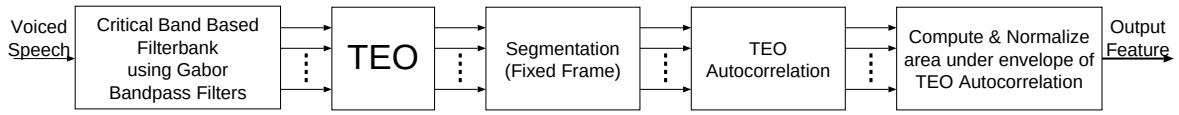


Figure 11: *Critical Band Based TEO Autocorrelation Envelope (TEO-CB-Auto-Env) feature extraction (from [Zhou et al., 2001]).*

In [Fernandez and Picard, 2003] a combination of a wavelet based subband approach and the TEO was used. For 21 subbands the TEO was calculated and then an inverse discrete cosine transform (DCT) was applied to the log of the energy coefficients to obtain TEO-based 'cepstrum coefficients', which are applied frame by frame to the speech signal. To reflect frame-to-frame correlations an autocorrelation measure between frames was also used for the final feature vector. Further information on the classification procedure and results can be found in [Fernandez and Picard, 2003] and section 4.3.3.

Other work using the TEO include [Rahurkar et al., 2002, Rahurkar and Hansen, 2003, Ruzanski et al., 2005].

4.2.4. Glottal Exitation

Cummings and Clements [Cummings and Clements, 1995] (see also [Cummings and Clements, 1990, Cummings and Clements, 1992]) investigated speech under stress with a focus on the glottal waveform. The goal of the research was to find a set of shape parameters that uniquely describes the glottal waveform for different speech styles.

They analyzed speech from two American speakers, from the Lincoln Labs style database, which is now part of the SUSAS database using the utterances 'fix', 'six', and 'destination' with the following speaking styles: *normal, angry, 50% tasking, 70% tasking, clear, fast, Lombard, loud, question, slow, and soft.* For our topic, the 50% and 70% tasking are of special interest. They are produced while the speaker is under a specified level of taskload (see section 3.3 and [Hansen et al., 1998, Hansen et al., 2000] for further details).

They performed inverse filtering of non-nasal vowels with linear predictive coding (LPC) parameters obtained during glottal closure, since theoretically at this moment the all-pole model assumption is best met. This procedure was performed semi-automatically that means manual work was involved in the process.

They came up with parameterization of the glottal waveform with six parameters (opening slope, closing slope, opening duration, top duration, closing duration, closed duration). In [Cummings and Clements, 1992] they used the beta function to model the glottal pulses:

$$y = A\left(\frac{x}{x_{max}}\right)^{b}\left(1-\frac{x}{x_{max}}\right)^{c},\ 0 < x \leq x_{max}$$

Table 4: *Results for TEO-CB-auto-env feature for multi-style stress classification*

	Correct detection rate (%)		**Distribution of Stress Detection**		
Test Class	Neutral	Stressed	Angry	Loud	Lombard
Neutral	70.6		3.8	27.5	68.8
Angry		96.3	65.0	29.2	5.8
Loud		97	34	51.9	14.1
Lombard		91.5	22.3	32.8	44.9
			stress detection sums up to 100%		

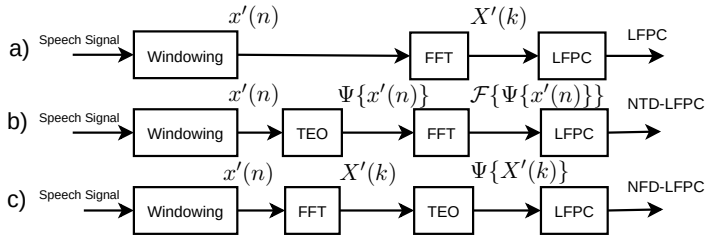


Figure 12: *a) Log frequency power coefficients (LFPC), b) Nonlinear time-domain LFPC, c) Nonlinear frequency domain LFPC (from [Nwe et al., 2003b]).*

They performed extensive statistical tests (Kolmogorov-Smirnoff test, Chi Square test) on the parameters to show differences between the different speech styles. First, pair-wise comparisons between the different speaking styles were performed to test which of the wave-shape parameters are not significantly different for different styles. They came up with the result that all waveforms, but the *clear* vs. *question*, had significantly different distributions in terms of the three waveform parameters closing slope, opening slope, and closed duration.

Assuming a Gaussian distribution of the parameters, the Mahalanobis distance, which is a distance measure between two multivariate Gaussian distributions, of the different speaking styles was calculated. A detailed study of the results can be found in the paper, here only the workload related results will be discussed. The *50% tasking* and *70% tasking* are confusable with each other, and the *50% tasking* is also confusable with *normal*, while the *70% tasking* is confusable with *fast*.

In a follow up study [Waters et al., 1994], stress detection was specially aimed at the *70% tasking* condition from the above mentioned database. The features used were the same as described in [Cummings and Clements, 1995, Cummings and Clements, 1992]. After statistical analysis they also came to the conclusion that between stressed and unstressed speech there is significant difference between the parameters, though speaker dependence was also reported. They performed classification tests using a multi layer perceptron (MLP) neural network with 8 nodes in the hidden layer. Over 25 different trials, they reported a mean classification rate of 71%, which is significantly better than chance. Unfortunately not many details are reported about the procedure.

Other work [Hansen et al., 2000] concentrated on the spectral slope of the glottal excitation. There some difference was shown between *normal, Lombard* and *angry* speech.

4.2.5. Vocal Tract Spectrum

The vocal tract spectrum is also a widely investigated area. A wide variety of features can be used to describe the vocal tract spectrum, such as Cepstrum based features, Fourier transform coefficients, or linear prediction coefficients.

In [Nwe et al., 2003b, Nwe et al., 2003a] experiments with different spectrum based features have been done. They used log frequency power coefficients (LFPC) with and without the TEO. Additionally, they included MFCC and linear prediction cepstral coefficient (LPCC) features for comparison.

For the LFPC they calculate a fast Fourier transform (FFT), then the log based power spectrum is found by accumulation of the result into 12 log-frequency filter banks (Fig. 12a). The approaches using the TEO either apply the TEO on the time-domain signal and then do an FFT (Fig. 12b), or first calculate the FFT and then apply the TEO (Fig. 12c) before they calculate the log-frequency power coefficients.

The classification is done using a 4 state continuous density HMM classifier with two Gaussian mixtures per state.

They evaluated the algorithm with two data sets, one was data from the SUSAS database, discriminating the styles *anger, clear, Lombard*, and *loud*, the other was a self recorded dataset of emotional speech including emotions *anger, disgust, fear, joy, sadness*, and *surprise*. Both pair-wise and multi-style classification was performed. Results are summarized in table 5. The results for the LFPC feature set with 87.8 % correct classification rate is quite high, if one keeps in mind the chance rate at 25 %.

MFCC was also used by others [Zhou et al., 2001], but only pair-wise classification results were reported, which are below the numbers in the work by [Nwe et al., 2003b].

Another study employed a codebook of neural networks, one for each phoneme group and stress condition using features, which are preselected to achieve best separability for each stress condition and phoneme group [Womack and Hansen, 1996]. Features used were:

Articulatory based: i.e., features derived from the articulatory cross section areas calculated from the speech signal.

Exitation based: pitch, phone class duration, intensity

Cepstrum based: mel-cepstral parameters and derivations, such as autocorrelation of mel-cepstral parameters.

The features which offered the best separability were autocorrelation of mel-cepstrum and pitch.

They report very good results (see Fig. 13) at the cost of high computational costs, though. First an HMM based recognizer is used to classify the different phoneme groups (silence, fricatives, vowels, affricates, nasals, semi-vowels, diphthongs). Then for every stress and phoneme group a neural network is used for stress classification.

4.3. Workload related Research

Workload induced stress is the focus of only a handful of papers, which will be discussed here in detail. To avoid the focus being to narrow also investigations on psychological stress, such as analysis of speech in out-of-control situations are included in this section.

4.3.1. Cognitive Workload

Lively *et al.* [Lively et al., 1993] investigated speech under workload stress. Five test subjects had to perform a visual compensatory tracking task while speaking test sentences of the form 'Say hVd again', where 'V' is a vowel. Acoustic measures were taken such as intensity, spectral tilt, F_0, F_0 variability, duration, and formants. While for amplitude measures no general trend related to workload stress was visible for all speakers, spectral tilt was distinctive for

Table 5: *Average stress and emotion classification accuracy.*

Features	Stress %		Emotion %	
	Multi-style	Pair-wise	Multi-style	Reduced set
LFPC	87.8	97.6	89.2	93.4
NFD-LFPC	86.7	97.6	85.8	93.6
NTD-LFPC	74.6	89.9	78.9	91.8
MFCC	66.6	91.9	77	90.5
LPCC	68.6	88.6	67.3	85.3

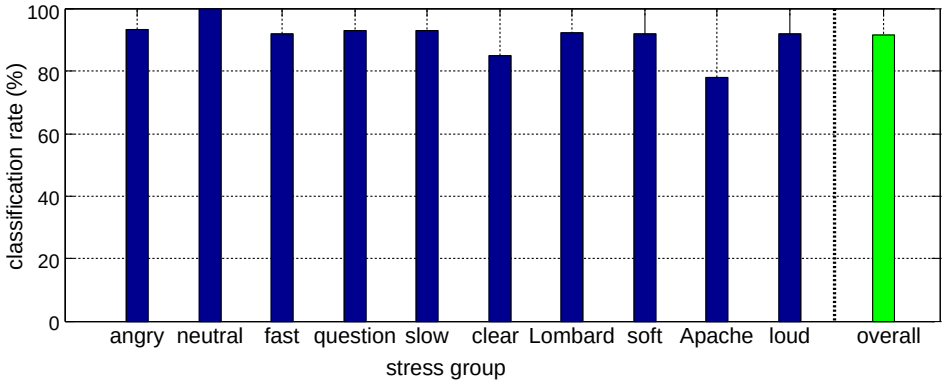


Figure 13: *Stress classification performance with 11 speakers from SUSAS database, 35 different words and 10 stress conditions using phoneme and stress group dependent neural networks (from [Womack and Hansen, 1996]).*

all speakers but one. No consistent effect of average F_0, but a general decrease of F_0 variability (mean F_0 standard deviation) was reported. The phrase duration decreased significantly for four of the five speakers, for the other speaker the opposite effect was observed. Formant frequencies did not significantly change, so the authors assume, that only voice source features are effected by increased workload.

They stress the difference between workload stress and psychological stress. The latter they relate to emotions such as fear (e.g., a pilot just before impact).

4.3.2. Taskload & Stress

Scherer *et al.* [Scherer et al., 2002] performed experiments, where participants were asked to perform a logical reasoning test under two different conditions:

- Focus on the task without disturbance
- Working on task while attending to auditory monitoring task

They presumed, that the single logical task only produced cognitive load, whereas the dual task produced – in addition to a greater cognitive load – psychological stress.

Acoustical analysis was performed on the following features:

- Speech rate
- Average gradient of energy decay over time
- Mean F_0
- Proportion of energy below 500 Hz

The results suggest the following conclusions (see figure 14).

Cognitive load: Cognitive load due to task engagement will reliably increase speech rate, energy attack and decay gradients, mean F_0, and the proportion of energy in the higher frequency range. It seems reasonable to expect that cognitive and attentional demands primarily affect fluency and speech rate.

Psychological stress: In contrast, only F_0 and the change in spectral energy distribution seem to respond to the induction of psychological stress. It seems reasonable to expect that sympathetic activation mostly affects F_0 parameters.

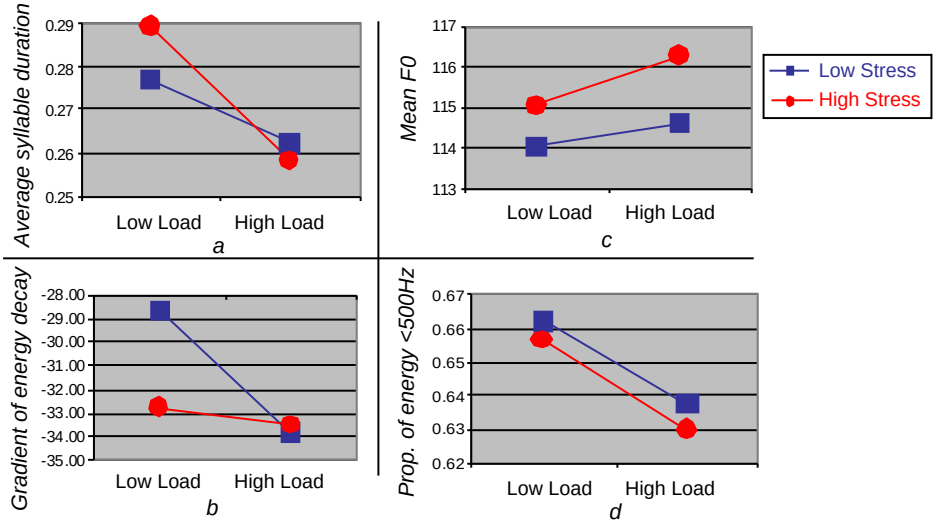


Figure 14: *Differences between load and stress levels. a) Speech rate. b) Average gradient of energy decay. c) Mean F_0. d) Proportion of energy below 500Hz (from [Scherer et al., 2002]).*

The average differences found in this study, even though they are often statistically significant, and thus reliable, cannot be used directly as on-line detectors for the effect of load or stress on speech, due to their strong individual differences.

4.3.3. Driver's Speech under Stress

In a study at the MIT Media Laboratory [Fernandez and Picard, 2003] different levels of stress are simulated by presenting a driver at different driving speed conditions (low/fast) with an arithmetic task (adding up two numbers, $x + y < 100$) either every 4 seconds or every 9 seconds. This gives a set of 4 different levels of cognitive stress. The arithmetic result utterances of four subjects are used for classifying the level of stress. They use a combination of features based on the TEO (see Section 4.2.3) applied to sub-bands after multi-resolution analysis via wavelet transforms. For automatic classification, they tried several different HMM structures, such as the basic HMM, autoregressive HMM, factorial HMM, hidden Markov decision tree (HMDT) and mixtures of HMMs, (M-HMM), see Fig. 15. Additionally classification at the utterance level without temporal features was performed using a support vector machine (SVM) and an artificial neural network (ANN).

The speech data was divided into a training (80 %) and a test (20 %) set and the different models were applied, both for subject-dependent and subject-independent experiments. A detailed discussion of the results can be found in the paper. We will only summarize the outcome here. First the M-HMM setup gives the best results, outperforming the other methods with a recognition rate for both the subject-dependent and the subject-independent test case of more than 10 percentage points higher than the other models. A further point of interest is the high subject dependence of the results, where for one subject a recognition rate of 96.15 % accuracy is reached for the test set, while for another one only 36.11 % is obtained. The results are summarized in table 6.

Compared to the 25 % chance level the results are significant, though there is still room for improvement. This study is particularly interesting, because they try to classify different levels of stress and not only stressed vs. normal speech.

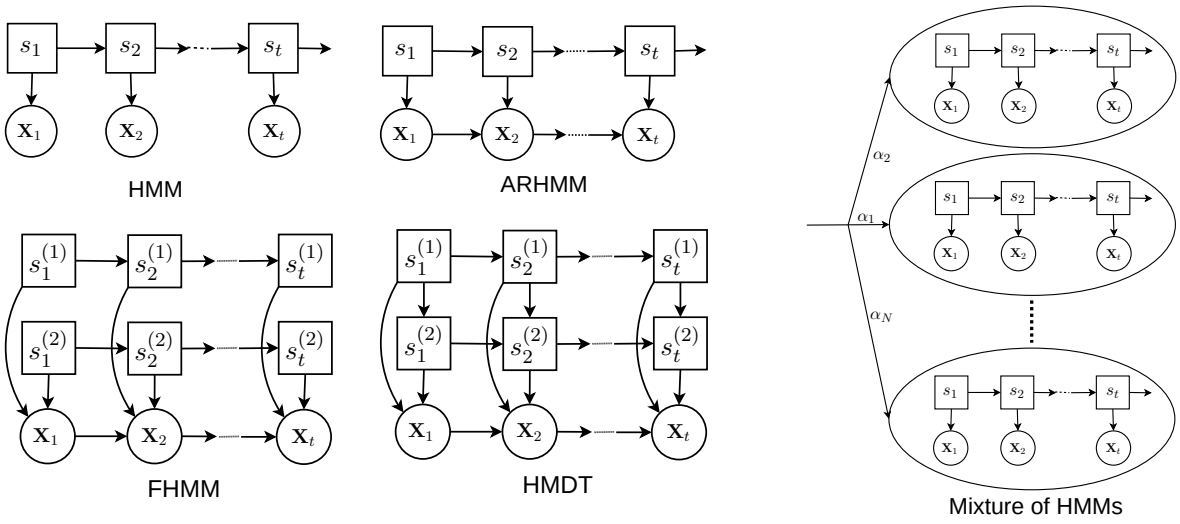


Figure 15: *Graphical representation of models compared in [Fernandez and Picard, 2003] with discrete states s_i and continuous observations $\mathbf{x}_i$.*

4.4. Work Underload, Boredom, Fatigue, and Sleep Deprivation

As it is often pointed out that, seemingly contradictory, many errors occur for non-complex control tasks like, e.g., light traffic (cf. [Majumdar and Ochieng, 2002, table 3]), it is worth considering the automatic detection of work underload. In particular, since physiological measures seem to be no significant indicator for work underload [Desmond and Hoyes, 1996], it may be a good choice to resort to speech signal analysis.

Not many investigations deal with the identification of situations of low workload or stress from the speech signal, since the concern commonly is to detect unusual, stressful situations. However, as noted for emotion recognition above, the identification of some emotional cues is closely related to stress recognition. Work underload, in particular, maybe highly correlated with boredom [Cowie and Cornelius, 2003] or with other 'low arousal' emotions (e.g., sadness). Attempts for identifying work underload thus should look at investigations on emotion recognition, specifically recognition attempts that utilize the arousal-valence emotion space (cf. [Cowie et al., 2001, Pantic and Rothkrantz, 2003]).

The effects of sleep deprivation result in a reduction of human work capacity, so the manageable taskload is much lower than usual. This results in a work overload at a much lower taskload level (see 2.1.1). Fatigue is the more general term, which means, that the human work capacity is reduced either because of sleep deprivation over a too long period of work, or due to a high workload level.

[Whitmore and Fisher, 1996] examined 14 pilots on a 36 hours flight performing a simulated bombing mission. Sleep was allowed for short times if the course of the mission allowed it. The aim was to find relations between subjective fatigue and vocal parameters (F_0 and word

Table 6: *Average classification results of M-HMM (from [Fernandez and Picard, 2003]).*

	Training	Testing
Subject dependent task	96.4 %	61.2 %
Subject independent recognition rate	80.4 %	51.2 %
Best subject	89.19 %	96.15 %
Worst subject	97.39 %	36.11 %

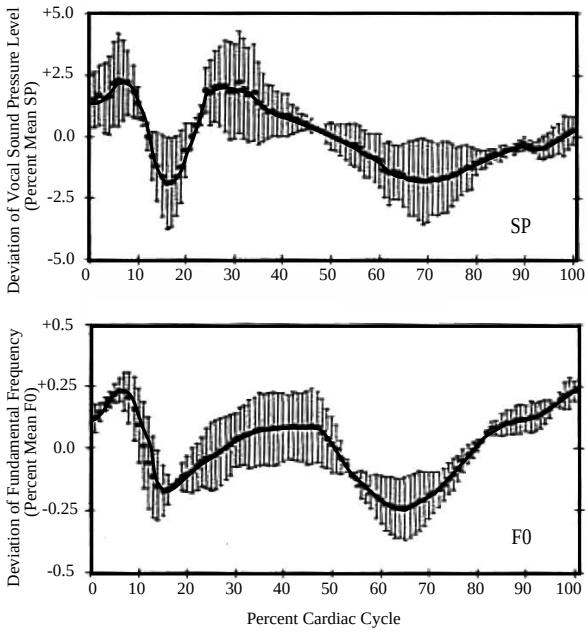


Figure 16: *Normalized vocal sound pressure and fundamental frequency data averaged across all heartbeats and SLP conditions (from [Orlikoff, 1990]).*

rate). Additionally a cognitive test was performed every three hours. Their findings were that the measurement of both the cognitive tests and the vocal parameters showed a dependency rather on the circadian cycle (showing differences specially in the early morning hours) than on the length of the mission.

4.5. Voice and Heartbeat

Heartbeat has been shown to be related to increased stress (see Section A.2), and investigations which show that there is a relation between heart rate and voice features seem very interesting [Orlikoff and Baken, 1989a, Orlikoff and Baken, 1989b, Orlikoff, 1990]. They report a dependence between F_0 and the normalized vocal sound pressure level on the cardiac cycle. See figure 16 for a display of the variation of F_0 and sound pressure level over one cardiac cycle averaged over all heartbeats and different sound pressure levels. The relation is not very strong, though, and only measurable in sustained vowels. For practical purposes this subject is not well enough explored yet. However, the indirect observation of stress-induced heart-rate changes through voice analysis appears to be a promising research area.

5. Speech Processing with Speech under Stress

Though observation of human speech under stress is the main focus of this work, a short view will be taken at the implications of influences of stress on the speech signal for other applications of speech technology.

5.1. Automatic Speech Recognition for Speech under Stress

The deviations from the neutral speech under stressed conditions can increase recognition error rates for automatic speech recognition (ASR) [Chen, 1988, Hansen, 1996, Sarikaya and Hansen, 2000, Womack and Hansen, 1999, Zhou et al., 2001]. Several approaches dealt with speech under stress and ways to improve recognition rates by compensating the influence of stress on the speech signal. Those influences are reflected in changes of signal features as described in section 4, for a limited vocabulary context, such as ATC, the use of illegal vocabulary may also be an issue. The degree of recognition degradation of course depends a lot on the type of stress. In one specific work, workload has been reported to cause degradation of recognition rates [Baber et al., 1996]. In the same paper it is proposed, that ASR by itself my lead to an increased workload.

To improve the performance of speech recognition algorithms under stress, several methods have been proposed. These fall into three main categories (Fig. 17, [Bou-Ghazale and Hansen, 2000]):

- Robust features
- Stress equalization methods
- Model adjustment or training methods

Improvement of speech recognition performance for speech degraded due to stress has, so far, only been marginally successful, so continued research is required.

5.2. Stressed Speech Synthesis

Though the main focus of this report is the observation of stress in the human voice, an understanding of these effects is also useful in synthesizing speech when it is desired to

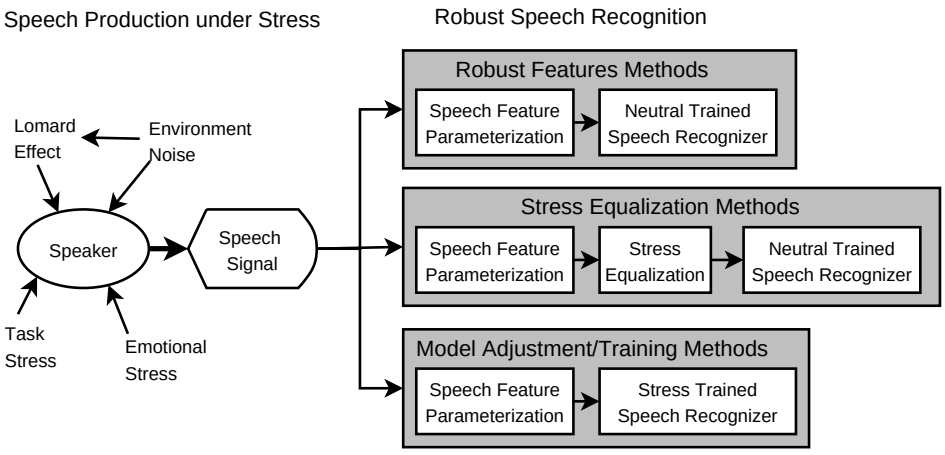


Figure 17: *Robust speech recognition for speech under stress (from [Bou-Ghazale and Hansen, 2000]).*

simulate stress or emotions. As an example, synthetic emotion may be added to cockpit or air traffic controller voice warnings in order to impart a sense of urgency to the pilot or the controller.

A short overview on the influence of stress on speech technologies can be found in [Hansen et al., 2000] or, more focused on emotional speech synthesis, in [Murray et al., 1996a], see also section 2.4. Work on synthesis of speech under stress for speaking style modification of neutral speech has been done using a code excited linear prediction (CELP) vocoder in [Bou-Ghazale and Hansen, 1996] or, using HMMs to represent the deviations of the speech signal from neutral conditions caused by stress in [Bou-Ghazale and Hansen, 1998].

5.3. Stressed Speaker Identification and Verification

Some preliminary experiments on speaker identification have been reported using two of the databases described in section 3. The results of the experiments indicate that speech spoken under stressful conditions degrades the performance of speaker recognition systems. The performance observed when testing a system which was trained with neutral speech can be less than half of the performance of a system trained with material collected under the same stress conditions [Hansen et al., 2000].

6. Summary & Conclusion

6.1. Databases

Although a range of databases with speech under stress exist, they cover a very wide range of sources of stress and are only partly useful for workload induced stress research. Therefore, a more focused database would be necessary for further research. Still, the currently available databases would provide a reasonable starting point for stress related speech research and for the definition of future data collection efforts.

6.2. Classification of Speech under Workload Induced Stress

The classification of the impact of workload related stress on speech is a very specific topic in the context of analysis of speech under stress. Ideally one would like to have a continuous workload estimation to be able to avoid overload and underload all together. The minimum requirement would be a classification of an appropriate workload versus undesired workload levels, such as underload or overload.

So far quite some reseach has been done on stress using a very broad definition, but much fewer work has been reported on classifying workload-induced stress specifically . The most promising approaches use a multitude of features, which are then classified by advanced machine learning techniques, such as ANNs or HMMs.

Section 5.1 and section 5.3 show that speech and speaker recognition degrade significantly for speech under stress, when trained on neutral data. This property could be exploited by using the error rate of the speech/speaker recognizer for stress detection.

6.3. Relation to Research on Emotional Speech

We see the need of incorporation research on emotional speech into research on workload induced stress. In a way because work over- and underload and emotions are closely related. In addition, since emotions can cause similar effects on the voice as workload related impacts, there is a need to include emotional speech knowledge to reduce the false alarm rate.

6.4. Discussion and Recommendations

The present study shows that research in workload-induced stress in speech has become a topic that attracts researchers on an international scale. On the other hand, this research is still in its very early stage with a lot of difficult problems to be addressed, where one of the key issues is the multitude of potential influence factors not related to workload-induced stress, which may result in similar changes of speech features. This non-uniqueness problem needs to be addressed by the collection of more data under varied, but well-controlled conditions, and by focussing the research on a very specific application scenario that may help to either rule out some extraneous influences or, at least, to characterize them in a definite way.

While the overall performance of speech-based stress classification is far from satisfactory it is still not so different from the performance reported for non-speech based methods reported in the appendix. As speech has the clear advantage of being non-invasive/contact-free/ non-intrusive, it seems well worthwhile to explore this approach further. Such research should not be considered as short-term applied research leading to the development of stress-monitoring equipment, but as long-term research that will allow to describe and separate the factors influencing the speech features and serve as the basis for future automatic stress classification systems. An adequate fundamental research program should be built on three pillars, i.e.,

1. Production of extended speech-under-stress databases with a detailed annotation of workload and individual/environmental circumstances, centered on a specific application scenario.

2. Development of feature extraction algorithms and their systematic test on the databases, with the goal of defining features with high discriminative power against non-workload related influences.

3. Evaluation/verification of the proposed methods in a simulation environment involving real-life users over extended time frames, with the possibility to conduct post-experimental interviews.

Such research could be considered as the pro-active basis for future developments towards advanced operator support systems for ATC safety enhancement through speech-based workload monitoring.

References

[Alter et al., 2003] Alter, K., Rank, E., Kotz, S. A., Toepel, U., Besson, M., Schirmer, A., and Friederici, A. D. (2003). Affective encoding in the speech signal and in event-related brain potentials. *Speech Communication*, 40(1–2):61–70.

[Amir and Ron, 1998] Amir, N. and Ron, S. (1998). Towards an automatic classification of emotions in speech. In *Proceedings of the International Conference on Spoken Language Processing*, volume 3, pages 555–558, Sydney, Australia.

[Anderson et al., 1991] Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., and Weinert, R. (1991). The HCRC map task corpus. *Language and Speech*, 34:351–366.

[Athénes et al., 2002] Athénes, S., Averty, P., Puechmorel, S., Delahaye, D., and Collet, C. (2002). ATC complexity and controller workload: Trying to bridge the gap. In *International Conference on Human-Computer Interaction in Aeronautics (HCI-Aero)*, Cambridge, USA.

[Baber et al., 1996] Baber, C., Mellor, B., Graham, R., Noyes, J. M., and Tunley, C. (1996). Workload and the use of automatic speech recognition: The effects of time and resource demands. *Speech Communication*, 20(1-2):37–53.

[Ball and Breese, 1999] Ball, G. and Breese, J. (1999). Modeling the emotional state of computer users. In *Workshop on 'Attitude, Personality and Emotions in User-Adapted Interaction'*.

[Barbato et al., 1995] Barbato, G., Ficca, G., Beatrice, M., Casiello, M., Muscettola, G., and Rinaldi, F. (1995). Effects of sleep deprivation on spontaneous eye blink rate and alpha eeg power. *Biological Psychiatry*, 38(5):340–341.

[Bard et al., 1996] Bard, E. G., Sotillo, C., Anderson, A. H., Thompson, H. S., and Taylor, M. M. (1996). The DCIEM map task corpus: Spontaneous dialogue under sleep deprivation and drug treatment. *Speech Communication*, 20(1-2):71–84.

[Bittner et al., 2000] Bittner, R., Smrcka, P., Vysoký, P., Haná, K., Pousek, L., and Scheib, P. (2000). Detecting of fatigue states of a car driver. In Brause, R. and Hanisch, E., editors, *Proceedings of Medical Data Analysis: First International Symposium, ISMDA*, volume 1933 of *Lecture Notes in Computer Science*, pages 260–274. Springer-Verlag, Heidelberg, Frankfurt, Germany.

[Bolia and Slyh, 2003] Bolia, R. S. and Slyh, R. E. (2003). Perception of stress and speaking style for selected elements of the SUSAS database. *Speech Communication*, 40(4):493–501.

[Bonner and Wilson, 2001] Bonner, M. A. and Wilson, G. F. (2001). Heart rate measures of flight test and evaluation. *International Journal of Aviation Psychology*, 12(1):63–77.

[Bou-Ghazale and Hansen, 1998] Bou-Ghazale, S. and Hansen, J. (1998). HMM-based stressed speech modeling with application to improved synthesis and recognition of isolated speech under stress. *IEEE Transactions on Speech and Audio Processing*, 6(3):201–216.

[Bou-Ghazale and Hansen, 2000] Bou-Ghazale, S. and Hansen, J. (2000). A comparative study of traditional and newly proposed features for recognition of speech under stress. *IEEE Transactions on Speech and Audio Processing*, 8(4):429–442.

[Bou-Ghazale and Hansen, 1996] Bou-Ghazale, S. E. and Hansen, J. H. L. (1996). Generating stressed speech from neutral speech using a modified celp vocoder. *Speech Communication*, 20(1-2):93–110.

[Bouraoui et al., 2003] Bouraoui, J.-L., Bothorel, G., and Vigouroux, N. (2003). Transcription and annotation of an apprenticeship corpus: Application to the correction and self-correction strategies. In *Proc. Error Handling in Spoken Dialogue Systems*, pages 35–40, Château d'Oex, Vaud, Switzerland.

[Brookings et al., 1996] Brookings, J. B., Wilson, G. F., and Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, 42(3):361–377.

[Byrne and Parasuraman, 1996] Byrne, E. A. and Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology*, 42(3):249–268.

[Cahn, 1990] Cahn, J. E. (1990). The generation of affect in synthesized speech. *Journal of the American Voice I/O Society*, 8:1–19.

[Cairns and Hansen, 1994] Cairns, D. A. and Hansen, J. H. L. (1994). Nonlinear analysis and classification of speech under stressed conditions. *Journal of the Acoustical Society of America*, 96(6):3392–3400.

[Chen, 1988] Chen, Y. (1988). Cepstral domain talker stress compensation for robust speech. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(4):433–439.

[Costa, 1995] Costa, G. (1995). Occupational dfstress and stress prevention in air traffic control. Working paper. International Labor Office, Geneva, Italy.

[Cowie and Cornelius, 2003] Cowie, R. and Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech Communication*, 40(1-2):5–32.

[Cowie et al., 2001] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1):32–80.

[Cox et al., 2000] Cox, T., Griffiths, A., and Rial-González, E. (2000). *Research on Work-related Stress*. European Agency for Safety and Health at Work, Luxembourg. ISBN 92-828-9255-7.

[Cummings and Clements, 1990] Cummings, K. and Clements, M. (1990). Analysis of glottal waveforms across stress styles. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 369–372, Albuquerque, NM USA.

[Cummings and Clements, 1992] Cummings, K. and Clements, M. (1992). Improvements to and applications of analysis of stressed speech. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 25–28, San Francisco, CA , USA.

[Cummings and Clements, 1995] Cummings, K. and Clements, M. (1995). Analysis of the glottal excitation of emotionally styled and stressed speech. *Journal of the Acoustical Society of America*, 98(1):88–98.

[David, 2001] David, H. (2001). Measuring people in control [air traffic controllers]. In *Human Interfaces in Control Rooms, Cockpits and Command Centres, 2001. People in Control. The Second International Conference on*, pages 285 – 290, Manchester UK.

[de la Torre et al., 2003] de la Torre, F., Rubio, C., and Martinez, E. (2003). Subspace eyetracking for driver warning. In *Proceedings of International Conference on Image Processing*, volume 2, pages III – 329–32.

[de Waard, 1996] de Waard, D. (1996). *The measurement of drivers' mental workload*. PhD thesis, University of Groningen. Haren, The Netherlands: University of Groningen, Traffic Research Centre.

[Desmond and Hoyes, 1996] Desmond, P. A. and Hoyes, T. W. (1996). Workload variation, intrinsic risk and utility in a simulated air traffic control task: Evidence for compensatory effects. *Safety Science*, 22(1-3):87–101.

[Ekman, 1992] Ekman, P. (1992). Are there basic emotions? *Psychological Review*, 99(3):550–553.

[Eriksson and Papanikotopoulos, 1997] Eriksson, M. and Papanikotopoulos, N. (1997). Eye-tracking for detection of driver fatigue. In *Proceedings of IEEE Conference on Intelligent Transportation System, ITSC 97*, pages 314 – 319, Boston, MA USA.

[Farmer et al., 2003] Farmer, E., Brownson, A., and QinetiQ (2003). Review of workload measurement, analysis and interpretation methods. Technical Report CARE-Integra-TRS-130-02-WP2, EUROCONTROL.

[Fernandez and Picard, 2003] Fernandez, R. and Picard, R. W. (2003). Modeling drivers' speech under stress. *Speech Communication*, 40(1-2):145–159.

[Fournier et al., 1999] Fournier, L. R., Wilson, G. F., and Swain, C. R. (1999). Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training. *International Journal of Psychophysiology*, 31(2):129–145.

[Gopalan, 2001] Gopalan, K. (2001). On the effect of stress on certain modulation parameters of speech. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 101–104, Salt Lake City, UT USA.

[Grace et al., 1998] Grace, R., Byrne, V., Bierman, D., Legrand, J.-M., Gricourt, D., Davis, B., Staszewski, J., and Carnahan, B. (1998). A drowsy driver detection system for heavy vehicles. In *Proceedings of the 17th Digital Avionics Systems Conference (DASC)*, volume 2, pages I36/1 – I36/8, Bellevue, WA USA.

[Hansen and Bou-Ghazale, 1997] Hansen, J. and Bou-Ghazale, S. (1997). Getting started with SUSAS: A speech under simulated and actual stress database. In *Proceedings of the European Conference on Speech Communication and Technology*, volume 4, pages 1743–1746, Rhodes, Greece.

[Hansen et al., 2000] Hansen, J., Swail, C., South, A., Moore, R., Steeneken, H., Cupples, E., Anderson, T., Vloeberghs, C., Trancoso, I., and Verlinde, P. (2000). The impact of speech under 'stress' on military speech technology. Technical Report AC/323(IST)TP/5 IST/TG-01, NATO Research & Technology Organization RTO-TR-10.

[Hansen, 1998] Hansen, J. H. (1998). NATO IST-03– Speech under stress homepage. http://cslr.colorado.edu/rspl/stress.html.

[Hansen et al., 1998] Hansen, J. H., Bou-Ghazale, S. E., Sarikaya, R., and Pellom, B. (1998). Getting started with the SUSAS: Speech under simulated and actual stress database. Technical Report RSPL-98-10, Robust Speech Processing Laboratory, Duke University.

[Hansen, 1996] Hansen, J. H. L. (1996). Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition. *Speech Communication*, 20(1-2):151–173.

[Hilburn and Jorna, 2001] Hilburn, B. and Jorna, P. G. (2001). Workload and air traffic control. In Hancock, P. and Desmond, P., editors, *Stress, Workload & Fatigue*, pages 384–394. Lawrence Erlbaum Associates, Mahwah, N.J.

[Huang et al., 1998] Huang, T., Chen, L., and Tao, H. (1998). Bimodal emotion recognition by man and machine. In *ATR Workshop on Virtual Communication Environments*, Kyoto, Japan.

[Huber et al., 2000] Huber, R., Batliner, A., Buckow, J., Noth, E., Warnke, V., and Niemann, H. (2000). Recognition of emotion in a realistic dialogue scenario. In *Proceedings of the International Conference on Spoken Language Processing*, pages 665–668, Beijing, China.

[Ji et al., 2004] Ji, Q., Zhu, Z., and Lan, P. (2004). Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE Transactions on Vehicular Technology*, 53(4):1052–1068.

[Johnstone and Scherer, 1999] Johnstone, T. and Scherer, K. R. (1999). The effects of emotions on voice quality. In *Proceedings of the International Congress of Phonetic Sciences*, pages 2029–2032.

[Kaiser, 1993] Kaiser, J. (1993). Some useful properties of Teager's energy operators. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 149–152, Minneapolis, USA.

[Kang et al., 2000] Kang, B.-S., Han, C.-H., Lee, S.-T., Youn, D.-H., and Lee, C. (2000). Speaker dependent emotion recognition using speech signals. In *Proceedings of the International Conference on Spoken Language Processing*, pages 383–386.

[Laine et al., 2002] Laine, T., Bauer, K., Lanning, J., Russell, C., and Wilson, G. (2002). Selection of input features across subjects for classifying crewmember workload using artificial neural networks. *IEEE Transactions on Systems, Man and Cybernetics, Part A,*, 32(6):691–704.

[Lang, 1995] Lang, P. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5):372–385.

[Li and Zhao, 1998] Li, Y. and Zhao, Y. (1998). Recognizing emotions in speech using short-term and long-term features. In *Proceedings of the International Conference on Spoken Language Processing*, volume 6, pages 2255–2258, Sydney, Australia.

[Lively et al., 1993] Lively, S. E., Pisoni, D. B., Van Summers, W., and Bernacki, R. H. (1993). Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *Journal of the Acoustical Society of America*, 93(5):2962–73.

[Majumdar and Ochieng, 2002] Majumdar, A. and Ochieng, W. Y. (2002). The factors affecting air traffic controller workload: A multivariate analysis based upon simulation modelling of controller workload. Technical report, Centre for Transport Studies Department of Civil and Environmental Engineering Imperial College of Science, Technology and Medicine, London.

[Marshall et al., 2003] Marshall, S., Pleydell-Pearce, C., and Dickson, B. (2003). Integrating psychophysiological measures of cognitive workload and eye movements to detect strategy shifts. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, pages 130–135, Hawaii, US.

[McCall et al., 2003] McCall, J., Mallick, S., and Trivedi, M. (2003). Real-time driver affect analysis and tele-viewing system. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, pages 372–377.

[Mellor and Graham, 1995] Mellor, B. and Graham, R. (1995). A corpus of read car number plates recorded under conditions of cognitive stress. Defence Evaluation and Research Agency (DERA), Malvern, UK. CD-ROM.

[Metaxas et al., 2004] Metaxas, D., Venkataraman, S., and Vogler, C. (2004). Image-based stress recognition using a model-based dynamic face tracking system. In Bubak, M. et al., editors, *4th International Conference Computational Science, Workshop on Dynamic Data Driven Applications Systems*, volume 3038 of *Lecture Notes in Computer Science*, pages 813–821. Springer-Verlag Heidelberg, Kraków, Poland.

[Morris and Miller, 1996] Morris, T. L. and Miller, J. C. (1996). Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology*, 42(3):343–360.

[Murata and Iwase, 1998] Murata, A. and Iwase, H. (1998). Evaluation of mental workload by fluctuation analysis of pupil area. In *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 6, pages 3094–3097, Hong Kong, China.

[Murray et al., 1996a] Murray, I. R., Arnott, J. L., and Rohwer, E. A. (1996a). Emotional stress in synthetic speech: Progress and future directions. *Speech Communication*, 20(1-2):85–91.

[Murray et al., 1996b] Murray, I. R., Baber, C., and South, A. (1996b). Towards a definition and working model of stress and its effects on speech. *Speech Communication*, 20(1-2):3–12.

[Nwe et al., 2003a] Nwe, T. L., Foo, S. W., and De Silva, L. (2003a). Classification of stress in speech using linear and nonlinear features. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages II – 9–12.

[Nwe et al., 2003b] Nwe, T. L., Foo, S. W., and De Silva, L. (2003b). Detection of stress and emotion in speech using traditional and FFT based log energy features. In *Fourth Pacific Rim Conference on Multimedia, Information, Communications and Signal Processing*, volume 3, pages 1619–1623.

[Orlikoff, 1990] Orlikoff, R. F. (1990). Vowel amplitude variation associated with the heart cycle. *Journal of the Acoustical Society of America*, 88(5):2091–2098.

[Orlikoff and Baken, 1989a] Orlikoff, R. F. and Baken, R. (1989a). The effect of the heartbeat on vocal fundamental frequency perturbation. *Journal of Speech and Hearing Research*, 32(3):576–82.

[Orlikoff and Baken, 1989b] Orlikoff, R. F. and Baken, R. (1989b). Fundamental frequency modulation of the human voice by the heart beat: Preliminary results and possible mechanims. *Journal of the Acoustical Society of America*, 85(2):888–893.

[Ortony and Turner, 1990] Ortony, A. and Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, 97(3):315–331.

[Pantic and Rothkrantz, 2003] Pantic, M. and Rothkrantz, L. J. M. (2003). Toward an affect-sensitive multimodal human–computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390.

[Pereira, 2000] Pereira, C. (2000). Dimensions of emotional meaning in speech. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, pages 1–4, Belfast, Northern Ireland.

[Petrushin, 2000] Petrushin, V. A. (2000). Emotion recognition in speech signal: Experimental study, development, and application. In *Proceedings of the International Conference on Spoken Language Processing*, pages 222–225.

[Protopapas and Lieberman, 1995] Protopapas, A. and Lieberman, P. (1995). Effects of vocal f0 manipulations on perceived emotional stress. In *Proceedings of the ESCA/NATO Tutorial and Research Workshop on Speech Under Stress*, pages 1–4, Lisbon, Portugal.

[Protopapas and Lieberman, 1997] Protopapas, A. and Lieberman, P. (1997). Fundamental frequency of phonation and perceived emotional stress. *Journal of the Acoustical Society of America*, 101(4):2267–2277.

[Rahurkar and Hansen, 2003] Rahurkar, M. A. and Hansen, J. H. (2003). Frequency distribution based weighted sub-band approach for classification of emotional/stressful content in speech. In *Proceedings of the European Conference on Speech Communication and Technology*, pages 721–724.

[Rahurkar et al., 2002] Rahurkar, M. A., Hansen, J. H. L., Meyerhoff, J., Saviolakis, G., and Koenig, M. (2002). Frequency band analysis for stress detection using a teager energy operator based feature. In *Proceedings of the International Conference on Spoken Language Processing*, pages 2021–2024, Denver, Colorado, USA.

[Russell, 1994] Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? *Psychol. Bull.*, 115(1):102–141.

[Ruzanski et al., 2005] Ruzanski, E., Hansen, J. H., Meyerhoff, J., Saviolakis, G., and Koenig, M. (2005). Effects of phoneme characteristics on TEO feature-based automatic stress detection in speech. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume I, pages 357–360, Philadelphia, USA.

[Sarikaya and Hansen, 2000] Sarikaya, R. and Hansen, J. (2000). High resolution speech feature parametrization for monophone-based stressed speech recognition. *IEEE Signal Processing Letters*, 7(7):182–185.

[Scerbo et al., 2001] Scerbo, M. W., Freeman, F. G., Mikulka, P. J., Parasuraman, R., Nocero, F. D., and III, L. J. P. (2001). The efficacy of psychophysiological measures for implementing adaptive technology. Technical Report NASA/TP-2001-211018, NASA.

[Scherer, 2003] Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1-2):227–256.

[Scherer et al., 2002] Scherer, K. R., Grandjean, D., Johnstone, T., Klasmeyer, G., and Bänziger, T. (2002). Acoustic correlates of task load and stress. In *Proceedings of the International Conference on Spoken Language Processing*, pages 2017–2020, Denver, Colorado, USA.

[Sloan et al., 1991] Sloan, R. P., Korten, J. B., and Myers, M. M. (1991). Components of heart rate reactivity during mental arithmetic with and without speaking. *Physiology & Behavior*, 50(5):1039–1045.

[Traunmüller, 2000] Traunmüller, H. (2000). Evidence for demodulation in speech perception. In *Proceedings of the 6th ICSLP*, volume 3, pages 790–793, Beijing, China.

[Turner and Ortony, 1992] Turner, T. J. and Ortony, A. (1992). Basic emotions: Can conflicting criteria converge? *Psychological Review*, 99(3):556–571.

[van Roon et al., 2004] van Roon, A. M., Mulder, L. J., Althaus, M., and Mulder, G. (2004). Introducing a baroreflex model for studying cardiovascular effects of mental workload. *Psychophysiology*, 41(6):961–981.

[Veltman, 2002] Veltman, J. A. (2002). A comparative study of psychophysiological reactions during simulator and real flight. *International Journal of Aviation Psychology*, 12(1):33–48.

[Veltman and Gaillard, 1998] Veltman, J. A. and Gaillard, A. W. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, 41(5):656–69.

[Veltman and Gaillard, 1996] Veltman, J. A. and Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology*, 42(3):323–342.

[Verschuere et al., 2004] Verschuere, B., Crombez, G., Clercq, A. D., and Koster, E. H. W. (2004). Autonomic and behavioral responding to concealed information: differentiating orienting and defensive responses. *Psychophysiology*, 41(3):461–6.

[Vincent et al., 1996] Vincent, A., Craik, F. I. M., and Furedy, J. J. (1996). Relations among memory performance, mental workload and cardiovascular responses. *International Journal of Psychophysiology*, 23(3):181–198.

[Waters et al., 1994] Waters, J., Nunn, S., Gillcrist, B., and VonColln, E. (1994). Detection of stress by voice: Analysis of the glottal pulse. Technical Report NRaD TR 1652, Space and Naval Warfare Systems Center San Diego.

[Whitmore and Fisher, 1996] Whitmore, J. and Fisher, S. (1996). Speech during sustained operations. *Speech Communication*, 20(1-2):55–70.

[Wilson, 2001] Wilson, G. F. (2001). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *International Journal of Aviation Psychology*, 12(1):3–18.

[Wilson and Russell, 2003] Wilson, G. F. and Russell, C. A. (2003). Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Human Factors*, 45(3):381–9.

[Womack and Hansen, 1999] Womack, B. and Hansen, J. (1999). N-channel hidden markov models for combined stressed speech classification and recognition. *IEEE Transactions on Speech and Audio Processing*, 7(6):668–677.

[Womack and Hansen, 1996] Womack, B. D. and Hansen, J. H. L. (1996). Classification of speech under stress using target driven features. *Speech Communication*, 20(1-2):131–150.

[Zeier et al., 1996] Zeier, H., Brauchli, P., and Joller-Jemelka, H. I. (1996). Effects of work demands on immunoglobulin a and cortisol in air traffic controllers. *Biological Psychology*, 42(3):413–423.

[Zhao et al., 2000] Zhao, L., Lu, W., Jiang, Y., and Wu, Z. (2000). A study on emotional feature recognition in speech. In *Proceedings of the International Conference on Spoken Language Processing*, pages 961–964.

[Zhou et al., 2001] Zhou, G., Hansen, J., and Kaiser, J. (2001). Nonlinear feature based classification of speech under stress. *IEEE Transactions on Speech and Audio Processing*, 9(3):201–216.

Appendix

A. Stress Analysis using Non-Speech Methods

The focus of this report was stress observation by analyzing human speech. As mentioned in section 2, speech is only one among many human parameters which change under the influence of stress. The big advantage of speech in the context of ATC is the availability of the signal without any additional sensors and that it is both non-invasive and non-intrusive.

Other methods, which are intrusive and some even invasive ones may offer a better reliability of the stress estimation, though. One has also to take into account that different measures are most sensitive in specific areas of demand. For example the cardiac measures heart rate (HR) and heart rate variability (HRV) are most sensitive in the regions of low demand and high demand, when the performance starts to decrease, whereas the 0.10 Hz band of the HRV is most sensitive in the region where increased effort is necessary to keep performance at a high level [de Waard, 1996, Ch.6]. To our knowledge, for speech no research results on similar specific effects exist until now.

To offer a more complete overview on the subject of human stress observation non-speech methods are briefly introduced in the following section. This section does not claim to be exhaustive, but should only give an overview of existing methods and their potential and problems. Where possible an emphasis will be given on literature focusing on air traffic control and workload induced stress.

Summary Literature An overview over a broader view of physiological measures for workload observation can be found in several papers. [Farmer et al., 2003] introduces different physiological measures and gives a brief analysis on reliability, validity, sensitivity, diagnosticity, practicality and intrusiveness and gives a reference paper for each method.

In [David, 2001] a table is given where a comparison of several measurements is given. Unfortunately, no discussion and background is provided.

A NASA study [Scerbo et al., 2001] gives an overview over the efficacy of physiological measures for adaptive technology [Byrne and Parasuraman, 1996]. They cover eye analysis, cardio-vascular measures, EEG, respiration and speech. They give an analysis on sensitivity, diagnosticity, ease of use, current real world/real time feasibility, cost and intrusiveness for each measure.

A short literature review on psychophysiological measures is given in [Laine et al., 2002]. Additionally an emphasis is placed on the selection of the most useful features for classification with an ANN.

We will start out with an overview of stress analysis based on visual analysis. Then we will look at cardiovascular measures followed by introduction of electromagnetic activity in the brain in relation to workload stress. Finally, we will also mention methods such as electrodermal activity, respiration, and the analysis of body fluids.

A.1. Visual Analysis

Visual analysis is, like speech, a contact-free method, but unlike speech, for ATCO applications an additional sensor, i.e., a camera has to be used. Another disadvantage is that the utilization of a camera might give the impression of being monitored at the work place and, therefore, induce some kind of intimidation. For camera-based approaches, problems

can arise when the system has to deal with real-life situations, such a difficult lighting or eye-glasses, which can cause reflections, or may be opaque.

Eye analysis: The eyes are influenced by stress, specially if the stress is induced by visual workload. There are several parameters, which describe the eye activity. Those used frequently are *eye movement*, *blink rate*, *eyelid closure rate*, *eye fixation*, *fluctuation of the pupil area* and derivatives of those features [Barbato et al., 1995, Bittner et al., 2000, Brookings et al., 1996, de la Torre et al., 2003, Eriksson and Papanikotopoulos, 1997, Murata and Iwase, 1998, Veltman and Gaillard, 1996, de Waard, 1996, Wilson, 2001]. Blink rate is reported to be correlated with *visual* demands, it decreases as visual demands increase. So it might be useful for ATCO monitoring [Laine et al., 2002, Wilson, 2001].

Very often eye analysis is used for driver fatigue monitoring. For a good state-of-the-art summary of this subject and a system, which does real-time monitoring and prediction of driver fatigue see [Ji et al., 2004]. See also [Bittner et al., 2000, Eriksson and Papanikotopoulos, 1997, Grace et al., 1998, de la Torre et al., 2003] for further reports on research.

Electrooculogram (EOG): This is not exactly a visual analysis method, but since it is a way of analyzing the eye movement it is included in this section. The EOG allows to measure eye parameters by analyzing electric activity picked up with electrodes placed around the eyes. The derived features are very similar to the above mentioned ones obtained by visual analysis including measures of blink rate, blink duration, long closure rate, blink amplitude, saccade velocity, saccade rate [Marshall et al., 2003, Morris and Miller, 1996]. [Fournier et al., 1999] and [Wilson and Russell, 2003] found a dependence of eye blink rate on different levels of workload.

Face Analysis: [McCall et al., 2003] analyze the whole face to derive features to detect the emotional state of a subject. [Metaxas et al., 2004] present a method to detect stress by analyzing sequences of facial images.

A.2. Cardiovascular Activity

Cardiovascular analysis is a very well established method of relating physiological measures to workload and stress. A number of papers have been published also for the application with ATC and aviation in general. A drawback for cardiovasular parameters is that they are influenced by other factors, e.g. physical activity. [Sloan et al., 1991] report influence of speaking on electrocardiogram (ECG) measures. [van Roon et al., 2004] propose a detailed model to describe coherent and time-dependent cardiovascular changes occurring during mental task performance.

Electrocardiogram (ECG): The cardiovascular activity is usually measured using an ECG, where electrodes are applied on the chest. Several parameters can be derived from the electrode signals, such as HR, HRV, interbeat intervals (IBI). Those measures have been reported to be useful both as an indicator for fatigue and workload [Bonner and Wilson, 2001, Brookings et al., 1996, Veltman and Gaillard, 1998, Veltman, 2002, Vincent et al., 1996, Wilson, 2001].

Blood pressure: Blood pressure can be measured with a tonometer, which measures the systolic and diastolic values with cuffs placed around the finger, though a high risk of failure is reported, when the subject bends its fingers. Blood pressure is reported to

rise with increased mental activity [Veltman and Gaillard, 1996, Veltman and Gaillard, 1998, Veltman, 2002, Vincent et al., 1996]

A.3. Brain Activity

Another method for workload monitoring is the measurement of the electrical signals emitted by the brain. The EEG records the signals using several electrodes at different locations on the scalp. Due to the large amount of data and the low signal-to-noise ratio (SNR) further processing is necessary to be able to enhance the interpretation of the measurement. Analysis of variance (ANOVA), principle components analysis (PCA) and ANNs are among the methods used (see [Laine et al., 2002] and [Brookings et al., 1996, Fournier et al., 1999, Wilson, 2001]). In [Wilson and Russell, 2003] a classification is done on two different workload qualities (volume, complexity) in three levels (low, medium, high), each plus overload. Classifiation rates using an ANN with 85 EEG features plus an ECG, an electrooculogram (EOG) and a respiration feature were more than 70 % for the different workload states (see Tab. 7).

An obvious disadvantage of EEG based methods is the rather complicated measurement procedure, since a lot of electrodes have to be applied to the human body, which makes it rather impractical for everyday continuous workload monitoring.

A.4. Others

Several other methods are reported, but none of them has been as widely been studied as the methods mentioned above.

Electrodermal activity (EDA): EDA refers to the changes of the electrical properties of the skin. Research on workload related stress is rather limited, though in [Wilson, 2001] promising results are reported. More research is done in association with emotion related stress as used for forensic applications [Verschuere et al., 2004, de Waard, 1996].

Respiration: Respiration can be related to the need of energy, which is also related to mental workload. In the context of workload measurement it is measured with belts around the chest and the abdomen, so the variation of the respiration volume can be calculated. Derived features are the frequency and the amplitude of the respiration [Brookings et al., 1996, Fournier et al., 1999, Veltman, 2002].

Body Fluids: Another possibility is to analyze human saliva to measure stress related substances like cortisol. A correlation between the cortisol concentration and the workload was reported [Zeier et al., 1996, Veltman, 2002]. However, no realtime measurement is possible, since taking the saliva interrupts the workflow and, therefore, cannot be done in very short intervals.

Table 7: *Percentage correct classification for different workload conditions using ANN classification with 88 features (from [Wilson and Russell, 2003]).*

	Volume			Complexity			
	Low	Medium	High	Low	Medium	High	Overload
ANN	83	74	79	82	71	80	91

B. Acronyms

AM *amplitude modulation*

ANN *artificial neural network*

ANOVA *analysis of variance*

ASR *automatic speech recognition*

ATC *air traffic control*

ATCO *air traffic control operator*

CELP *code excited linear prediction*

DCT *discrete cosine transform*

DERA *Defence Evaluation Research Agency*

ECG *electrocardiogram.*
Measures the cardiac cycle

EDA *electrodermal activity*

EEG *electroencephalogram.*
Measures electrical activity of the brain

EMG *electromyogram.*
Measures electrical activity of a muscle

EOG *electrooculogram.*
Electrodes around eyes, can detect eye movement, closure, blink rate

F_0 *fundamental frequency*

FM *frequency modulation*

FFT *fast Fourier transform*

HMM *hidden Markov model*

HR *heart rate*

HRV *heart rate variability*

IBI *interbeat intervals*

ICAO *International Civil Aviation Organization*

LDC *Linguistic Data Consortium*

LFPC *log frequency power coefficients*

LPC *linear predictive coding*

LPCC *linear prediction cepstral coefficient*

MFCC *mel frequency cepstral coefficient*

MLP *multi layer perceptron.*
A neural network using multiple layers of neurons.

NATO *North Atlantic Treaty Organization*

NASA *National Aeronautics and Space Administration*

PERCLOS *percentage of slow-eye closure over one minute*

PCA *principle components analysis*

R/T *radio telephony*

SNR *signal-to-noise ratio*

SUSAS *speech under simulated and actual stress*

SVM *support vector machine*

TEO *Teager energy operator*

TTS *text to speech*

VDT *video display terminal*

C. Overview of Speech Databases

Table 8: *Databases of speech under stress. Technical detail summary. Adapted from [Hansen et al., 2000]*

	SUSC-0			DERA	SUSAS	ATC	DCIEM
	Fighter Cockpit & Controller	Aircraft Lost in Fog	F-16 Engine Out	License Plates	Computer Tracking Task	apprentice-ship	Map task
Stressor	Workload	Anxiety	Anxiety Exertion	Pressure	Time Pressure Workload	Workload Pressure	Sleep Deprivation
Stressor Type	Psych.	Psych.	Psych.	Psych.	Psych.	Psych.	Physiol.
Stressor Order	3	3	3	3	3	3	1
Language	English and Dutch	Mostly Dutch, Some English	US English	UK English	US English	French and English	Canadian English
Task	Command and Control	Fighter Cockpit	Fighter Cockpit Air Traffic Control	Prompted Phrases	Isolated Words	Air Traffic Control	Map Task
Annotated stress level	Yes			Yes	Partly	No	Yes
Quantity	3 hours	23 mins	15 mins	2 hours	50min	19 hours	12 hours
Number of Speakers	9	2	3	16	8	32	35
Gender	Male	Male	Male	4 Female 11+1 Male	4 Female 4 Male	Male	33 Male 2 Female
Native Language	Dutch	Dutch	US English	UK English	US English	French	Canadian English
Population	Air Force Controllers	Pilots Pilots	Pilots and ATC	Researchers	Graduate Students	ATC apprentice Pseudo-Pilots	Military reservists
Microphone		Oxygen Mask	Oxygen Mask	Shure SM-10	Shure 512		Shure SM-10A
Sampling Rate	16 kHz	16 kHz	16 kHz	20 kHz	8kHz	16kHz	20kHz
Recording Quality	Good	Poor	Poor	Good	Good	Poor	Good

D. Extended Literature References

[1] Jans Aasman, Gijsbertus Mulder, and Lambertus J. M Mulder. Operator effort and the measurement of heart-rate variability. *Human Factors*, 29(2):161–170, 1987.

Abstract: This paper discusses the usefulness of heart-rate variability (sinus arrhythmia) as an index of operator effort. Effort is involved if task performance requires the use of attention-demanding, controlled processing. This form of processing heavily uses a capacity-limited mechanism: working memory. Effort is also required if the current state of the subject deviates from the target or task optimal state because of fatigue, circadian rhythm, time of day, sleep deprivation, time on task, drugs, heat, or noise. Effort is involved whenever an attempt to resolve mismatch of target and current state takes the form of active manipulation of cognitive resources. We argue that spectral analysis of sinus arrhythmia is useful to obtain more insight into the physiological mechanisms that underlie heart-rate variability. In the present study, it is shown that the amplitude of the 0. 10-Hz component of this signal systematically decreases as the load on working memory increases. Some suggestions are made to apply the methods and techniques discussed in this investigation

[2] M.G. Abbott. Study into the detection of stress in air traffic control speech. Technical Report DERA/S&P/SPI/374/047/01/STECR1, Defence Evaluation Research Agency (DERA), 1998.

[3] Kai Alter, Erhard Rank, Sonja A. Kotz, Ulrike Toepel, Mireille Besson, Annett Schirmer, and Angela D. Friederici. Affective encoding in the speech signal and in event-related brain potentials. *Speech Communication*, 40(1–2):61–70, April 2003.

Abstract: A number of perceptual features have been utilized for the characterization of the emotional state of a speaker. However, for automatic recognition suitable objective features are needed. We have examined several features of the speech previous termsignal innext term relation to accentuation and traces of previous termevent-related brain potentialsnext term (ERPs) during previous termaffectivenext term speech perception. Concerning the features of the speech previous termsignalnext term we focus on measures related to breathiness and roughness. The objective measures used were an estimation of the harmonics-to-noise ratio, the glottal-to-noise excitation ratio, a measure for spectral flatness, as well as the maximum prediction gain for a speech production model computed by the mutual information function and the ERPs. Results indicate that previous terminnext term particular the maximum prediction gain shows a good differentiation between neutral and non-neutral emotional speaker state. This differentiation is partly comparable to the ERP results that show a differentiation of neutral, positive and negative affect. Other objective measures are more related to accentuation than to emotional state of the speaker

[4] N. Amir and S. Ron. Towards an automatic classification of emotions in speech. In *Proceedings of the International Conference on Spoken Language Processing*, volume 3, pages 555–558, Sydney, Australia, 1998.

Abstract: In this paper we discuss a method for extracting emotional state from the speech signal. We describe a methodology for obtaining emotive speech. and a method for identifying the emotion present in the signal. This method is based on analysis of the signal over sliding windows. and extracting a representative parameter set. A set of basic emotions is defined, and for each such emotion a reference point is computed. At each instant the distance of the measured parameter set from the reference points is calculated. and used to compute a fuzzy membership index for each emotion, which we term the "emotional index". Preliminary results are presented, which demonstrate the discriminative abilities of this method when applied to a number of speakers

[5] A. Anderson, M. Bader, E. Bard, E. Boyle, G. M. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson, and R. Weinert. The HCRC map task corpus. *Language and Speech*, 34:351–366, 1991.

[6] Jeremy Ang, Rajdip Dhillon, Ashley Krupski, Elizabeth Shriberg, and Andreas Stolcke. Prosody-based automatic detection of annoyance and frustration in human-computer dialog. In *Proceedings of the International Conference on Spoken Language Processing*, pages 2037–2040, Denver, Colorado, USA, September 16-20 2002.

Abstract: We investigate the use of prosody for the detection of frustration and annoyance in natural human-computer dialog. In addition to prosodic features, we examine the contribution of language model information and speaking "style". Results show that a prosodic model can predict whether an utterance is neutral versus "annoyed or frustrated" with an accuracy on par with that of human interlabeler agreement. Accuracy increases when discriminating only "frustrated" from other utterances, and when using only those utterances on which labelers originally agreed. Furthermore, prosodic model accuracy degrades only slightly when using recognized versus true words. Language model features, even if based on true words, are relatively poor predictors of frustration. Finally, we find that hyperarticulation is not a good predictor of emotion; the two phenomena often occur independently

[7] Luigi Anolli and Rita Ciceri. The voice of deception: Vocal strategies of naive and able liars. *Journal of Nonverbal Behavior*, 21(4):259–284, 1997.

Abstract: The aim of this research was to analyze the main vocal cues and strategies used by a liar. 31 male university students were asked to raise doubts in an expert in law about a picture. The subjects were required to describe the picture in three experimental conditions: telling the truth (T) and lying to a speaker when acquiescent (L1) and when suspicious (L2). The utterances were then subjected to a digitized acoustic analysis in order to measure nonverbal vocal variables. Verbal variables were also analyzed (number of words, eloquency and disfluency index). Results showed that deception provoked an increment in F0, a greater number of pauses and words, and higher eloquency and fluency indexes. The F0 related to the two types of lie - prepared and unprepared - identified three classes of liars: good liars, tense liars (more numerous in L1), and overcontrolled liars (more numerous in L2). It is argued that these differences are correlated to the complex task of lying and the need to control one's emotions during deception. The liar's effort to control his/her voice, however, can lead to his/her tone being overcontrolled or totally lacking in control (leakage). Finally, the research forwards an explanation on the strategies used by the good liar and in particular treats the self-deception hypothesis

[8] Sylvie Athénes, Philippe Averty, Stéphane Puechmorel, Daniel Delahaye, and Christian Collet. ATC complexity and controller workload: Trying to bridge the gap. In *International Conference on Human-Computer Interaction in Aeronautics (HCI-Aero)*, Cambridge, USA, 2002.

Abstract: Considerable research has been conducted to identify a useful set of Air Traffic Control complexity factors. It is now necessary to determine, on the one hand, how these factors affect ATC complexity and controller workload and, on the other hand, how ATC complexity and controller workload interact. This line of research should lead to elaboration of guidelines to improve sector configuration and traffic flow as well as to produce automation tools and procedures to reduce controller workload This paper addresses the problem of formulating a functional relationship between ATC complexity and workload using a parameter reflecting intrinsic air traffic complexity -a measure of disorder of aircraft trajectories assumed to estimate cognitive difficulty-, a computed index, the Traffic Load Index, and psychophysiological parameters to characterize workload. Preliminary results concerning workload assessment are reported.

[9] P. Averty, S. Athenes, C. Collet, and A. Dittmar. Evaluating a new index of mental workload in real atc situation using psychophysiological measures. In *Proc. DASC*, volume 2, pages 7A4–1–7A4–13 vol.2, 2002.

Abstract: One way to improve both capacity and safety in air traffic control (ATC) implies to assist the human operator, while he remains in the decision loop. Designing high performance tools requires a precise tuning in order to interface them efficiently and non-intrusively with the controller, especially in heavy workload contexts. To this end, knowledge of workload sources and fluctuations is essential and has triggered a large number of studies. Surprisingly, there is no universally accepted definition of mental workload: a consensus exists merely to define it as the "cost" of a given task for the operator. Concept of mental load is related to information processing theory Limited capacity of cognitive resources and sequential processing of information define what can be overload. an excessive demand on perceptual and cognitive resources (visual and auditory perception, memory and attention, among others) with respect to information processing capacities. In the present study, we would like to propose a different way to model the link between aircraft configuration and perceived workload

[10] C. Baber, B. Mellor, R. Graham, J. M. Noyes, and C Tunley. Workload and the use of automatic speech recognition: The effects of time and resource demands. *Speech Communication*, 20(1-2):37–53, 1996.

Abstract: Previous research has indicated that workload can have an adverse effect on the use of speech recognition systems. In this paper, the relationship between workload and speech is discussed, and two studies are reported. In the first study, time-stress is considered. In the second study, dual-task performance is considered. Both studies show workload to significantly reduce recognition accuracy and user performance. The nature of the impairment is shown to differ between individuals and types of workload. Furthermore, it appears that workload affects the selection of words to use, the articulation of the words and the relationship between speaking to ASR and performing other tasks. It is proposed that speaking to ASR is, in itself, demanding and that as workload increases so the ability to perform the task within the limits required by ASR suffers

[11] R.W. Backs. Going beyond heart rate: autonomic space and cardiovascular assessment of mental workload. *International Journal of Aviation Psychology*, 5(1):25–48, 1995.

Abstract: Psychophysiological assessment of pilot mental workload using heart rate should be augmented with an autonomic space model of cardiovascular function. This model proposes that autonomic nervous system influences on the heart may change with psychological processing in ways that are not evident in heart rate. A method of mental-workload assessment was proposed that used multiple psychophysiological measures of cardiovascular responsivity to derive the underlying sympathetic and parasympathetic information needed to represent the autonomic space for heart rate. Principal-components analysis was used to extract Sympathetic and Parasympathetic components from heart period, residual heart period. respiratory sinus arrhythmia, and Traube-Hering-Mayer wave in three experiments that manipulated perceptual/central processing and physical task demands. This initial evaluation of the method concluded that the autonomic components were valid and that the components had greater diagnosticity, and for some manipulations greater sensitivity, than heart rate. These results support the contention that the Sympathetic and Parasympathetic components provided increased precision for mental-workload assessment

[12] Gene Ball and Jack Breese. Modeling the emotional state of computer users. In *Workshop on 'Attitude, Personality and Emotions in User-Adapted Interaction'*, 1999.

Abstract: We describe the structure of a Bayesian network designed to monitor the behavior of a user interacting with a conversational computer and use that information to estimate the user's emotional state. Our model of emotional state uses two discrete dimensions, valence (bad to good) and arousal (calm to excited), to capture the dominant aspects of physical emotional response. Emotion is expressed behaviorally in a variety of ways, including by linguistic choices, qualities of vocal expression, and movement. In this paper, we focus on vocal expression, and its correlation with emotional state, using the psychological literature to suggest the appropriate parameterization of our Bayesian model

[13] Giuseppe Barbato, Gianluca Ficca, Michele Beatrice, Margherita Casiello, Giovanni Muscettola, and Franco Rinaldi. Effects of sleep deprivation on spontaneous eye blink rate and alpha eeg power. *Biological Psychiatry*, 38(5):340–341, 1995.

[14] E. G. Bard, C. Sotillo, A. H. Anderson, H. S. Thompson, and M. M Taylor. The DCIEM map task corpus: Spontaneous dialogue under sleep deprivation and drug treatment. *Speech Communication*, 20(1-2):71–84, 1996.

Abstract: This paper describes a resource designed for the general study of spontaneous speech under the stress of sleep deprivation. It is a corpus of 216 unscripted task-oriented dialogues produced by normal adults in the course of a major sleep deprivation study. The study itself examined continuous task performance through baseline, sleepless and recovery periods by groups treated with placebo or one of two drugs (Modafinil, d-amphetamine) reputed to counter the effects of sleep deprivation. The dialogues were all produced while carrying out the route communication task used in the HCRC Map Task Corpus. Pairs of talkers collaborated to reproduce on one partner's schematic map a route preprinted on the other's Controlled differences between the maps and use of labelled imaginary locations limit genre, vocabulary and effects of real-world knowledge. The designs for the construction of maps and the allocation of subjects to maps make the corpus a controlled elicitation experiment. Each talker participated in 12 dialogues over the course of the study. Preliminary examinations of dialogue length and task performance measures indicate effects of drug treatment, sleep deprivation and number of conversational partners. The corpus is available to researchers interested in all levels of speech and dialogue analysis, in both normal and stressed conditions

[15] Roman Bittner, Pavel Smrcka, Petr Vysoký, Karel Hána, Lubomir Poušek, and Petr Scheib. Detecting of fatigue states of a car driver. In R.W. Brause and E. Hanisch, ed-

itors, *Proceedings of Medical Data Analysis: First International Symposium, ISMDA*, volume 1933 of *Lecture Notes in Computer Science*, pages 260–274. Springer-Verlag, Heidelberg, Frankfurt, Germany, 2000.

Abstract: This paper deals with research on fatigue states of car drivers on freeways and similar roads. All experiments are performed on-the-road. The approach is based on the assumption that fatigue indicators can be derived from driver+car system behaviour by measuring and processing appropriate factors. For our experiments we designed an array of devices to measure selected physiological and technical parameters. On the basis of experiments already performed and described in the literature, we selected signals that carry appropriate information about fatigue states of a driver. Two data sets are compared: the first set was obtained from alert drivers, and the second one from provably sleep-deprived drivers. We are trying to calibrate "fatigue states" of the driver through trained rater estimates, and to extract reliable symptoms from the measured signals; we are seeking relations between the symptoms and raters fatigue estimates. The results of experiments already performed indicate that the values of symptoms evolve in cycles, which is essential to take into account during design of classifiers in the future

[16] Robert S. Bolia and Raymond E. Slyh. Perception of stress and speaking style for selected elements of the SUSAS database. *Speech Communication*, 40(4):493–501, 2003.

Abstract: The Speech Under Simulated and Actual Stress (SUSAS) database is a collection of utterances recorded under conditions of simulated or actual stress, the purpose of which is to allow researchers to study the effects of stress and speaking style on the speech waveform. The aim of the present investigation was to assess the perceptual validity of the simulated portion of the database by determining the extent to which listeners classify its utterances according to their assigned labels. Seven listeners performed an eight-alternative, forced-choice response, judging whether monosyllabic or disyllabic words spoken by talkers from three different regional accent classes (Boston, General American, New York) were best classified as, clear, fast, loud, neutral, question, slow, or soft. Mean percentages of "correct" judgments were analysed using a 3 (regional accent class) x 2 (number of syllables) x 8 (speaking style) repeated measures analysis of variance. Results indicate that, overall, listeners correctly classify the utterances only 58% of the time, and that the percentage of correct classifications varies as a function of all three independent variables.

[17] Malcolm A. Bonner and Glenn F. Wilson. Heart rate measures of flight test and evaluation. *International Journal of Aviation Psychology*, 12(1):63–77, 2001.

Abstract: One of the goals of aircraft test and evaluation is to determine whether the crew can operate a new system safely and effectively. Because flying is a complex task, several measures are required to derive the best evaluation. This article describes the use of heart rate to augment the typical performance and subjective measures used in test and evaluation. Heart rate can be nonintrusively collected and provides additional information to the test team. Example data illustrate the nature of the results provided by heart rate during the test and evaluation of a transport aircraft. Comparison with subjective workload estimates shows discrepancies that provide valuable insights into the crews' responses to the demands of the test missions. Heart rate should be considered as an additional measure in the test and evaluation tool kit

[18] Sahar E. Bou-Ghazale and John H. L. Hansen. Generating stressed speech from neutral speech using a modified celp vocoder. *Speech Communication*, 20(1-2):93–110, 1996.

Abstract: The problem of speech modeling for generating stressed speech using a source generator framework is addressed in this paper. In general, stress in this context refers to emotional or task induced speaking conditions. Throughout this particular study, the focus will be limited to speech under angry, loud and Lombard effect (i.e., speech produced in noise) speaking conditions. Source generator theory was originally developed for equalization of speech under stress for robust recognition (Hansen, 1993, 1994). It was later used for simulated stressed training token generation for improved recognition (Bou-Ghazale, 1993; Bou-Ghazale and Hansen, 1994). The objective here is to generate stressed perturbed speech from neutral speech using a source generator framework previously employed for stressed speech recognition. The approach is based on (i) developing a mathematical model that provides a means for representing the change in speech production under stressed conditions for perturbation, and (ii) employing this framework in an isolated word speech processing system to produce emotional/stressed perturbed speech from neutral speech. A stress perturbation algorithm is formulated based on a CELP (code-excited linear prediction) speech synthesis structure. The algorithm is evaluated using four different speech feature perturbation sets. The stressed speech parameter evaluations from this study revealed that pitch is capable of reflecting the emotional state of the speaker, while formant information alone is not as good a correlate of stress. However, the combination

of formant location, pitch and gain information proved to be the most reliable indicator of emotional stress under a CELP speech model. Results from formal listener evaluations of the generated stressed speech show successful classification rates of 87% for angry speech, 75% for Lombard effect speech and 92% for loud speech

[19] S.E. Bou-Ghazale and J.H.L. Hansen. A source generator based modeling framework for synthesis of speech. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 664–667, Detroit, MI, USA, May9-12 1995.

Abstract: The objective of this paper is to formulate an algorithm to generate stressed synthetic speech from neutral speech using a source generator framework previously employed for stressed speech recognition. The following goals are addressed (i) identify the most visible indicators of stress as perceived by the listener in stressed speaking styles such as loud, Lombard effect and angry, (ii) develop a mathematical model for representing speech production under stressed conditions, and (iii) employ the above model to produce emotional/stressed synthetic speech from neutral speech. The stress modeling scheme is applied to an existing low-bit rate CELP speech coder in order to investigate (i) the coder's ability and limitations reproducing stressed synthetic speech, and (ii) our ability to perturb coded neutral speech parameters at the synthesis stage so that resulting speech is perceived as being under stress. Two stress perturbation algorithms are proposed and evaluated. Results from formal listener evaluations show that 87% of neutral perturbed speech was indeed perceived as stressed

[20] S.E. Bou-Ghazale and J.H.L Hansen. HMM-based stressed speech modeling with application to improved synthesis and recognition of isolated speech under stress. *IEEE Transactions on Speech and Audio Processing*, 6(3):201–216, 1998.

Abstract: A novel approach is proposed for modeling speech parameter variations between neutral and stressed conditions and employed in a technique for stressed speech synthesis and recognition. The proposed method consists of modeling the variations in pitch contour, voiced speech duration, and average spectral structure using hidden Markov models (HMMs). While HMMs have traditionally been used for recognition applications, here they are employed to statistically model the characteristics needed for generating pitch contour and spectral perturbation contour patterns to modify the speaking style of isolated neutral words. The proposed HMM models are both speaker and word-independent, but unique to each speaking style. While the modeling scheme is applicable to a variety of stress and emotional speaking styles, the evaluations presented focus on angry speech, the Lombard (1911) effect, and loud spoken speech in three areas. First, formal subjective listener evaluations of the modified speech confirm the HMMs ability to capture the parameter variations under stressed conditions. Second, an objective evaluation using a separately formulated stress classifier is employed to assess the presence of stress imparted on the synthetic speech. Finally, the stressed speech is also used for training and shown to measurably improve the performance of an HMM-based stressed speech recognizer

[21] S.E. Bou-Ghazale and J.H.L Hansen. A comparative study of traditional and newly proposed features for recognition of speech under stress. *IEEE Transactions on Speech and Audio Processing*, 8(4):429–442, 2000.

Abstract: It is well known that the performance of speech recognition algorithms degrade in the presence of adverse environments where a speaker is under stress, emotion, or Lombard (1911) effect. This study evaluates the effectiveness of traditional features in recognition of speech under stress and formulates new features which are shown to improve stressed speech recognition. The focus is on formulating robust features which are less dependent on the speaking conditions rather than applying compensation or adaptation techniques. The stressed speaking styles considered are simulated angry and loud. Lombard effect speech, and noisy actual stressed speech from the SUSAS database which is available on a CD-ROM through the NATO IST/TG-01 research group and LDC. In addition, this study investigates the immunity of the linear prediction power spectrum and fast Fourier transform power spectrum to the presence of stress. Our results show that unlike fast Fourier transform's (FFT) immunity to noise, the linear prediction power spectrum is more immune than FFT to stress as well as to a combination of a noisy and stressful environment. Finally, the effect of various parameter processing such as fixed versus variable preemphasis, liftering, and fixed versus cepstral mean normalization are studied. Two alternative frequency partitioning methods are proposed and compared with traditional mel-frequency cepstral coefficients (MFCC) features for stressed speech recognition. It is shown that the alternate filterbank frequency partitions are more effective for recognition of speech under both simulated and actual stressed conditions

[22] Jean-Léon Bouraoui, Gwenael Bothorel, and Nadine Vigouroux. Transcription and annotation of an apprenticeship corpus: Application to the correction and self-correction

strategies. In *Proc. Error Handling in Spoken Dialogue Systems*, pages 35–40, Château d'Oex, Vaud, Switzerland, August 28-31 2003.

Abstract: This paper presents a corpus-based study devoted to corrections, self-corrections strategies to recover errors during dialogues. The corpus used results from exercises where air-traffic controllers being formed interacts with people simulating pilots in practice. Considering correction strategies as an answer to errors, we present a fine-grained typology of these strategies, the errors they aims to compensate, and their markers. We also give various statistics about their distribution in the corpus, and comment them, in regards with their application to spoken dialog systems

[23] C. D. Braby, D. Harris, and H. C. Muir. A psychophysiological approach to the assessment of work underload. *Ergonomics*, 36(9):1035, September 1993.

Abstract: The findings of a laboratory investigation of the relationship between the subjective and physiological components of work underload are reported. The subjective component is described in terms of the subjective work underload checklist, mental effort, and cognitive arousal. The physiological component is defined in terms of heart rate and heart rate variability. Evidence for an increase in work underload with a decrease in heart rate is provided. The relevance of this research to the aerospace environment is discussed and the need to investigate the behavioural component of work underload emphasized.

[24] M Brenner, ET Doherty, and T Shipp. Speech measures indicating workload demand. *Aviation, Space and Environmental Medicine*, 65(1):21–6, January 1994.

Abstract: Heart rate and six speech measures were evaluated using a manual tracking task under different workload demands. Following training, 17 male subjects performed three task trials: a difficult trial, with a $50 incentive for successful performance at a very demanding level; an easy trial, with a $2 incentive for successful performance at a simple level; and a baseline trial, in which there was physiological monitoring but no tracking performance. Subjects counted aloud during the trials. It was found that heart rate, speaking fundamental frequency (pitch), and vocal intensity (loudness) increased significantly with workload demands. Speaking rate showed a marginal increase, while vocal jitter and vocal shimmer did not show reliable changes. A derived speech measure, which statistically combined information from all other speech measures except shimmer, was also evaluated. It increased significantly with workload demands and was surprisingly robust in showing differences for individual subjects. It appears that speech analysis can provide practical workload information

[25] Jeffrey B. Brookings, Glenn F. Wilson, and Carolyne R Swain. Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, 42(3):361–377, 1996.

Abstract: In this investigation, eight Air Force air traffic controllers (ATCs) performed three scenarios on TRACON (Terminal Radar Approach Control), a computer-based air traffic control (ATC) simulation. Two scenarios were used each with three levels of difficulty. One scenario varied traffic volume by manipulating the number of aircraft to be handled and the second scenario varied traffic complexity by manipulating arriving to departing flight ratios, pilot skill and mixture of aircraft types. A third scenario, overload, required subjects to handle a larger number of aircraft in a limited amount of time. The effects of the manipulations on controller workload were assessed using performance, subjective (TLX), and physiological (EEG, eye blink, heart rate, respiration, saccade) measures. Significant main effects of difficulty level were found for TRACON performance, TLX, eye blink, respiration and EEG measures. Only the EEG was associated with main effects for the type of traffic. The results provide support for the differential sensitivity of a variety of workload measures in complex tasks, underscore the importance of traffic complexity in ATC workload, and support the utility of TRACON as a tool for studies of ATC workload

[26] Evan A. Byrne and Raja Parasuraman. Psychophysiology and adaptive automation. *Biological Psychology*, 42(3):249–268, 1996.

Abstract: Adaptive automation is an approach to automation design where tasks are dynamically allocated between the human operator and computer systems. Psychophysiology has two complementary roles in research on adaptive automation: first, to provide information about the effects of different forms of automation thus promoting the development of effective adaptive logic; and second, psychophysiology may yield information about the operator that can be integrated with performance measurement and operator modelling to aid in the regulation of automation. This review discusses the basic tenets of adaptive automation and the role of psychophysiological measures in the study of adaptive automation. Empirical results from studies of flight simulation are presented. Psychophysiological

measures may prove especially useful in the prevention of performance deterioration in underload conditions that may accompany automation. Individual differences and the potential for learned responses require research to understand their influence on adaptive algorithms. Adaptive automation represents a unique domain for the application of psychophysiology in the work environment

[27] P. Cabon, B. Farbos, and R. Mollard. Gaze analysis and psychophysiological parameters: A tool for the design and the evaluation of man-machine interfaces. Technical Report 350, Eurocontrol Experimental Centre, July 2000.

Abstract: This study demonstrates the value of gaze analysis with simultaneous physiological data recording for the assessment of the human-computer interface in air traffic control. The literature on visual scanning was reviewed. Two studies using different eye-tracking systems, one head-mounted (I VIEW) and one remote (ASL 504) were conducted at EEC. In each study, a TRACON II simulator was used to present controllers with a representative ATC task and a set of relevant electro-physiological measures (EEG, HRV etc..) was employed. This study shows that selected eye-view and physiological parameters may be used together to evaluate strain, fatigue, operational strategies of controllers in highly interactive situations. Further studies should investigate the effect of changes in the nature and extent of automation of the ATC task, such electronic stripping, flat displays etc

[28] Janet E. Cahn. The generation of affect in synthesized speech. *Journal of the American Voice I/O Society*, 8:1–19, 1990.

Abstract: Synthesized speech need not be expressionless. By identifying the effects of emotion on speech and choosing an appropriate representation, the generation of affect is possible and can become computational. I describe a program - the Affect Editor - which implements an acoustical model of speech and generates synthesizer instructions to produce the desired affect. The authenticity of the affect is limited by synthesizer capabilities and by incomplete descriptions of the acoustical and perceptual phenomena. However, the results of an experiment show that this approach produces synthesized speech with recognizable, and, at times, natural, affect.

[29] Douglas A. Cairns and John H. L. Hansen. Nonlinear analysis and classification of speech under stressed conditions. *Journal of the Acoustical Society of America*, 96(6):3392–3400, December 1994.

Abstract: The speech production system is capable of conveying an abundance of information with regards to sentence text, speaker identity, prosodics, as well as emotion and speaker stress. In an effort to better understand the mechanism of human voice communication, researchers have attempted to determine reliable acoustic indicators of stress using such speech production features as fundamental frequency (F0), intensity, spectral tilt, the distribution of spectral energy, and others. Their findings indicate that more work is necessary to propose a general solution. In this study, we hypothesize that speech consists of a linear and nonlinear component, and that the nonlinear component changes markedly between normal and stressed speech. To quantify the changes between normal and stressed speech, a classification procedure was developed based on the nonlinear Teager Energy operator. The Teager Energy operator provides an indirect means of evaluating the nonlinear component of speech. The system was tested using VC and CVC utterances from native speakers of English across the following speaking styles; neutral, loud, angry, Lombard effect, and clear. Results of the system evaluation show that loud and angry speech can be differentiated from neutral speech, while clear speech is more difficult to differentiate. Results also show that reliable classification of Lombard effect speech is possible, but system performance varies across speakers

[30] John A. Caldwell, Kecia K. Hall, and Bradley S. Erickson. EEG data collected from helicopter pilots in flight are sufficiently sensitive to detect increased fatigue from sleep deprivation. *International Journal of Aviation Psychology*, 12(1):19–32, 2002.

Abstract: This investigation determined whether the electroencephalographic (EEG) changes associated with sleep deprivation could be reliably recorded from aviators flying standardized maneuvers in an aircraft. In-flight EEG data were recorded from 10 UH-60 helicopter pilots who were kept awake for approximately 26 hr. In addition, resting EEGs and mood data were collected in the laboratory between flights. Results indicate that EEG theta activity, and to some extent delta activity, increases as a function of sleep deprivation in both settings. In addition, mood decrements were associated with the fatigue from sleep loss. These results indicate it is possible to monitor a pilot's general fatigue levels via the EEG without interfering with the primary duty of flying the aircraft

[31] J.G. Casali and W.W. Wierwille. On the measurement of pilot perceptual workload: a comparison of assessment techniques addressing sensitivity and intrusion issues. *Ergonomics*, 27(10):1033–50, October 1984.

[32] Y. Chen. Cepstral domain talker stress compensation for robust speech. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(4):433–439, 1988.

Abstract: A study of talker-stress-induced intraword variability and an algorithm that compensates for the systematic changes observed are presented. The study is based on hidden Markov models trained by speech tokens spoken in various talking styles. The talking styles include normal speech, fast speech, loud speech, soft speech, and taking with noise injected through earphones; the styles are designed to simulate speech produced under real stressful conditions. Cepstral coefficients are used as the parameters in the hidden Markov models. The stress compensation algorithm compensates for the variations in the cepstral coefficients in a hypothesis-driven manner. The functional form of the compensation is shown to correspond to the equalization of spectral tilts. Substantial reduction of error rates has been achieved when the cepstral domain compensation techniques were tested on the simulated-stress speech database. The hypothesis-driven compensation technique reduced the average error rate from 13.9% to 6.2%. When a more sophisticated recognizer was used, it reduced the error rate from 2.5% to 1.9%

[33] J.-J. Congleton, W.-A. Jones, S.-G. Shiflett, K.-P. McSweeney, and R.-D. Huchingson. An evaluation of voice stress analysis techniques in a simulated awacs environment. *International Journal of Speech Technology*, 2(1):61–69, 1997.

Abstract: Fundamental frequency, frequency jitter, and amplitude shimmer voice algorithms were employed to measure the effects of stress in crew member communications data in simulated AWACS mission scenarios. Two independent workload measures were used to identify levels of stress: (1) a predictor model developed by the simulation author based upon scenario generated stimulus events, and (2) the duration of communication for each weapons director, representative of the individual's response to the induced stress. Results identified fundamental frequency and frequency jitter as statistically significant vocal indicators of stress, while amplitude shimmer showed no signs of any significant relationship with workload or stress. Consistent with previous research, the frequency algorithm was identified as the most reliable measure. However, the results did not reveal a sensitive discrimination measure between levels of stress, but rather, did distinguish between the presence or absence of stress

[34] Randolph R. Cornelius. Theoretical approaches to emotion. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, pages 1–8, Belfast, Northern Ireland, September 2000.

Abstract: Four major theoretical perspectives on emotion in psychology are described. Examples of the ways in which research on emotion and speech utilize aspects of the various perspectives are presented and a plea is made for students of emotion and speech to consider more self-consciously the place of their research within each of the perspectives. integration of theories from the four perspectives

[35] G. Costa. Evaluation of workload in air traffic controllers. *Ergonomics*, 36(9):1111–1120, September 1993.

Abstract: This study examined 20 air traffic controllers from the Rome Regional Air Control Centre for three successive work shifts: afternoon (13:00-20:00), morning (07:00-13:00) and night (20:00-07:00). The number of aircraft under control per hour was recorded as index of workload. Recordings involved subjective ratings (mood, physical fitness, fatigue) and objective measures (heart rate, vanillyl mandelic acid excretion, reaction times, critical flicker fusion, oral temperature). In addition, the subjects filled out questionnaires for personality traits (extroversion, neuroticis, anxiety) and behavioural characteristics (morningness, rigidity of sleeping habits, vigourness to overcome drowsiness). The volume of air traffic varied greatly with peaks during the day and low levels at night. Nevertheless, the heart rates of the group members showed quite constant levels in all the three shifts, irrespective of the workload. The same pattern appeared in the controllers' excretion of VMA, which remained high during both day and night shifts, regardless of the reduced workload. The subjective mood and physical fitness decreased similarly, while feelings of fatigue increased on all three shifts, particularly on the night shift. The circadian rhythm of the oral temperature showed a slight modification of the nocturnal depression during the night shift, caused by the state of awakeness and activity. However, the rhythm was not altered in its normal circadian phase, due to the fast shift rotation adopted. The psychophysiological responses were affected by personal characteristics, in particular morningness and ability to overcome drowsiness

[36] Giovanni Costa. Occupational dfstress and stress prevention in air traffic control. Working paper. International Labor Office, Geneva, Italy, 1995.

[37] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1):32–80, January 2001.

Abstract: Two channels have been distinguished in human interaction: one transmits explicit messages, which may be about anything or nothing; the other transmits implicit messages about the speakers themselves. Both linguistics and technology have invested enormous efforts in understanding the first, explicit channel, but the second is not as well understood. Understanding the other party's emotions is one of the key tasks associated with the second, implicit channel. To tackle that task, signal processing and analysis techniques have to be developed, while, at the same time, consolidating psychological and linguistic analyses of emotion. This article examines basic issues in those areas. It is motivated by the PKYSTA project, in which we aim to develop a hybrid system capable of using information from faces and voices to recognize people's emotions

[38] Roddy Cowie and Randolph R Cornelius. Describing the emotional states that are expressed in speech. *Speech Communication*, 40(1-2):5–32, 2003.

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

[39] Tom Cox, Amanda Griffiths, and Eusebio Rial-González. *Research on Work-related Stress*. European Agency for Safety and Health at Work, Luxembourg, 2000. ISBN 92-828-9255-7.

Abstract: The European Agency for Safety and Health at Work commissioned this Status Report on stress at work within the framework of the Topic Centre on Research – Work and Health (TC/WH). The Report considers early and contemporary scientific studies on the nature of stress at work, on its effects on health and on the way in which such knowledge is being applied in attempts to manage this problem. The Topic Centre on Good Practice – Stress at Work (TC/GP-ST) collects and evaluates good practice information on stress at work both within the EU and beyond. Consequently, this Report deals with the research evidence regarding the assessment and management of stress at work: it does not review stress management in practice. However, it discusses the conceptual frameworks implied in the practice of stress management at work and in current health and safety legislation, focusing in particular on the utility of the 'control cycle' and problem-solving approaches to the management of stress at work.

[40] J.H. Crump. Review of stress in air traffic control: its measurement and effects. *Aviation, Space and Environmental Medicine*, 50(3):243–248, March 1979.

[41] K.E. Cummings and M.A. Clements. Analysis of glottal waveforms across stress styles. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 369–372, Albuquerque, NM USA, April 3-6 1990.

Abstract: Results from a study of glottal waveforms derived from 11 different styles of stressed speech are described. The general goal of this research is to investigate the differences in the glottal excitation across specific speaking styles. Using a glottal inverse filtering method tailored specifically for the speech under study, an analysis of glottal waveforms from voiced speech spoken under various conditions is performed. The extracted waveforms are parameterized using six descriptors, and statistical analysis of the resulting database allows very good characterization of each style with respect to these parameters. Specifically, each style is shown to have a unique profile which will allow high performance in a stress style identification task

[42] K.E. Cummings and M.A. Clements. Improvements to and applications of analysis of stressed speech. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 25–28, San Francisco, CA , USA, March 23-26 1992.

Abstract: A continuation of the authors' previous research (1990) in the area of statistical analysis and modeling of glottal waveforms of emotionally stressed speech is presented. A simple glottal model is defined which, by varying four parameters, can be used to model the glottal excitation of many styles of speech. An algorithm which automatically decouples these model glottal waveforms from the vocal tract information in a given speech segment is presented. Several applications of decoupling the model glottal waveforms are discussed. In particular, a method of automatic formant tracking of stressed speech which is superior to traditional formant tracking procedures is described. Also, a method of modifying the style of speech by removing some of the effects of stress on the speech waveform is presented

[43] K.E. Cummings and M.A. Clements. Analysis of the glottal excitation of emotionally styled and stressed speech. *Journal of the Acoustical Society of America*, 98(1):88–98, July 1995.

Abstract: The problems of automatic recognition of and synthesis of multistyle speech have become important topics of research in recent years. This paper reports an extensive investigation of the variations that occur in the glottal excitation of eleven commonly encountered speech styles. Glottal waveforms were extracted from utterances of non-nasalized vowels for two speakers for each of the eleven speaking styles. The extracted waveforms were parametrized into four duration-related and two slope-related values. Using these six parameters, the glottal waveforms from the eleven styles of speech were analyzed both qualitatively and quantitatively. The glottal waveforms from each style speech were analyzed both qualitatively and quantitatively. The glottal waveforms from each style of speech have been shown to be significantly and identifiably different from all other styles, thereby confirming the importance of the glottal waveform in conveying speech style information and in causing speech waveform variations. The degree of variation in styled glottal waveforms has been shown to be consistent when trained on one speaker and compared with another.

[44] Hugh David. Measuring people in control [air traffic controllers]. In *Human Interfaces in Control Rooms, Cockpits and Command Centres, 2001. People in Control. The Second International Conference on*, pages 285 – 290, Manchester UK, June 19-21 2001.

Abstract: Summarises the experience of EUROCONTROL (The European Organisation for the Safety of Air Navigation) in measuring people engaged in air traffic control simulations. By analogy with the engineering usage, the authors define stress as the external forces on the controller, mainly the workload imposed, and strain as the response of the controller to this stress

[45] F. de la Torre, C.J.G. Rubio, and E. Martinez. Subspace eyetracking for driver warning. In *Proceedings of International Conference on Image Processing*, volume 2, pages III – 329–32, September 14-17 2003.

Abstract: Driver's fatigue/distraction is one of the most common causes of traffic accidents. The aim of this paper is to develop a real time system to detect anomalous situations while driving. In a learning stage, the user will sit in front of the camera and the system will learn a person-specific facial appearance model (PSFAM) in an automatic manner. The PSFAM will be used to perform gaze detection and eye-activity recognition in a real time based on subspace constraints. Preliminary experiments measuring the PERCLOS index (average time that the eyes are closed) under a variety of conditions are reported

[46] Dick de Waard. *The measurement of drivers' mental workload.* PhD thesis, University of Groningen. Haren, The Netherlands: University of Groningen, Traffic Research Centre, 1996.

Abstract: Driving a vehicle may seem to be a fairly simple task. After some initial training many people are able to handle a car safely. Nevertheless, accidents do occur and the majority of these accidents can be attributed to human failure. At present there are factors that may even lead to increased human failure in traffic. Firstly, owing in part to increased welfare, the number of vehicles on the road is increasing. Increased road intensity leads to higher demands on the human information processing system and an increased likelihood of vehicles colliding. Secondly, people continue

to drive well into old age. Elderly people suffer from specific problems in terms of divided attention performance, a task that is more and more required in traffic. One of the causes of these increased demands is the introduction of new technology into the vehicle. It began with a car radio, was followed by car-phones and route guidance systems, and will soon be followed by collision avoidance systems, intelligent cruise controls and so on. All these systems require drivers' attention to be divided between the system and the primary task of longitudinal and lateral vehicle control. Thirdly, drivers in a diminished state endanger safety on the road. Longer journeys are planned and night time driving increases for economic purposes and/or to avoid congestions. Driver fatigue is currently an important factor in accident causation. But not only lengthy driving affects driver state, a diminished driver state can also be the result of the use of alcohol or (medicinal) sedative drugs. The above-mentioned examples have in common that in all cases driver workload is affected. An increase in traffic density increases the complexity of the driving task. Additional systems in the vehicle add to task complexity. A reduced driver state affects the ability to deal with these demands. How to assess this, i.e. how to assess driver mental workload is the main theme of this thesis

[47] F. Dellaert, T. Polzin, and A. Waibel. Recognizing emotion in speech. In *Proceedings of the International Conference on Spoken Language Processing*, volume 3, pages 1970–1973, 1996.

Abstract: This paper explores several statistical pattern recognition techniques to classify utterances according to their emotional state. We have recorded a corpus containing emotional speech with over 1000 utterances from different speakers. We present a new method of extracting prosodic features from speech, based on a smoothing spline approximation of the pitch contour. To make maximal use of the limited amount of training data available, we introduce a novel pattern recognition techniwue: majority voting of subspace specialists. Using this techniwue, we obtain classification performance that is close to human performance on the task

[48] Paula A. Desmond and Thomas W. Hoyes. Workload variation, intrinsic risk and utility in a simulated air traffic control task: Evidence for compensatory effects. *Safety Science*, 22(1-3):87–101, 1996.

[49] David Dinges. Optical computer recognition of behavioral stress. Project Technical Summary, Univ. of Pennsylvania, 2000–2004.

Abstract: Optical computer recognition algorithm of the face to detect the presence of stress for astronauts. In colaboration with Rutgers University, Prof. D. Metaxas

[50] P. Ekman. Universals and cultural differences in facial expressions of emotion. In J. Cole, editor, *Proc. Nebraska Symp. Motivation*, pages 207–283, 1972.

[51] P. Ekman. Are there basic emotions? *Psychological Review*, 99(3):550–553, 1992.

[52] M. Eriksson and N.P. Papanikotopoulos. Eye-tracking for detection of driver fatigue. In *Proceedings of IEEE Conference on Intelligent Transportation System, ITSC 97*, pages 314 – 319, Boston, MA USA, November 9-12 1997.

Abstract: In this paper, we describe a system that locates and tracks the eyes of a driver. The purpose of such a system is to perform detection of driver fatigue. By mounting a small camera inside the car, we can monitor the face of the driver and look for eye-movements which indicate that the driver is no longer in condition to drive. In such a case, a warning signal should be issued. This paper describes how to find and track the eyes. We also describe a method that can determine if the eyes are open or closed. The primary criterion for the successful implementation of this system is that it must be highly nonintrusive. The system should start when the ignition is turned on without having the driver initiate the system. Nor should the driver be responsible for providing any feedback to the system. The system must also operate regardless of the texture and the color of the face. It must also be able to handle diverse conditions, such as changes in light, shadows, reflections, etc

[53] Eric Farmer, Adam Brownson, and QinetiQ. Review of workload measurement, analysis and interpretation methods. Technical Report CARE-Integra-TRS-130-02-WP2, EUROCONTROL, 2003.

Abstract: This report was prepared as part of a project being conducted under the EUROCONTROL INTEGRA programme. The aim of the project is to derive principles of workload measurement in

man-in-the-loop simulations from experience in non-air traffic management (ATM) domains. This report describes the outcome of Work Package 2. Types of workload measure performancebased, subjective, and physiological/biochemical are critically reviewed, and advice is given on methods of selecting the best set of measures for ATM simulations. In the next Work Package, the development of sound experimental designs incorporating these measures will be considered.This report was prepared as part of a project being conducted under the EUROCONTROL INTEGRA programme. The aim of the project is to derive principles of workload measurement in man-in-the-loop simulations from experience in non-air traffic management (ATM) domains. This report describes the outcome of Work Package 2. Types of workload measure performancebased, subjective, and physiological/biochemical are critically reviewed, and advice is given on methods of selecting the best set of measures for ATM simulations. In the next Work Package, the development of sound experimental designs incorporating these measures will be considered

[54] W.A. Fellenz, J.G. Taylor, E. Cowie, R.; Douglas-Cowie, F. Piat, S. Kollias, C. Orovas, and B. Apolloni. On emotion recognition of faces and of speech using neural networks, fuzzy logic and the ASSESS system. In *IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN)*, volume 2, pages 93–98, 2000.

Abstract: We propose a framework for the processing of face image sequences and speech, using different dynamic techniques to extract appropriate features for emotion recognition. The features will be used by a hybrid classification procedure, employing neural network techniques and fuzzy logic, to accumulate the evidence for the presence of an emotional expression of the face and the speaker's voice

[55] R. Fernandez and R. W. Picard. Signal processing for recognition of human frustration. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 6, pages 3773–3776, 1998.

Abstract: In this work, inspired by the application of human-machine interaction and the potential use that human-computer interfaces can make of knowledge regarding the affective state of a user, we investigate the problem of sensing and recognizing typical affective experiences that arise when people communicate with computers. In particular, we address the problem of detecting "frustration" in human computer interfaces. By first sensing human biophysiological correlates of internal affective states, we proceed to stochastically model the biological time series with hidden Markov models to obtain user-dependent recognition systems that learn affective patterns from a set of training data. Labeling criteria to classify the data are discussed, and generalization of the results to a set of unobserved data is evaluated. Significant recognition results (greater than random) are reported for 21 of 24 subjects

[56] Raul Fernandez and Rosalind W. Picard. Analysis and classification of stress categories from drivers' speech. Technical Report 513, MIT Media Lab, Perceptual Computing Section, 1999.

Abstract: In this paper we explore the use of features derived from multiresolution analysis of speech and the Teager Energy Operator for classification of drivers' speech under stressed conditions. The potential stress categories are determined by driving speed and the frequency with which the driver has to solve a mental task while driving. We first use an unsupervised approach to gain some understanding as to whether the discrete stress categories form meaningful clusters in feature space, and use the clustering results to build a user-dependent recognition system which combines local discriminants of 4 discreet stress categories. Recognition results are reported for 4 subjects

[57] Raul Fernandez and Rosalind W. Picard. Modeling drivers' speech under stress. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, pages 1–6, Belfast, Northern Ireland, September 2000.

Abstract: In this paper we explore the use of features derived from multiresolution analysis of speech and the Teager Energy Operator for classification of drivers' speech under stressed conditions. We apply this set of features to a database of short speech utterances to create user-dependent discriminants of four stress categories. In addition we address the problem of choosing a suitable temporal scale for representing categorical di erences of the data. This leads to two sets of modeling techniques. In the first approach, we model the dynamics of the feature set within the utterance with a family of dynamic classifiers. In the second approach, we model the mean value of the features across the utterance with a family of static classifiers. We report and compare classification performances on the sparser and full dynamic representations for a set of four subjects

[58] Raul Fernandez and Rosalind W. Picard. Modeling drivers' speech under stress. *Speech Communication*, 40(1-2):145–159, 2003.

Abstract: We explore the use of features derived from multiresolution analysis of speech and the Teager energy operator for classification of drivers' speech under stressed conditions. We apply this set of features to a database of short speech utterances to create user-dependent discriminants of four stress categories. In addition we address the problem of choosing a suitable temporal scale for representing categorical differences in the data. This leads to two modeling approaches. In the first approach, the dynamics of the feature set within the utterance are assumed to be important for the classification task. These features are then classified using dynamic Bayesian network models as well as a model consisting of a mixture of hidden Markov models (M-HMM). In the second approach, we define an utterance-level feature set by taking the mean value of the features across the utterance. This feature set is then modeled with a support vector machine and a multilayer perceptron classifier. We compare the performance on the sparser and full dynamic representations against a chance-level performance of 25% and obtain the best performance with the speaker-dependent mixture model (96.4% on the training set, and 61.2% on a separate testing set). We also investigate how these models perform on the speaker-independent task. Although the performance of the speaker-independent models degrades with respect to the models trained on individual speakers, the mixture model still outperforms the competing models and achieves significantly better than random recognition (80.4% on the training set, and 51.2% on a separate testing set)

[59] Lisa R. Fournier, Glenn F. Wilson, and Carolyne R Swain. Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training. *International Journal of Psychophysiology*, 31(2):129–145, 1999.

Abstract: This study examined whether alpha event-related desynchronization (ERD) and theta event-related synchronization (ERS) could successfully measure changes in cognitive workload and training while an operator was engaged in a continuous, interactive, control task(s). Alpha 1 (8-10 Hz) ERD, alpha 2 (10-12 Hz) ERD, and theta (3-7 Hz) ERS were determined for a communications event that occurred during multiple task workload conditions or as a single task. Other measures (alpha and theta EEG power, heart rate, respiration, eye blinks, behavioral performance, and subjective workload ratings) were also evaluated. Results showed that alpha 2 EEG, heart rate, behavioral, and subjective measures were sensitive to changes in workload in the multiple tasks. In addition, eye blink rate and behavioral measures were sensitive to training. Alpha ERD and theta ERS were not sensitive to workload and training in our interactive, multiple task environment. However, they were effective indexes of cognitive/behavioral demands within an interactive single task

[60] A. W. K. Gaillard. Comparing the concepts of mental load and stress. *Ergonomics*, 36(9):991, September 1993.

Abstract: This paper delineates mental load and stress as two related concepts that originate from different theoretical frameworks. A proper distinction between the two concepts is important, not only for theory building, but because it may lead also to different interpretations of experimental results, and, consequently, to different recommendations in applied situations. High workload is regarded as an important but not a critical factor in the development of stress symptoms. It is quite possible to work hard in difficult and complex tasks, even under unfavourable conditions, without cognitive strain, psychosomatic complaints, or adverse physiological effects. High task demands can be met by mobilizing extra energy through mental effort. This 'trying harder' reaction is a normal and healthy coping strategy to adapt to situational demands. In contrast, stress is regarded as a state in which the equilibrium between cognitive and energetical processes is disturbed by ineffective energy mobilization and negative emotions. Stress typically is characterized by inefficient behaviour, overreactivity, and the incapacity to recover from work. Stress is regarded as a state in which the physiological system is disorganized, which results in decreased well-being, sleeping problems, psychosomatic complaints, and increased health risks

[61] Anthony W.K. Gaillard, Wolfgang Boucsein, and John Stern. Psychophysiology of workload. *Biological Psychology*, 42(3):245–247, 1996.

[62] K. Gopalan. On the effect of stress on certain modulation parameters of speech. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 101–104, Salt Lake City, UT USA, May 7-11 2001.

Abstract: This paper reports the results of correlation between demodulated amplitude and frequency variations from the AM-FM speech model and the heart rate of a fighter aircraft flight controller. It has been found that the peak frequencies in the spectrum of the amplitude envelope of the

model follow F0 regardless of the center frequency of analysis. This tracking of F0 gives a qualitative estimate of stress level-as measured by the heart rate-relative to a neutral state of stress, with no a priori knowledge of F0 or formants. Additionally, the mean values of F0 estimates at low and high heart rates increased significantly from those at a neutral state. Formant tracking showed an increase in F3 at both high and low heart rates while F4 generally varied directly with the heart rate

[63] K. Gopalan, S. Wenndt, and E.J. Cupples. An analysis of speech under stress using certain modulation. In *Industrial Electronics Society, 1999. IECON '99 Proceedings. The 25th Annual Conference of the IEEE*, volume 3, pages 1193–1197, 1999.

Abstract: An analysis of speech under stress using its amplitude and frequency modulation (AM and FM) characteristics has been carried out. From the preliminary results using actual stressed speech from an operational aviation emergency, stress, in general, appears to increase the modulation behavior in both AM and FM. In particular, the modulation characteristics become significantly more if the analysis (carrier) frequency is centered at one of the resonant frequencies of the vocal tract. Moreover, the spectra of AM envelopes show increasing peak frequencies with stress. The highest peak in the spectra of the demodulated instantaneous frequency and the AM envelope appear to track the fundamental frequency F0. Linear prediction model of the AM envelope also shows a peak at F0 regardless of the analysis (center) frequency; the peak does not be distinct at all center frequencies, however

[64] Richard Grace, V.E. Byrne, D.M. Bierman, J.-M. Legrand, D. Gricourt, B.K. Davis, J.J. Staszewski, and B. Carnahan. A drowsy driver detection system for heavy vehicles. In *Proceedings of the 17th Digital Avionics Systems Conference (DASC)*, volume 2, pages I36/1 – I36/8, Bellevue, WA USA, October 31–November 7 1998.

Abstract: Driver drowsiness/fatigue is an important cause of combination-unit truck crashes. Drowsy driver detection methods can form the basis of a system to potentially reduce accidents related to drowsy driving. We report on efforts performed at the Carnegie Mellon Driving Research Center to develop such in vehicle driver monitoring systems. Commercial motor vehicle truck drivers were studied in actual fleet operations. The drivers operated vehicles that were equipped to measure vehicle performance and driver psychophysiological data. Based on this work, two drowsiness detection methods are being considered. The first is a video-based system that measures PERCLOS, a scientifically supported measure of drowsiness associated with slow eye closure. The second detection method is based on a model to estimate PERCLOS based on vehicle performance data. A non-parametric (neural network) model was used to estimate PERCLOS using measures associated with lane keeping, steering wheel movements and lateral acceleration of the vehicle

[65] G.R. Griffin and C.E. Williams. The effects of different levels of task complexity on three vocal measures. *Aviation, Space and Environmental Medicine*, 58(12):1165–1170, December 1987.

Abstract: Interactive voice query and control systems are being considered as an alternative method of interfacing pilots with aircraft systems in high performance aviation environments. Before engineers integrate interactive voice systems into aircraft, knowledge is needed concerning what types of vocal changes, if any, are likely to occur as a result of environmental and task-induced stress. Recordings of the vocal utterances of 20 subjects, obtained as the subjects performed psychomotor and dichotic listening tasks, were subjected to acoustical analyses to evaluate the effects of task complexity on three measures: fundamental frequency (fo), peak amplitude (intensity), and word duration. The analyses revealed significant increases in fo (p less than 0.05) and amplitude (p less than 0.01) and a significant decrease in word duration (p less than 0.01) as a function of increased task complexity. These findings suggest that task-induced vocal changes must be considered in the design of voice recognition systems for use in complex, high performance aviation environments.

[66] A. Gundel and G. F. Wilson (editors). Special issue: Workload stress. *Ergonomics*, 36(9), September 1993.

[67] D.-M. Haddad and R.-J. Ratley. Voice stress analysis and evaluation. In *Proceedings of the SPIE The International Society for Optical Engineering*, volume 4232, pages 281–292, 2001.

Abstract: Voice stress analysis (VSA) systems are marketed as computer based systems capable of measuring stress in a person's voice as an indicator of deception. They are advertised as being less expensive, easier to use, less invasive in use, and less constrained in their operation then polygraph

technology. The National Institute of Justice have asked the Air Force Research Laboratory for assistance in evaluating voice stress analysis technology. Law enforcement officials have also been asking questions about this technology. If VSA technology proves to be effective, its value for military and law enforcement application is tremendous

[68] J.H.L. Hansen and S. Bou-Ghazale. Getting started with SUSAS: A speech under simulated and actual stress database. In *Proceedings of the European Conference on Speech Communication and Technology*, volume 4, pages 1743–1746, Rhodes, Greece, September 1997.

Abstract: It is well known that the introduction of acoustic background distortion and the variability resulting from environmentally induced stress causes speech recognition algorithms to fail. In this paper, we discuss SUSAS: a speech database collected for analysis and algorithm formulation of speech recognition in noise and stress. The SUSAS database refers to Speech Under Simulated and Actual Stress, and is intended to be employed in the study of how speech production and recognition varies when speaking during stressed conditions. This paper will discuss (i) the formulation of the SUSAS database, (ii) baseline speech recognition using SUSAS data, and (iii) previous research studies which have used the SUSAS data base. The motivation for this paper is to familiarize the speech community with SUSAS, which was released April 1997 on CD-ROM through the NATO RSG.10

[69] J.H.L. Hansen, C. Swail, A.J. South, R.K. Moore, H. Steeneken, E.J. Cupples, T. Anderson, C.R.A. Vloeberghs, I. Trancoso, and P. Verlinde. The impact of speech under 'stress' on military speech technology. Technical Report AC/323(IST)TP/5 IST/TG-01, NATO Research & Technology Organization RTO-TR-10, March 2000.

Abstract: Introduction: Military operations are often conducted under conditions of physical and mental stress which are detrimental to the effectiveness of speech communications. This stress is induced by high workload, sleep deprivation, frustration over contradictory information, emotions such as fear, pain, psycho- logical tension, and other modern battlefield conditions. These conditions are known to affect the physical and cognitive production of human speech. The change in speech production is likely to be disastrous not only to human understanding of coded speech, but also to the performance of communication equipment, and weapon systems equipped with vocal interfaces (e.g. advanced cockpits, C3 Systems - Command, Control, and Communication systems, and information warfare). The IST-TG01 (Formerly RSG.10) recognized the need to perform research and conduct studies on this topic to better understand, detect, and mitigate the e ects of stress on speech production. Thereby identifying and supporting future military requirements. In order to address the di erent scienti c aspects of this topic, Project 4 began with the organization, in cooperation with ESCA (European Speech Communication Association), of an international workshop on "Speech Under Stress" held in Lisbon, Portugal in September 1995. This very successful workshop underlined the necessity of a coordinated international effort to support NATO interests in this area. A primary outcome of this project is to create speech systems which are robust to stress indicative of the military speech environment. In order to share the most recent advances in this field the NATO IST-TG01 established a Speech Under Stress webpage1. This page contains an overview of the on-going activities, col- lected/available speech databases, international research groups, and an extensive updated set of references. This report is organized into five chapters, with conclusions and recommendations in Chapter 7.

[70] J.H.L. Hansen and B.D Womack. Feature analysis and neural network-based classification of speech. *IEEE Transactions on Speech and Audio Processing*, 4(4):307–313, 1996.

Abstract: It is well known that the variability in speech production due to task-induced stress contributes significantly to loss in speech processing algorithm performance. If an algorithm could be formulated that detects the presence of stress in speech, then such knowledge could be used to monitor speaker state, improve the naturalness of speech coding algorithms, or increase the robustness of speech recognizers. The goal in this study is to consider several speech features as potential stress-sensitive relayers using a previously established stressed speech database (SUSAS). The following speech parameters are considered: mel, delta-mel, delta-delta-mel, auto-correlation-mel, and cross-correlation-mel cepstral parameters. Next, an algorithm for speaker-dependent stress classification is formulated for the 11 stress conditions: angry, clear, cond50, cond70, fast, Lombard, loud, normal, question, slow, and soft. It is suggested that additional feature variations beyond neutral conditions reflect the perturbation of vocal tract articulator movement under stressed conditions. Given a robust set of features, a neural network-based classifier is formulated based on an extended delta-bar-delta learning rule. The performance is considered for the following three test scenarios: monopartition (nontargeted) and tripartition (both nontargeted and targeted) input feature vectors

[71] John H. L Hansen. Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition. *Speech Communication*, 20(1-2):151–173, 1996.

Abstract: It is well known that the introduction of acoustic background distortion and the variability resulting from environmentally induced stress causes speech recognition algorithms to fail. In this paper, several causes for recognition performance degradation are explored. It is suggested that recent studies based on a Source Generator Framework can provide a viable foundation in which to establish robust speech recognition techniques. This research encompasses three inter-related issues: (i) analysis and modeling of speech characteristics brought on by workload task stress, speaker emotion/stress or speech produced in noise (Lombard effect), (ii) adaptive signal processing methods tailored to speech enhancement and stress equalization, and (iii) formulation of new recognition algorithms which are robust in adverse environments. An overview of a statistical analysis of a Speech Under Simulated and Actual Stress (SUSAS) database is presented. This study was conducted on over 200 parameters in the domains of pitch, duration, intensity, glottal source and vocal tract spectral variations. These studies motivate the development of a speech modeling approach entitled Source Generator Framework in which to represent the dynamics of speech under stress. This framework provides an attractive means for performing feature equalization of speech under stress. In the second half of this paper, three novel approaches for signal enhancement and stress equalization are considered to address the issue of recognition under noisy stressful conditions. The first method employs (Auto:I,LSP:T) constrained iterative speech enhancement to address background noise and maximum likelihood stress equalization across formant location and bandwidth. The second method uses a feature enhancing artificial neural network which transforms the input stressed speech feature set during parameterization for keyword recognition. The final method employs morphological constrained feature enhancement to address noise and an adaptive Mel-cepstral compensation algorithm to equalize the impact of stress. Recognition performance is demonstrated for speech under a range of stress conditions, signal-to-noise ratios and background noise types

[72] John H.L. Hansen. *Analysis and Compensation of Stressed and Noisy Speech with Application to Robust Automatic Recognition.* PhD thesis, Georgia Institute of Technology, 1988.

[73] John H.L. Hansen. NATO IST-03 – Speech under stress homepage. http://cslr.colorado.edu/rspl/stress.html, 1998.

[74] John H.L. Hansen, Sahar E. Bou-Ghazale, Ruhi Sarikaya, and Bryan Pellom. Getting started with the SUSAS: Speech under simulated and actual stress database. Technical Report RSPL-98-10, Robust Speech Processing Laboratory, Duke University, April 1998.

[75] S.G. Hart and J.R .Hauser. Inflight application of three pilot workload measurement techniques. *Aviation, Space and Environmental Medicine*, 58(5):402–10, May 1987.

Abstract: Three measures of workload were tested during 11 routine missions conducted by the NASA Kuiper Airborne Observatory: communications performance, subjective ratings, and heart rate. The activities that contributed to crewmember workload varied; the commander was responsible for aircraft control and navigation whereas the copilot handled communications with ATC and the astronomers. Ratings of workload, stress, and effort given by the two crewmembers were highly correlated and varied across flight segments, peaking during takeoff and landing. Since the pilots performed different tasks during each segment, their ratings appeared to reflect overall crew workload, rather than experiences specific to each pilot. Subjective fatigue increased significantly from takeoff to landing for all flights, although the increase was significantly greater as landing times shifted from 10:00 p.m. to 9:00 a.m. The type, source, number, and frequency of communications varied significantly across flight segments, providing an objective indicator of pilot workload. Heart rate was significantly higher for the aircraft commander than for the copilot. Although heart rate peaked for both positions during takeoff and landing, the amount of change was significantly greater for the aircraft commander. Subjective ratings of stress, workload, and mental effort were significantly correlated with heart rate and communications frequency but were unrelated to mission duration, rated fatigue, or pilot evaluation of performance

[76] J. Healey and R. Picard. Smartcar: detecting driver stress. In *Proceedings of 15th International Conference on Pattern Recognition, 2000*, volume 4, pages 218–221, Barcelona Spain, September 3-7 2000.

Abstract: Smart physiological sensors embedded in an automobile afford a novel opportunity to capture naturally occurring episodes of driver stress. In a series of ten ninety minute drives on public

roads and highways, ECG, EMG, respiration and skin conductance sensors were used to measure the autonomic nervous system activation. The signals were digitized in real time and stored on the SmartCar's Pentium class computer. Each drive followed a pre-specified route through fifteen different events, from which four stress level categories were created according to the results of the subjects self report questionnaires. In total, 545 one minute segments were classified. A linear discriminant function was used to rank each feature individually based on the recognition performance, and a sequential forward floating selection algorithm was used to find an optimal set of features for recognizing patterns of driver stress. Using multiple features improved performance significantly over the best single feature performance

[77] J. A. Healey. *Wearable and Automotive Systems for Affect Recognition from Physiology.* Tr 526, Massachusetts Institute of Technology, Cambridge, MA, May 2000.

Abstract: Novel systems and algorithms have been designed and built to recognize affective patterns in physiological signals. Experiments were conducted for evaluation of the new systems and algorithms in three types of settings: a highly constrained laboratory setting, a largely unconstrained ambulatory environment, and a less unconstrained automotive environment. The laboratory experiment was designed to test for the presence of unique physiological patterns in each of eight different emotions given a relatively motionless seated subject, intentionally feeling and expressing these states. This experiment generated a large dataset of physiological signals containing many day-to-day variations, and the proposed features contributed to a success rate of 81% for discriminating all eight emotions and rates of up to 100% for subsets of emotion based on similar emotion qualities. New wearable computer systems and sensors were developed and tested on subjects who walked, jogged, talked, and otherwise went about daily activities. Although in the unconstrained ambulatory setting, physical motion often overwhelmed affective signals, the systems developed in this thesis are currently useful as activity monitors, providing an image diary correlated with physiological signals. Automotive systems were used to detect physiological stress during the natural but physically driving task. This generated a large database of physiological signals covering over 36 hours of driving. Algorithms for detecting driver stress achieved a recognition rates of 96% using stress ratings based on task conditions for validation and 89% accuracy using questionnaires analysis for validation. Additional analysis shows that highly significant correlations (up to r = .77 for over 4000 samples) exist between physiological features and a metric of observed stressors obtained from a second by second annotation of video tape records of the drives. Together, these three experiments show a range of success in recognizing affect from physiology, showing high recognition rates in somewhat constrained conditions and highlighting the need for more automatic context sensing in unconstrained conditions. The recognition rates obtained thus far lend support to the hypothesis that many emotional differences can be automatically discriminated in patterns of physiological changes

[78] Brian Hilburn and Peter G.A.M. Jorna. Workload and air traffic control. In Peter Hancock and Paula Desmond, editors, *Stress, Workload & Fatigue*, pages 384–394. Lawrence Erlbaum Associates, Mahwah, N.J., 2001.

[79] T. Huang, L. Chen, and H. Tao. Bimodal emotion recognition by man and machine. In *ATR Workshop on Virtual Communication Environments*, Kyoto, Japan, 1998.

Abstract: In this paper, we report preliminary results of recognizing emotions using both speech and video by machine. We applied automatic feature analysis and extraction algorithms to the same data used by De Silva et al. [5] in their human subjective studies. We compare machine recognition results to human performances in three tests:(1) video only, (2) audio only, (3) combined audio and video. In machine-assisted analyses, we found these two modalities to be complementary. Emotion categories that have similar features in one modality have very different features in the other modality. By using both, we show it is possible to achieve higher recognition rates than either modality alone

[80] R. Huber, A. Batliner, J. Buckow, E. Noth, V. Warnke, and H. Niemann. Recognition of emotion in a realistic dialogue scenario. In *Proceedings of the International Conference on Spoken Language Processing*, pages 665–668, Beijing, China, October 16-20 2000.

Abstract: Nowadays modern automatic dialogue systems are able to understand complex sentences instead of only a few com- mands like Stop or No. In a call-center, such a system should be able to determine in a critical phase of the dialogue if the call should be passed over to a human operator. Such a critical phase can be indicated by the customer's vocal expression. Other studies prooved that it is possible to distinguish between anger and neutral speech with prosodic features alone. Subjects in these studies were mostly people acting or simulating emotions like anger. In this paper we use data from a so-called Wizard of Oz (WoZ) scenario to get more realistic data instead of simulated

anger. As shown below, the classi cation rate for the two classes "emotion" (class E) and "neutral" (class :E) is signi cantly worse for these more realistic data. Furthermore the classi cation results are heavily speaker dependent. Prosody alone might thus not be su cient and has to be supplemented by the use of other knowledge sources such as the detection of repetitions, reformulations, swear words, and dialogue acts

[81] Qiang Ji and Xiaojie Yang. Real time visual cues extraction for monitoring driver vigilance. In B. Schiele and G. Sagerer, editors, *Computer Vision Systems: Second International Workshop, ICVS*, volume 2095 of *Lecture Notes in Computer Science*, pages 107–124. Springer, Vancouver, Canada, 2001.

Abstract: This paper describes a real-time prototype computer vision system for monitoring driver vigilance. The main components of the system consists of a specially-designed hardware system for real time image acquisition and for controlling the illuminator and the alarm system, and various computer vision algorithms for real time eye tracking, eyelid movement monitoring, face pose discrimination, and gaze estimation. Specific contributions include the development of an infrared illuminator to produce the desired bright/dark pupil effect, the development a digital circuitry to perform real time image subtraction, and the development of numerous real time computer vision algorithms for eye tracking, face orientation discrimination, and gaze tracking. The system was tested extensively in a simulating environment with subjects of different ethnic backgrounds, different genders, ages, with/without glasses, and under different illumination conditions, and it was found very robust, reliable and accurate

[82] Qiang Ji, Zhiwei Zhu, and Peilin Lan. Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE Transactions on Vehicular Technology*, 53(4):1052–1068, 2004.

Abstract: This paper describes a real-time online prototype driver-fatigue monitor. It uses remotely located charge-coupled-device cameras equipped with active infrared illuminators to acquire video images of the driver. Various visual cues that typically characterize the level of alertness of a person are extracted in real time and systematically combined to infer the fatigue level of the driver. The visual cues employed characterize eyelid movement, gaze movement, head movement, and facial expression. A probabilistic model is developed to model human fatigue and to predict fatigue based on the visual cues obtained. The simultaneous use of multiple visual cues and their systematic combination yields a much more robust and accurate fatigue characterization than using a single visual cue. This system was validated under real-life fatigue conditions with human subjects of different ethnic backgrounds, genders, and ages; with/without glasses; and under different illumination conditions. It was found to be reasonably robust, reliable, and accurate in fatigue characterization

[83] Tom Johnstone and Klaus R. Scherer. The effects of emotions on voice quality. In *Proceedings of the International Congress of Phonetic Sciences*, pages 2029–2032, 1999.

Abstract: Two studies are presented, in which emotional vocal recordings were made using a computer emotion induction task and an imagination technique. Concurrent recordings were made of a variety of physiological parameters, including electroglottograph, respiration, electrocardiogram, and surface electromyogram (muscle tension). Acoustic parameters pertaining to voice quality, including F0 floor, F0 range, jitter and spectral energy distribution were analysed and compared to the physiological parameters. The range of parameters was also statistically compared across different emotions. Although methodological problems still hamper such research, the combined analysis of physiological and acoustic parameters across emotional speaker states promises a clearer interpretation of the effects of emotion on voice quality

[84] Jean-Claude Junqua. The Lombard reflex and its role on human listeners and automatic speech recognizers. *Journal of the Acoustical Society of America*, 93(1):510–524, January 1993.

Abstract: Automatic speech recognition experiments show that, depending on the task performed and how speech variability is modeled, automatic speech recognizers are more or less sensitive to the Lombard reflex. To gain an understanding about the Lombard effect with the prospect of improving performance of automatic speech recognizers, (1) an analysis was made of the acoustic-phonetic changes occurring in Lombard speech, and (2) the influence of the Lombard effect on speech perception was studied. Both acoustic and perceptual analyses suggest that the influence of the Lombard effect on male and female speakers is different. The analyses also bring to light that, even if some tendencies across speakers can be observed consistently, the Lombard reflex is highly variable from

speaker to speaker. Based on the results of the acoustic and perceptual studies, some ways of dealing with Lombard speech variability in automatic speech recognition are also discussed.

[85] J.F. Kaiser. Some useful properties of Teager's energy operators. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 149–152, Minneapolis, USA, April 27-30 1993.

Abstract: Teager's energy operators are defined in both the continuous and discrete domains and are very useful tools for analyzing single component signals from an energy point of view. A number of important properties of these operators are shown that make it possible to determine the energy functions of quite complicated functions, provided these functions can be expressed as products of simpler functions, this operation of function multiplication being typical of a modulation process. Some of the eigenfunction properties of the energy operator that illustrate the special role of the trigonometric, Gaussian, and single soliton functions are also given

[86] Bong-Seok Kang, Chul-Hee Han, Sang-Tae Lee, Dae-Hee Youn, and Chungyong Lee. Speaker dependent emotion recognition using speech signals. In *Proceedings of the International Conference on Spoken Language Processing*, pages 383–386, 2000.

Abstract: This paper examines three algorithms to recognize speaker's emotion using the speech signals. Target emotions are happiness, sadness, anger, fear, boredom and neutral state. MLB(Maximum-Likelihood Bayes), NN(Nearest Neighbor) and HMM(Hidden Markov Model) algorithms are used as the pattern matching techniques. In all cases, pitch and energy are used as the features. The feature vectors for MLB and NN are composed of pitch mean, pitch standard deviation, energy mean, energy standard deviation, etc. For HMM, vectors of delta pitch with delta-delta pitch and delta energy with delta-delta energy are used. A corpus of emotional speech data was recorded and the subjective evaluation of the data was performed by 23 untrained listeners. The subjective recognition result was 56% and was compared with the classifiers recognition rates. MLB, NN, and HMM classifiers achieved recognition rates of 68.9%, 69.3%, and 89.1%, respectively, for the speaker dependent and context-independent classification

[87] Miriam Kienast and Walter F. Sendlmeier. Acoustical analysis of spectral and temporal changes in emotional speech. In *Proc. ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, pages 92–97, Newcastle, Northern Ireland, UK, September 5-7 2000.

Abstract: In the present study, the vocal expressions of the emotions anger, happiness, fear, boredom and sadness are acoustically analyzed in relation to neutral speech. The emotional speech material produced by actors is investigated especially with regard to spectral and segmental changes which are caused by different articulatory behavior accompanying emotional arousal. The findings are interpreted in relation to temporal variations

[88] Tetsuro Kitazoe, Sung-Ill Kim, Yasunari Yoshitomi, and Tatsuhiko Ikeda. Recognition of emotional states using voice, face image and thermal image of face. In *Proceedings of the International Conference on Spoken Language Processing*, pages 653–656, 2000.

Abstract: A new integration method is presented to recognize the emotional expressions. We attempted to use both voices and facial expressions. For voices, we use such prosodic parameters as pitch signals, energy, and their derivatives, which are trained by Hidden Markov Model (HMM) for recognition. For facial expressions, we use feature parameters from thermal images in addition to visible images, which are trained by neural networks (NN) for recognition. The thermal images are observed by infrared ray which is not influenced by lighting conditions. The total recognition rates show better performance than each performance rate obtained from isolated experiment. The results are compared with the recognition by human questionnaire

[89] T.I. Laine, K.W. Bauer, J.W. Lanning, C.A. Russell, and G.F Wilson. Selection of input features across subjects for classifying crewmember workload using artificial neural networks. *IEEE Transactions on Systems, Man and Cybernetics, Part A,*, 32(6):691–704, November 2002.

Abstract: The issue of crewmember workload is important in complex system operation because operator overload leads to decreased mission effectiveness. Psychophysiological research on mental workload uses measures such as electroencephalogram (EEG), cardiac, eye-blink, and respiration measures to identify mental workload levels. This paper reports a research effort whose primary objective was to determine if one parsimonious set of salient psychophysiological features can be identified to accurately classify mental workload levels across multiple test subjects performing a multiple task battery. To accomplish this objective, a stepwise multivariate discriminant analysis heuristic and artificial neural network feature selection with a signal-to-noise ratio (SNR) are used. In general, EEG power in the 31-40-Hz frequency range and ocular input features appeared highly salient. The second objective was to assess the feasibility of a single model to classify mental workload across different subjects. A classification accuracy of 87% was obtained for seven independent validation subjects using neural network models trained with data from other subjects. This result provides initial evidence for the potential use of generalized classification models in multitask workload assessment

[90] P. Lang. The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5):372–385, 1995.

Abstract: Emotions are action dispositions, states of vigilant readiness that vary widely in reported affect, physiology, and behavior. They are driven, however, by only 2 opponent motivational systems, appetitive and aversive subcortical circuits that mediate reactions to primary reinforcers. Using a large emotional picture library, reliable affective psychophysiologies are shown, defined by the judged valence (appetitive/pleasant or aversive/unpleasant) and arousal of picture percepts. Picture-evoked affects also modulate responses to independently presented startle probe stimuli. In other words, they potentiate startle reflexes during unpleasant pictures and inhibit them during pleasant pictures, and both effects are augmented by high picture arousal. Implications are elucidated for research in basic emotions, psychopathology, and theories of orienting and defense. Conclusions highlight both the approach's constraints and promising paths for future study

[91] Yang Li and Yunxin Zhao. Recognizing emotions in speech using short-term and long-term features. In *Proceedings of the International Conference on Spoken Language Processing*, volume 6, pages 2255–2258, Sydney, Australia, 1998.

Abstract: The acoustic characteristics of speech are influenced by speakers' emotional status. In this study, we attempted to recognize the emotional status of individual speakers by using speech features that were extracted from short-time analysis frames as well as speech features that represented entire utterances. Principal component analysis was used to analyze the importance of individual features in representing emotional categories. Three classification methods including vector quantization, artificial neural networks and Gaussian mixture density model were used. Classifications using short-term features only, long-term features only and both short-term and long-term features were conducted. The best recognition performance of 62% accuracy was achieved by using the Gaussian mixture density method with both short-term and long-term features.

[92] Scott E. Lively, David B. Pisoni, W. Van Summers, and Robert H. Bernacki. Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *Journal of the Acoustical Society of America*, 93(5):2962–73, May 1993.

Abstract: The present investigation examined the effects of cognitive workload on speech production. Workload was manipulated by having talkers perform a compensatory visual tracking task while speaking test sentences of the form "Say hVd again." Acoustic measurements were made to compare utterances produced under workload with the same utterances produced in a control condition. In the workload condition, some talkers produced utterances with increased amplitude and amplitude variability, decreased spectral tilt and F0 variability and increased speaking rate. No changes in F1, F2, or F3 were observed across conditions for any of the talkers. These findings indicate both laryngeal and sublaryngeal adjustments in articulation, as well as modifications in the absolute timing of articulatory gestures. The results of a perceptual identification experiment paralleled the acoustic measurements. Small but significant advantages in intelligibility were observed for utterances produced under workload for talkers who showed robust changes in speech production. Changes in amplitude and amplitude variability for utterances produced under workload appeared to be the major factor controlling intelligibility. The results of the present investigation support the assumptions of Lindblom's ["Explaining phonetic variation: A sketch of the H&H theory," in Speech Production and Speech Modeling (Klewer Academic, The Netherlands, 1990)] H&H model: Talkers adapt their speech to suit the demands of the environment and these modifications are designed to maximize intelligibility

[93] Arnab Majumdar and Washington Y. Ochieng. The factors affecting air traffic controller workload: A multivariate analysis based upon simulation modelling of con-

troller workload. Technical report, Centre for Transport Studies Department of Civil and Environmental Engineering Imperial College of Science, Technology and Medicine, London, 2002.

[94] S.P. Marshall, C.W. Pleydell-Pearce, and B.T. Dickson. Integrating psychophysiological measures of cognitive workload and eye movements to detect strategy shifts. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, pages 130–135, Hawaii, US, January 6-9 2003.

Abstract: This paper presents a case study to introduce a new technique for identifying and comparing cognitive strategies. The technique integrates eye movements with the Index of Cognitive Activity, a psychophysiological measurement of cognitive workload derived from changes in pupil dilation. The first part of the paper describes the technique and its constituent elements. The second part describes the task used in this study. The task has been used extensively in previous studies with electrophysiological recordings of EEG, EOG, and EMG and is know to elicit different levels of cognitive workload. The third part of the paper shows the results and provides detail about three separate strategies that emerged in the subject's performance. The strategies are first identified from differences in the Index of Cognitive Activity and corroborated through detailed analyses of the participant's eye movements as he performed the task

[95] J.C. McCall, S.P. Mallick, and M.M. Trivedi. Real-time driver affect analysis and tele-viewing system. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, pages 372–377, June 2003.

Abstract: This paper deals with the development of novel sensory systems and interfaces to enhance the safety of a driver who may be using a cell phone. It describes a system for real time affect analysis and Tele-viewing in order to bring the context of the driver to remote users. The system is divided into two modules. The affect recognition module recognizes six emotional states of the driver. The face-modeling module generates a 3D model of the driver's face using two images and synthesizes the emotions on it. Depending on bandwidth, either a raw video, 3D warped model, or an iconic representation is sent to the remote user

[96] Brian Mellor and Robert Graham. A corpus of read car number plates recorded under conditions of cognitive stress. Defence Evaluation and Research Agency (DERA), Malvern, UK, December 1995. CD-ROM.

Abstract: This memorandum provides details of a corpus of spoken car number plates, recorded at the SRU during the summer of 1992. The corpus comprises recordings of read car number plates dictated in the ICAO alphabet form (alpha, bravo, charlie etc), plus digits. Fifteen speakers were recorded, 11 male and four female covering a range of ages and featuring a number of non- extreme British regional accents. Recordings were carried out under two levels of cognitive stress, achieved by varying the presentation rate of the number plates. The subjects' perceived stress during the dictation sessions was recorded to provide data for a corpus of speech under limited stress conditions. This task was chosen partially to fulfil the demand for an initial speaker independent test database of spoken alpha-numerics, and partly to examine the effects of cognitive stress on speech performance and any concurrent effects on automatic speech recognition performance. Two stress levels were provided by displaying the data at high and low rates. The fast data display rate was aimed at requiring a dictation performance above the speakers ability, and so lead to performance failure and hence stress. The database was recorded in conjunction with the Industrial Ergonomics Group at Birmingham University, whose interest was in examining dictation accuracy, speaker coping strategies and the effects of stress on the performance of speech recognisers. The SRU component was funded by package 85T within RO 3.3

[97] E Mendoza and G Carballo. Acoustic analysis of induced vocal stress by means of cognitive workload tasks. *J Voice*, 12(3):263–73, September 1998.

Abstract: The purpose of this study was to determine the acoustic effects on voice of three tasks of cognitive workload and their possible relationship to stress. Acoustic analysis was used to measure stress and workload in four experimental tasks and two experiments. In the first experiment, subjects performed cognitive workload tasks under a stressful condition, performing the tasks as rapidly as possible without errors and with the knowledge that any errors committed would reduce their grade in a course. The second condition was to perform the same tasks but without the condition of stress related to the final grade. Four testing conditions were included. One was a baseline measure in which subjects spelled the Spanish alphabet. The second was the reading of a tongue twister, the third was the reading of a tongue twister with delayed auditory feedback, and the fourth was spelling the Spanish

alphabet in reverse order. In each condition the subjects prolonged the vowel /a/ for, approximately 5 sec. All subjects performed a test to determine their overall level of anxiety. The results suggest that in conditions of experimentally induced stress there is an increase in the fundamental frequency (F0) relative to baseline, an increase in jitter and shimmer, an increase in the high-frequency harmonic energy, and a decrease in spectral noise

[98] E Mendoza and G Carballo. Vocal tremor and psychological stress. *J Voice*, 13(1):105–112, March 1999.

Abstract: This study examines vocal tremor and its decrease in high versus low experimentally induced stress situations. We have analyzed the Amplitude Tremor Intensity Index (ATRI) and Frequency Tremor Intensity Index (FTRI) from the prolongation of vowel /a/ for approximately 5 seconds, under baseline conditions and under 3 different test conditions (reading of tongue twister, reading of tongue twister with delayed auditory feedback [DAF], and spelling of alphabet in reverse order), in a 2-test series, with and without demanding experimental instructions (Experiments 1 and 2, respectively). Inclusion of experimental instructions was considered as making the first test situation more stressful than the second one. Our results show a significant decrease in the ATRI variable when reading a tongue twister with DAF in relation to the baseline for the first test but not in the second, which suggests a suppression or significant reduction of amplitude tremor only in high-stress situations.

[99] Dimitris Metaxas, Sundara Venkataraman, and Christian Vogler. Image-based stress recognition using a model-based dynamic face tracking system. In Marian Bubak et al., editors, *4th International Conference Computational Science, Workshop on Dynamic Data Driven Applications Systems*, volume 3038 of *Lecture Notes in Computer Science*, pages 813–821. Springer-Verlag Heidelberg, Kraków, Poland, 2004.

Abstract: Stress recognition from facial image sequences is a subject that has not received much attention although it is an important problem for a host of applications such as security and human-computer interaction. This class of problems and the related software are instances of Dynamic Data Driven Application Systems (DDDAS). This paper presents a method to detect stress from dynamic facial image sequences. The image sequences consist of people subjected to various psychological tests that induce high and low stress situations. We use a model-based tracking system to obtain the deformations of different parts of the face (eyebrows, lips, mouth) in a parameterized form. We train a Hidden Markov Model system using these parameters for stressed and unstressed situations and use this trained system to do recognition of high and low stress situations for an unlabelled video sequence. Hidden Markov Models (HMMs) are an effective tool to model the temporal dependence of the facial movements. The main contribution of this paper is a novel method of stress detection from image sequences of a person's face

[100] S. Milosevic. Drivers' fatigue studies. *Ergonomics*, 40(3):381–389, March 1997.

Abstract: This paper summarizes the results of several fatigue studies of bus and truck drivers using different approaches. It presents findings from 24 city bus drivers obtained by the use of biochemical and psychophysiological tests before and after 7 h of driving. It presents an acoustic speech analysis conducted on 34 bus drivers before and after driving. Also, for a small group of city bus drivers, a continual examination of heart rate was carried out by electrocardiorecorder. Questionnaire studies on fatigue are related to the responses of 200 long-distance truck drivers and 107 dump-truck drivers who work in one copper mine. The last approach deals with the analysis of long-distance truck drivers' activities (driving, loading-unloading, resting and sleeping) during trips, based on their individual records. On the basis of these different approaches to bus and truck drivers' studies, clear psychophysiological, speech and subjective changes have been demonstrated and, on a descriptive level, certain symptoms observed during prolonged driving have been interpreted as effects of drivers' fatigue

[101] R.K. (Editor) Moore. Speech under stress (special issue). *Speech Communication*, 20(1–2), 1996.

[102] T. L. Morris and J. C Miller. Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology*, 42(3):343–360, 1996.

Abstract: The investigation evaluated specific components of eye and eyelid movement as predictors of performance decrements resulting from pilot fatigue. Ten partially sleep deprived pilots flew a GAT-1 moving-base flight simulator on a 4.5-h sortie. The scored flight portion consisted of eight legs, each leg made up of two segments, a flight maneuvers task (FMT) and a straight and level flying task (SLT). Error scores were calculated across altitude, airspeed, heading, and vertical velocity. An electrooculogram provided measures of blink rate (BR), blink duration, long closure rate (LCR),

blink amplitude (BA), saccade velocity, saccade rate, and peak saccade velocity. Subjective fatigue, workload and sleepiness were estimated using the USAFSAM seven-point forced-choice scales, the Stanford Sleepiness Scale, and the USAFSAM Sleep Survey Form. Error scores increased significantly during the first seven legs of the sortie and decreased slightly for the last leg. Subjective reports of fatigue increased significantly over time and were positively correlated with increased error. For the combined data set and for FMTs alone, BA was the best predictor of changes in error with decreased amplitude corresponding to increased error. BR and LCR were the second and third best predictors, respectively. For SLTs alone, LCR and BA were the first and second best predictors of increased error, respectively. The investigation demonstrated that measurable flying performance decrements do occur due to changes in fatigue and that one can measure physiological correlates of those performance decrements

[103] A. Murata and H. Iwase. Evaluation of mental workload by fluctuation analysis of pupil area. In *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 6, pages 3094–3097, Hong Kong, China, October 29–November 1 1998.

Abstract: It is generally known that the pupil is regulated by the autonomic nervous system. In this study, an attempt was made to assess mental workload on the basis of fluctuation rhythm of pupil area. Controlling the respiration interval, the pupil area during mental tasking was measured for 1 min. The respiration curve was measured simultaneously to monitor the respiration interval. Two mental tasks were used. One was a division task, the work level of which was set to two, three, four and five number of digits of the dividend. The other was a Sternberg memory search task, which had four work levels defined by the number of memory set. In the Sternberg memory search, the number of memory sets changed from five to eight. In such a way, the mental workload induced by mental loading was changed. As a result of calculating an AR power spectrum, we could observe two peaks which corresponded to the blood pressure variation and respiratory sinus arrhythmia under a low workload condition. With increasing workload, the spectral peak related to the respiratory sinus arrhythmia disappeared. The ratio of the power at the low frequency band (from 0.05-0.15 Hz) to the power at the respiration frequency band (from 0.35-0.4 Hz) increased with work level. In conclusion, the fluctuation of the pupil area can be used as a promising means to evaluate mental workload or autonomic function to supplement well-used heart rate variability measures

[104] Iain R. Murray, John L. Arnott, and Elizabeth A. Rohwer. Emotional stress in synthetic speech: Progress and future directions. *Speech Communication*, 20(1-2):85–91, 1996.

[105] Iain R. Murray, Chris Baber, and Allan South. Towards a definition and working model of stress and its effects on speech. *Speech Communication*, 20(1-2):3–12, 1996.

Abstract: Discussions at the ESCA-NATO Workshop on Speech Under Stress (Lisbon, Portugal, September 1995) centred on definitions and models of stress and its effects. Based on the Workshop discussions, in this paper we attempt to produce a definition of stress, and propose a number of stress models which clarify issues in this field, and which might be adopted by the speech community. The concept of stress is very broad and used differently in a number of domains – a definition of stress has thus remained elusive. Similarly, our understanding of the processes of stress and the ways in which it affects speech is incomplete, and any models used to describe stress are somewhat elementary. Greater separation of Stressors (the causes of stress) and strain (the effects of stress) is proposed, and methods for relating Stressors to strains are presented. Suggestions for future research directions in this field are also made

[106] R. Nakatsu, A. Solomides, and N. Tosa. Emotion recognition and its application to computer agents with spontaneous interactive capabilities. In *IEEE 3rd Workshop on Multimedia Signal Processing*, pages 439–444, 1999.

Abstract: Introduction: Nonverbal information such as emotions plays an essential role in human communications. Unfortunately, only the verbal aspect of communications has been focused upon in communications between humans and computer agents. In the future, however, it is expected that computer agents will have nonverbal as well as verbal communication capabilities. In this paper, we first study the recognition of emotions involved in human speech. Then we tried to apply this emotion recognition algorithm to a computer agent that plays a character role in the interactive movie system we are developing

[107] R. Nakatsu, A. Solomides, and N. Tosa. Emotion recognition and its application to computer agents with spontaneous interactive capabilities. In *IEEE International Conference on Multimedia Computing and Systems*, volume 2, pages 804–808, 1999.

Abstract: Nonverbal information such as emotion plays an essential role in human communication. We first study the recognition of emotions involved in human speech. We propose an emotion recognition algorithm based on a neural network. An emotion recognition rate of approximately 50% was obtained in a speaker-independent mode for eight emotion states. We tried to apply this emotion recognition algorithm to a computer agent that plays a character role in the interactive movie system we are developing. We propose to use emotion recognition as key technology for an architecture of the computer characters with both narrative-based and spontaneous interaction capabilities

[108] Tin Lay Nwe, Say Wei Foo, and L.C. De Silva. Classification of stress in speech using linear and nonlinear features. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages II – 9–12, April 6-10 2003.

Abstract: Three systems for the classification of stress in speech are proposed. The first system makes use of linear short time log frequency power coefficients (LFPC), the second employs a Teager energy operator (TEO) based nonlinear frequency domain LFPC features (NFD-LFPC) and the third uses TEO based nonlinear time domain LFPC features (NTD-LFPC). The systems were tested using the SUSAS (speech under simulated and actual stress) database to categorize five stress conditions individually. Results show that the system using LFPC gives the highest accuracy, followed by the system using NFD-LFPC features, while the system using NTD-LFPC features gives the worst performance. For the system using linear LFPC features, average accuracy of 84% and best accuracy of 95% were obtained in classifying five stress categories

[109] Tin Lay Nwe, Say Wei Foo, and L.C. De Silva. Detection of stress and emotion in speech using traditional and FFT based log energy features. In *Fourth Pacific Rim Conference on Multimedia, Information, Communications and Signal Processing*, volume 3, pages 1619–1623, December 15-18 2003.

Abstract: In this paper, a novel system for detection of human stress and emotion in speech is proposed. The system makes use of FFT based linear short time log frequency power coefficients (LFPC) and TEO based nonlinear LFPC features in both time and frequency domains. The performance of the proposed system is compared with the traditional approaches which use features of LPCC and MFCC. The comparison of each approach is performed using SUSAS (speech under simulated and actual stress) and ESMBS (emotional speech of Mandarin and Burmese speakers) databases. It is observed that proposed system outperforms the traditional systems. Results show that, the system using LFPC gives the highest accuracy (87.8% for stress, 89.2% for emotion classification) followed by the system using NFD-LFPC feature. While the system using NTD-LFPC feature gives the lowest accuracy

[110] Robert F. Orlikoff. Vowel amplitude variation associated with the heart cycle. *Journal of the Acoustical Society of America*, 88(5):2091–2098, 1990.

Abstract: Modulation of the acoustic amplitude of a sustained vowel across the cardiac (ECG) cycle was examined by signal-averaging techniques. Ten normal men prolonged [a] at a comfortable Fo maintained within three SPL ranges: 60-68, 70-78, and 80-88 dB. Peak-to-peak amplitude variation associated with the heart cycle averaged 8.5% (s.d. = 5.4) re: mean, varying from about 14low SPLs to approximately 3% at high SPLs. The amplitude modulation was estimated to account for 11.8% of the measured short-term amplitude perturbation (shimmer), ranging from about 5% to almost 22% for individual samples. The mean deterministic shimmer (Sd) was 0.036 dB (s.d. = 0.019), with a trend toward decreasing Sd with increasing SPL. Additionally, fundamental frequency variation across the heart cycle within these phonations was comparable to that observed by Orlikoff and Baken [J. Acoust. Soc. Am. 85, 888-893 (1989)], and was shown to be uninfluenced by vocal SPL, although deterministic jitter (Jd) did decrease with vocal intensity. The results are discussed in terms of how the phonovascular relationship may affect the reliability and interpretation of acoustic shimmer measures

[111] Robert F. Orlikoff and R.J. Baken. The effect of the heartbeat on vocal fundamental frequency perturbation. *Journal of Speech and Hearing Research*, 32(3):576–82, September 1989.

Abstract: Signal-averaging and autocorrelation analysis revealed that the cardiovascular system exerts a modest but consistent influence on vocal fundamental frequency (F0), accounting for approximately 0.5% to 20% of the absolute Fo perturbation (jitter) measured during a sustained phonation. There was also a marked trend for this percentage to decrease with increasing vocal Fo. Estimated mean "deterministic jitter" (Jd) values of 3.7 microsec (SD = 3.2) and 0.9 microsec (SD = 0.5) were derived from 6 normal male and 6 normal female subjects, respectively, with an overall mean of 2.3 microsec (SD = 2.7). These values represent approximately 6.9% of the mean total jitter for men and 2.4% of the mean total jitter for women, or about 4.6% for all subjects. The results are discussed in terms of their significance regarding more reliable vocal jitter measurement

[112] Robert F. Orlikoff and R.J. Baken. Fundamental frequency modulation of the human voice by the heart beat: Preliminary results and possible mechanims. *Journal of the Acoustical Society of America*, 85(2):888–893, 1989.

Abstract: Signal-averaging techniques reveal that the vocal fundamental frequency (F0) of a sustained vowel is modulated over a period equal to that of the speaker's heart cycle. Average F0 deviation varies in an orderly way from about 0.5% to about 1.0% as F0 changes. Location of the peak deviation in the time frame of the heart cycle also changes systematically with vocal F0. Modulation of the vocal F0 is likely to be caused by pressure-related changes in the stiffness of the vascular bed of the vocal folds and by alterations of the geometry of the thyroarytenoid muscle produced by periodic vascular engorgement

[113] Andrew Ortony and Terence J. Turner. What's basic about basic emotions? *Psychological Review*, 97(3):315–331, 1990.

[114] Astrid Paeschke and Walter F. Sendlmeier. Prosodic characteristics of emotional speech: Measurements of fundamental frequency movements. In *Proc. ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, pages 75–80, Newcastle, Northern Ireland, UK, September 5-7 2000.

Abstract: In the present study, the vocal expressions of the emotions anger, happiness, fear, boredom and sadness are acoustically analyzed in relation to neutral speech. The emotional speech material produced by actors is investigated especially with regard to spectral and segmental changes which are caused by different articulatory behavior accompanying emotional arousal. The findings are interpreted in relation to temporal variations

[115] M. Pantic, I. Patras, and L. Rothkrantz. Facial action recognition in face profile image sequences. In *IEEE International Conference on Multimedia and Expo (ICME)*, volume 1, pages 37–40, August 2002.

Abstract: A robust way to discern facial gestures in images of faces, insensitive to scale, pose, and occlusion, is still the key research challenge in the automatic facial-expression analysis domain. A practical method recognized as the most promising one for addressing this problem is through a facial-gesture analysis of multiple views of the face. Yet, current systems for automatic facial-gesture analysis utilize mainly portrait or nearly frontal views of faces. To advance the existing technological framework upon which research on automatic facial-gesture analysis from multiple facial views can be based, we developed an automatic system as to analyze subtle changes in facial expressions based on profile-contour reference points in a profile-view video. A probabilistic classification method based on statistical modeling of the color and motion properties of the profile in the scene is proposed for tracking the profile face. From the segmented profile face, we extract the profile contour and from it, we extract 10 profile-contour reference points. Based on these, 20 individual facial muscle actions occurring alone or in a combination are recognized by a rule-based method. A recognition rate of 85% is achieved

[116] M. Pantic and L.J.M. Rothkrantz. Facial gesture recognition in face image sequences: A study on facial gestures typical for speech articulation. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 6, page 6 pp., 2002.

Abstract: Automatic analysis of facial gestures is rapidly becoming an area of intense interest in computer science and human-computer interaction design communities. However, the basic goal of this area of research translating detected facial changes into a human-like description of shown, facial expression is yet to be achieved. One of the main impediments to achieving this aim is the fad that human interpretations of a facial expression differ depending upon whether the observed

person is speaking or not. A first step in tackling this problem is to achieve automatic detection of facial gestures that are typical for speech articulation. This paper presents our approach to seizing this step in the research on automatic facial expression analysis. It presents: a robust and flexible method for recognition of 22 facial muscle actions from face image sequences, a method for automatic determination of whether the observed subject is speaking or not, and an experimental study on facial muscle actions typical for speech articulation

[117] M. Pantic and L.J.M. Rothkrantz. Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man and Cybernetics, Part B,*, 34(3):1449–1461, June 2004.

Abstract: Automatic recognition of facial gestures (i.e., facial muscle activity) is rapidly becoming an area of intense interest in the research field of machine vision. In this paper, we present an automated system that we developed to recognize facial gestures in static, frontal- and/or profile-view color face images. A multidetector approach to facial feature localization is utilized to spatially sample the profile contour and the contours of the facial components such as the eyes and the mouth. From the extracted contours of the facial features, we extract ten profile-contour fiducial points and 19 fiducial points of the contours of the facial components. Based on these, 32 individual facial muscle actions (AUs) occurring alone or in combination are recognized using rule-based reasoning. With each scored AU, the utilized algorithm associates a factor denoting the certainty with which the pertinent AU has been scored. A recognition rate of 86% is achieved

[118] Maja Pantic and Leon J. M. Rothkrantz. Toward an affect-sensitive multimodal human–computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390, 2003.

Abstract: he ability to recognize affective states of a person we are communicating with is the core of emotional intelligence. Emotional intelligence is a facet of human intelligence that has been argued to be indispensable and perhaps the most important for successful interpersonal social interaction. This paper argues that next-generation human-computer interaction (HCI) designs need to include the essence of emotional intelligence - the ability to recognize a user's affective states-in order to become more human-like, more effective, and more efficient. Affective arousal modulates all nonverbal communicative cues (facial expressions, body movements, and vocal and physiological reactions). In a face-to-face interaction, humans detect and interpret those interactive signals of their communicator with little or no effort. Yet design and development of an automated system that accomplishes these tasks is rather difficult. This paper surveys the past work in solving these problems by a computer and provides a set of recommendations for developing the first part of an intelligent multimodal HCI-an automatic personalized analyzer of a user's nonverbal affective feedback

[119] R. Parasuraman, T.B. Sheridan, and C.D Wickens. A model for types and levels of human interaction with automation. *Systems, Man and Cybernetics, Part A, IEEE Transactions on*, 30(3):286–297, 2000.

Abstract: We outline a model for types and levels of automation that provides a framework and an objective basis for deciding which system functions should be automated and to what extent. Appropriate selection is important because automation does not merely supplant but changes human activity and can impose new coordination demands on the human operator. We propose that automation can be applied to four broad classes of functions: 1) information acquisition; 2) information analysis; 3) decision and action selection; and 4) action implementation. Within each of these types, automation can be applied across a continuum of levels from low to high, i.e., from fully manual to fully automatic. A particular system can involve automation of all four types at different levels. The human performance consequences of particular types and levels of automation constitute primary evaluative criteria for automation design using our model. Secondary evaluative criteria include automation reliability and the costs of decision/action consequences, among others. Examples of recommended types and levels of automation are provided to illustrate the application of the model to automation design

[120] I. Patras and M. Pantic. Particle filtering with factorized likelihoods for tracking facial features. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 97–102, 2004.

Abstract: In the recent years particle filtering has been the dominant paradigm for tracking facial and body features, recognizing temporal events and reasoning in uncertainty. A major problem associated with it is that its performance deteriorates drastically when the dimensionality of the state space is high. In this paper, we address this problem when the state space can be partitioned in groups of random variables whose likelihood can be independently evaluated. We introduce a novel proposal density which is the product of the marginal posteriors of the groups of random variables. The proposed

method requires only that the interdependencies between the groups of random variables (i.e. the priors) can be evaluated and not that a sample can be drawn from them. We adapt our scheme to the problem of multiple template-based tracking of facial features. We propose a color-based observation model that is invariant to changes in illumination intensity. We experimentally show that our algorithm clearly outperforms multiple independent template tracking schemes and auxiliary particle filtering that utilizes priors

[121] C'ecile Pereira. Dimensions of emotional meaning in speech. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, pages 1–4, Belfast, Northern Ireland, September 2000.

Abstract: A number of psychologists have conceptualized the communication of affect as three-dimensional (e.g. Davitz, 1964). The three dimensions proposed are arousal, pleasure and power. This paper reports the findings of a perception study where 31 normally-hearing subjects rated utterances said in the emotion of happiness, sadness, two forms of anger (hot and cold) and a neutral state on the dimensional scales of arousal, pleasure and power. Findings show that the concept of the dimensions of emotion is useful to describe and distinguish emotions; and that emotions with a similar level of arousal, and sometimes a similar level of power, share acoustic characteristics in terms of F0 range and mean, and particularly intensity mean. It is suggested that this contributes to perceived similarity between emotions, and consequently confusions, especially in the hearing-impaired

[122] Valery A. Petrushin. Emotion recognition in speech signal: Experimental study, development, and application. In *Proceedings of the International Conference on Spoken Language Processing*, pages 222–225, 2000.

Abstract: The paper describes an experimental study on vocal emotion expression and recognition and the development of a computer agent for emotion recognition. The study deals with a corpus of 700 short utterances expressing five emotions: happiness, anger, sadness, fear, and normal (unemotional) state, which were portrayed by thirty subjects. The utterances were evaluated by twenty three subjects, twenty of whom participated in recording. The accuracy of recognition emotions in speech is the following: happiness - 61.4%, anger - 72.2%, sadness - 68.3%, fear - 49.5%, and normal - 66.3%. The human ability to portray emotions is approximately at the same level (happiness - 59.8%, anger - 71.7%, sadness - 68.1%, fear - 49.7%, and normal - 65.1%), but the standard deviation is much larger. The human ability to recognize their own emotions has been also evaluated. It turned out that people are good in recognition anger (98.1%), sadness (80%) and fear (78.8%), but are less confident for normal state (71.9%) and happiness (71.2%). A part of the corpus was used for extracting features and training computer based recognizers. Some statistics of the pitch, the first and second formants, energy and the speaking rate were selected and several types of recognizers were created and compared. The best results were obtained using the ensembles of neural network recognizers, which demonstrated the following accuracy: normal state - 55-75%, happiness - 60-70%, anger - 70-80%, sadness - 75-85%, and fear - 35-55%. The total average accuracy is about 70%. An emotion recognition agent was created that is able to analyze telephone quality speech signal and distinguish between two emotional states –"agitation" and "calm" – with the accuracy of 77%. The agent was used as a part of a decision support system for prioritizing voice messages and assigning a proper human agent to response the message at call center environment. The architecture of the system is presented and discussed

[123] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. Toward machine emotional intelligence: Analysis and affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(12):1175–1191, October 2001.

Abstract: The ability to recognize emotion is one of the hallmarks of emotional intelligence, an aspect of human intelligence that has been argued to be even more important than mathematical and verbal intelligences. This paper proposes that machine intelligence needs to include emotional intelligence and demonstrates results toward this goal: developing a machine's ability to recognize the human affective state given four physiological signals. We describe difficult issues unique to obtaining reliable affective data and collect a large set of data from a subject trying to elicit and experience each of eight emotional states, daily, over multiple weeks. This paper presents and compares multiple algorithms for feature-based recognition of emotional state from this data. We analyze four physiological signals that exhibit problematic day-to-day variations: The features of different emotions on the same day tend to cluster more tightly than do the features of the same emotion on different days. To handle the daily variations, we propose new features and algorithms and compare their performance. We find that the technique of seeding a Fisher Projection with the results of sequential floating forward search improves the performance of the Fisher Projection and provides the highest recognition rates reported to date for classification of affect from physiology: 81 percent recognition accuracy on eight classes of emotion, including neutral

[124] T.S. Polzin. Verbal and non-verbal cues in the communication of emotions. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 6, pages 2429–2432, 2000.

Abstract: Several studies have shown that humans extend their interpersonal behavioral patterns onto their interaction with computers. Based on this finding, reserach in human-computer interaction acknowledges the need to detect, the users' expressions of emotion. However, we discovered that most of the current research is confined to emotion synthesis. In this investigation, we explored the role of verbal and non-verbal information in the communication of emotions. By training emotion-specific language and prosodic models on a corpus consisting of several thousands of sad, angry, or neutral speech segments from English movies, we showed that a classification system based on these models achieved an accuracy compareable to the accuracy of human listeners performing the same task

[125] C.A. Porubcansky. Speech technology: Present and future applications in the airborne environment. *Aviation, Space and Environmental Medicine*, 56(2):138–43, February 1985.

Abstract: Advanced speech technology systems, specifically voice recognition and voice synthesis systems are being considered as viable solutions to the problem of distributing pilot workload in today's increasingly complex airborne environment. The U.S. Air Force has funded various projects in this area over the past few years including studies defining and establishing priorities for candidate tasks to be accomplished by voice control, as well as a developmental hardware program which entered a flight test phase in December 1982. The applications of this technology are constrained only by the rate of its evolution. Certain tasks have already been identified as excellent candidates for control by a voice system and the present hardware is capable of being configured to perform these tasks. Other tasks have been identified as potential candidates. Although these candidates are tasks which would impact pilot workload, present technology is incapable fo supporting them

[126] Athanassios Protopapas and Philip Lieberman. Effects of vocal F0 manipulations on perceived emotional stress. In *Proceedings of the ESCA/NATO Tutorial and Research Workshop on Speech Under Stress*, pages 1–4, Lisbon, Portugal, September 14–15 1995.

Abstract: In this study we investigated the effect of jitter and longterm F0 measures on perceived emotional stress using stimuli synthesized with the LPC coefficients of a steady vowel and varying F0-tracks. The original F0 tracks were taken from naturally occurring speech in stressful and not stressful conditions. Stimuli with more jitter were rated as sounding more hoarse but not more stressed. Mean and maximum F0 within an utterance correlated highly (r¿0.8) with stress ratings but range of F0 did not correlate significantly with the stress ratings, especially after the effect of maximum F0 was removed by stepwise regression

[127] Athanassios Protopapas and Philip Lieberman. Fundamental frequency of phonation and perceived emotional stress. *Journal of the Acoustical Society of America*, 101(4):2267–2277, April 1997.

Abstract: Nonlinguistic information about the speaker's emotional state is conveyed in spoken utterances by means of several acoustic characteristics and listeners can often reliably identify such information. In this study we investigated the effect of short- and long-term F0 measures on perceived emotional stress using stimuli synthesized with the LPC coefficients of a steady vowel and varying F0 tracks. The original F0 tracks were taken from naturally occurring speech in highly stressful (contingent on terror) and nonstressful conditions. Stimuli with more jitter were rated as sounding more hoarse but not more stressed, i.e., a demonstrably perceptible amount of jitter did not seem to play a role in perceived emotional stress. Reversing the temporal pattern of F0 did not affect the stress ratings, suggesting that the directionality of variations in F0 does not convey emotional stress information. Mean and maximum F0 within an utterance correlated highly with stress ratings, but the range of F0 did not correlate significantly with the stress ratings, especially after the effect of maximum F0 was removed in stepwise regression. It is concluded that the range of F0 per se does not contribute to the perception of emotional stress, whereas maximum F0 constitutes the primary indicator. The observed effects held across several voices that were found to sound natural (three male voices and one of two female ones). An effect of the formant frequencies was also observed in the stimuli with the lowest F0; it is hypothesized that formant frequency structure dominated the F0 effect in the one voice that gave discrepant results

[128] Mandar A. Rahurkar, John H. L. Hansen, James Meyerhoff, George Saviolakis, and Michael Koenig. Frequency band analysis for stress detection using a teager energy

operator based feature. In *Proceedings of the International Conference on Spoken Language Processing*, pages 2021–2024, Denver, Colorado, USA, September 16–20 2002.

Abstract: Studies have shown that the performance of speech recognition algorithms severely degrade due to the presence of task and emotional induced stress in adverse conditions. This paper addresses the problem of detecting the presence of stress in speech by analyzing nonlinear feature characteristics in specific frequency bands. The framework of the previously derived Teager Energy Operator(TEO) based feature TEO-CB-AutoEnv is used. A new detection scheme is proposed based on weighted TEO features derived from critical bands frequencies. The new detection framework is evaluated on a military speech corpus collected in a Soldier of the Quarter (SOQ) paradigm. Heart rate and blood pressure measurements confirm subjects were under stress. Using the traditional TEO-CB-AutoEnv feature with an HMM trained stressed speech classifier, we show error rates of 22.5% and 13% for stress and neutral speech detection. With the new weighted sub-band detection scheme, detection error rates are reduced to 4.7% and 4.6% for stress and neutral detection, a relative error reduction of 79.1% and 64.6% respectively. Finally we discuss issues related to generation of stress anchor models and speaker dependency

[129] Mandar A. Rahurkar and John H.L. Hansen. Frequency distribution based weighted sub-band approach for classification of emotional/stressful content in speech. In *Proceedings of the European Conference on Speech Communication and Technology*, pages 721–724, 2003.

Abstract: In this paper we explore the use of nonlinear Teager Energy Operator based features derived from multiresolution sub-band analysis for classification of motional/stressful speech. We propose a novel scheme for automatic sub-band weighting in an effort towards developing a generic algorithm for understanding emotion or stress in speech. We evaluate the proposed algorithm using a corpus of audio material from a military stressful Soldier of the Quarter Board evaluation panel. We establish classification performance of emotional/stressful speech using an open speaker set with open test tokens. With the new frequency distribution based scheme, we obtain a relative detection error reduction of 81.3% in stress speech, and a 75.4% relative detection rate reduction in neutral speech detection error rate. The results suggest a important step forward in establishing an effective processing scheme for developing generic models of neutral and emotional speech

[130] P. Rajasekaran, G. Doddington, and J. Picone. Recognition of speech under stress and in noise. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 11, pages 733–736, Dallas, Texas, USA, April 1986.

Abstract: Speech recognizers trained in one condition but operating in a different condition degrade in performance. Typical of this situation is when the recognizer is trained under normal conditions but operated in a stressful and noisy environment as in military applications. This paper reports on recognition experiments conducted with a "simulated stress" data base using a baseline algorithm and its modifications. These algorithms perform acceptably well (1% substitution rate) for a vocabulary of 105 words under normal conditions, but degrade by an order of magnitude under the "stress" conditions. The experiments also show that the speech production variation caused by noise exposure at the ear is far more deleterious than ambient acoustic noise with a noise cancelling microphone

[131] A.H. Roscoe. Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biol Psychol*, 34(2–3):259–287, November 1992.

Abstract: The application of heart rate and respiratory measures to the human factors of flight is discussed. The concept of pilot workload is related to the concept of arousal and distinguished from physical workload. Finds from studies of pilot workload using heart rate measures are reviewed for flight, simulated flight and related real-life challenges. Measurement techniques and transducers are discussed from the perspective of field measurement. Recommended procedures are presented as are directions for future work

[132] L.J.M. Rothkrantz, P. Wiggers, J.W.A. van Wees, and R.J. van Vark. Voice stress analysis. In Petr Sojka, Ivan Kopeček, and Karel Pala, editors, *Proceedings of the 7th International Conference on Text, Speech and Dialogue—TSD 2004*, Lecture Notes in Artificial Intelligence LNCS/LNAI 3206, pages 449–456. Springer-Verlag, Brno, Czech Republic, September 2004.

Abstract: The nonverbal content of speech carries information about the physiological and psychological condition of the speaker. Psychological stress is a pathological element of this condition, of which the cause is accepted to be 'workload'. Objective, quantifiable correlates of stress are searched for by means of measuring the acoustic modifications of the voice brought about by workload. Different voice features from the speech signal to be influenced by stress are: loudness, fundamental frequency, jitter, zero-crossing rate, speech rate and high-energy frequency ratio. To examine the effect of workload on speech production an experiment was designed. 108 native speakers of Dutch were recruited to participate in a stress test (Stroop test). The experiment and the analysis of the test results will be reported in this paper

[133] Robert Ruiz, Emmanuelle Absil, Bernard Harmegnies, Claude Legros, and Dolors Poch. Time- and spectrum-related variabilities in stressed speech under laboratory and real conditions. *Speech Communication*, 20(1-2):111–129, 1996.

Abstract: Stress induced by various types of situation leads to vocal signal modifications. Previous studies have indicated that stressed speech is associated with a higher fundamental frequency and noticeable changes in vowel spectrum. This paper presents pitch- and spectral-based analyses of stressed speech corpora drawn from both artificial and real situations. The laboratory corpus is obtained by means of the Stroop test, the real-case corpus is extracted from the Cockpit Voice Recording of a crashed aeroplane. Analyses relative to pitch are presented and an index of microprosodic variation, [mu], is introduced. Spectrum-related indicators of stress are issued from a cumulative histogram of sound level and from statistical analyses of formant frequencies. Distances to the F1-F2-F3 centre are also investigated. All these variations, throughout the two different situations, show the direct link between some new vocal parameters and stress appearances. The results confirm the validity of laboratory experiments on stress, but emphazise quantitative as well as qualitative differences between the situations and the speakers involved

[134] J. A. Russell. Is there universal recognition of emotion from facial expression? *Psychol. Bull.*, 115(1):102–141, 1994.

[135] Evan Ruzanski, John H.L. Hansen, James Meyerhoff, George Saviolakis, and Michael Koenig. Effects of phoneme characteristics on TEO feature-based automatic stress detection in speech. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume I, pages 357–360, Philadelphia, USA, March 18-23 2005.

Abstract: A major challenge of automatic speech recognition systems foundin many areas of todays society is the ability to overcome naturalphoneme conditions that potentially degrade performance. In thisstudy, we discuss the effects of two critical phonemecharacteristics, decreased vowel duration and mismatched voweltype, on the performance of automatic stress detection in speechusing Teager Energy Operator features. We determine the scopeand magnitude of these effects on stress detection performanceand propose an algorithm to compensate for vowel type andduration shortening on stress detection performance using acomposite phoneme decision scheme, which results in relativeerror reductions of 24% and 39% in the non-stress and stressconditions, respectively.

[136] A. Samel, H.M. Wegmann, M. Vejvoda, J. Drescher, A. Gundel, D. Manzey, and J. Wenzel. Two-crew operations: stress and fatigue during long-haul night flights. *Aviation, Space and Environmental Medicine*, 68(8):679–87, August 1997.

Abstract: BACKGROUND: As part of a research program concerning legal aspects of two-pilot operations on long-haul routes, the purpose of the study was to investigate two-crew extended range operations during a flight roster with two consecutive night flights and a short layover. HYPOTHESIS: Present flight time regulations may not be adequate for two-crew minimum operations. METHODS: The study was conducted in cooperation with a German airline company on the route Frankfurt (FRA)-Mahe (SEZ). There were 11 rotations (22 flights) that were investigated by pre-, in- and post-flight data collection each time from the two pilots. Recordings included sleep, taskload, fatigue and stress by measurement of EEG, ECG, motor activity, and subjective ratings. The average actual flight times were 9:15 h (FRA-SEZ) and 9:53 h (SEZ-FRA). All flights took place at night. The layover duration in Mahe was 13:30 h during day-time. RESULTS: During layover, sleep was shortened by 2 h on average compared with 8-h baseline sleep. The two consecutive night duties resulted in a sleep loss of 9.3 h upon return to home base. Inflight ratings of taskload showed moderate grades, but for fatigue ratings an increasing level was observed. Fatigue was more pronounced during the return flight and several pilots scored their fatigue at a critical level. Motor activity, brainwave activity (occurrences of micro-events) and heart rate indicated drowsiness and a low state of vigilance and alertness during both night flights, but these effects were more pronounced during the second flight. CONCLUSIONS:

From the findings it is concluded that a duty roster, as conducted in this study, may impose excessive demands on mental and physiological capacity

[137] R. Sarikaya and J.N. Gowdy. Wavelet based analysis of speech under stress. In *Proceedings. IEEE Southeastcon, 'Engineering new New Century'*, pages 92–96, Blacksburg, VA USA, April 12-14 1997.

Abstract: Stress and its effects in speech signals have been studied by many researchers. A number of studies have been attempted to determine reliable acoustic indicators of stress using such speech production features as fundamental frequency (F0), intensity, spectral tilt, the distribution of spectral energy and others. The findings indicate that more work is necessary to propose a general solution. The goal of this study is to propose a new set of speech features based on dyadic wavelet transform (DyWT) coefficients as potential stress sensitive parameters. The parameters' ability to capture different stress types is evaluated on the basis of separability distance measure between parameters of each stress class for four stress conditions: neutral, loud, question and angry. After an extensive number of scatter distributions were considered a number of clear trends emerged that confirmed that the new speech parameters are well suited for stress classification

[138] R. Sarikaya and J.N. Gowdy. Subband based classification of speech under stress. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 569–572, Seattle, WA USA, May 12-15 1998.

Abstract: This study proposes a new set of feature parameters based on subband analysis of the speech signal for classification of speech under stress. The new speech features are scale energy (SE), autocorrelation-scale-energy (ACSE), subband based cepstral parameters (SC), and autocorrelation-SC (ACSC). The parameters' ability to capture different stress types is compared to widely used mel-scale cepstrum based representations: mel-frequency cepstral coefficients (MFCC) and autocorrelation-mel-scale (AC-mel). Next, a feedforward neural network is formulated for speaker-dependent stress classification of 10 stress conditions: angry, clear, cond50/70, fast, loud, lombard, neutral, question, slow, and soft. The classification algorithm is evaluated using a previously established stressed speech database (SUSAS) (Hansen and Bou-Ghazale 1997). Subband based features are shown to achieve +7.3% and +9.1% increase in the classification rates over the MFCC based parameters for ungrouped and grouped stress closed vocabulary test scenarios respectively. Moreover the average scores across the simulations of new features are +8.6% and +13.6% higher than MFCC based features for the ungrouped and grouped stress test scenarios respectively

[139] R. Sarikaya and J.H.L. Hansen. High resolution speech feature parametrization for monophone-based stressed speech recognition. *IEEE Signal Processing Letters*, 7(7):182–185, 2000.

Abstract: This letter investigates the impact of stress on monophone speech recognition accuracy and proposes a new set of acoustic parameters based on high resolution wavelet analysis. The two parameter schemes are entitled wavelet packet parameters (WPP) and subband-based cepstral parameters (SBC). The performance of these features is compared to traditional Mel-frequency cepstral coefficients (MFCC) for stressed speech monophone recognition. The stressed speaking styles considered are neutral, angry, loud, and Lombard effect speech from the SUSAS database. An overall monophone recognition improvement of 20.4% and 17.2% is achieved for loud and angry stressed speech, with a corresponding increase in the neutral monophone rate of 9.9% over MFCC parameters

[140] N. Sarkar. Psychophysiological control architecture for human-robot coordination-concepts and initial experiments. In *Proceedings of IEEE International Conference on Robotics and Automation, ICRA '02*, volume 4, pages 3719–3724, May 11-15 2002.

Abstract: The use of robots is expected to be pervasive in many spheres of society: in hospitals, homes, offices and battlefields, where the robots will need to interact and cooperate closely with a variety of people. The paper proposes an innovative approach to human-robot cooperation where the robot will be able to recognize the psychological state of the interacting human and modify its (i.e., robot's) own action to make the human feel comfortable in working with the robot. Wearable biofeedback sensors are used to measure a variety of physiological indices to infer the underlying psychological states (affective states) of the human. The eventual idea is to correlate the psychological states with the actions of the robot to determine which action(s) is responsible for a particular affective state. The robot controller will then modify that action if there is a need to alter the affective state. A concept of such a control architecture, a requirement analysis, and initial results from human experiments for stress detection are presented

[141] H. Sato, Y. Mitsukura, M. Fukumi, and N.; Akamatsu. Emotional speech classification with prosodic parameters by using neural networks. In *The Seventh Australian and New Zealand Intelligent Information Systems Conference*, pages 395–398, November 2001.

Abstract: Interestingly, in order to achieve a new Human Interface such that digital computers can deal with the KASEI information, the study of the KANSEI information processing recently has been approached. In this paper, we propose a new classification method of emotional speech by analyzing feature parameters obtained from the emotional speech and by learning them using neural networks, which is regarded as a KANSEI information processing. In the present research, KANSEI information is usually human emotion. The emotion is classified broadly into four patterns such as neutral, anger, sad and joy. The pitch as one of feature parameters governs voice modulation, and can be sensitive to change of emotion. The pitch is extracted from each emotional speech by the cepstrum method. Input values of neural networks (NNs) are then emotional pitch patterns, which are time-varying. It is shown that NNs can achieve classification of emotion by learning each emotional pitch pattern by means of computer simulations

[142] Mark W. Scerbo, Frederick G. Freeman, Peter J. Mikulka, Raja Parasuraman, Francesco Di Nocero, and Lawrence J. Prinzel III. The efficacy of psychophysiological measures for implementing adaptive technology. Technical Report NASA/TP-2001-211018, NASA, 2001.

Abstract: Adaptive automation refers to technology that can change its mode of operation dynamically. Further, both the technology and the operator can initiate changes in the level or mode of automation. The present paper reviews research on adaptive technology. The paper is intended as a guide and review for those seeking to use psychophysiological measures in design and assessing adaptively automated systems. It is divided into four primary sections. In the first section, issues surrounding the development and implementation of adaptive automation are presented. Because physiological-based measures show much promise for implementing adaptive automation, the second section is devoted to examining candidate indices and reviews some of the current research on these measures as they relate to workload. In the third section, detailed discussion is devoted to electroencephalogram (EEG) and eventrelated potentials (ERPs) measures of workload. The final section provides an example of how psychophysiological measures can be used in adaptive automation design

[143] Klaus R. Scherer. Adding the affective dimension: A new look in speech analysis and synthesis. In *Proceedings of the International Conference on Spoken Language Processing*, pages 1811–1814, 1996.

Abstract: This introduction to a special session on 'Emotion in recognition and synthesis' highlights the need to understand the effects of affective speaker states on voice and speech on a psychophysiological level. It is argued that major advances in speaker verification, speech recognition, and natural-sounding speech synthesis depend on increases in our knowledge of the mechanisms underlying voice and speech production under emotional arousal or other attitudinal states, as well as on a more adequate understanding of listener decoding of affect from vocal quality. A brief review of the current state of the art is provided

[144] Klaus R. Scherer. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1-2):227–256, April 2003.

Abstract: The current state of research on emotion effects on voice and speech is reviewed and issues for future research efforts are discussed. In particular, it is suggested to use the Brunswikian lens model as previous termanext term base for research on the previous termvocal communication of emotion.next term This approach allows one to model the complete process, including both encoding (expression), transmission, and decoding (impression) of previous termvocalnext term emotion communication. Special emphasis is placed on the conceptualization and operationalization of the major elements of the model (i.e., the speaker's emotional state, the listener's attribution, and the mediating acoustic cues). In addition, the advantages and disadvantages of research previous termparadigmsnext term for the induction or observation of emotional expression in voice and speech and the experimental manipulation of previous termvocalnext term cues are discussed, using pertinent examples drawn from past and present research

[145] Klaus R. Scherer, D. Grandjean, Tom Johnstone, Gudrun Klasmeyer, and Thomas Bänziger. Acoustic correlates of task load and stress. In *Proceedings of the Inter-*

national Conference on Spoken Language Processing, pages 2017–2020, Denver, Colorado, USA, September 16–20 2002.

Abstract: It is argued that reliable acoustic profiles of speech under stress can only be found if different types of stress are clearly distinguished and experimentally induced. We report first results of a study with 100 speakers from three language groups, using a computer-based induction procedure that allows distinguishing cognitive load due to task engagement from psychological stress. Findings show significant effects of load, and partly of stress, for speech rate, energy contour, F0, and spectral parameters. It is further suggested that the mean results for the complete sample of speakers do not reflect the amplitude of stress effects on the voice. Future research should isolate and focus on speakers for whom the psychological stress induction has been successful

[146] Klaus R. Scherer, Tom Johnstone, Gudrun Klasmeyer, and Thomas Bänziger. Can automatic speaker verification be improved by training the algorithms on emotional speech? In *Proceedings of the International Conference on Spoken Language Processing*, volume 2, pages 807–810, Beijing, China, October 16-20 2000.

Abstract: The ongoing work described in this contribution attempts to demonstrate the need to train ASV algorithms on emotional speech, in addition to neutral speech, in order to achieve more robust results in real life verification situations. A computerized induction program with 6 different tasks, producing different types of stressful or emotional speaker states, was developed, pretested, and used to record French, German, and English speaking participants. For a subset of these speakers, physiological data were obtained to determine the degree of physiological arousal produced by the emotion inductions and to determine the correlation between physiological responses and voice production as revealed in acoustic parameters. In collaboration with a commercial ASV provider (Ensigma Ltd.), a standard verification procedure was applied to this speech material. This paper reports the first set of preliminary analyses for the subset of 30 German speakers. It is concluded that an evaluation of the promise of training ASV material on emotional speech requires in-depth analyses of the individual differences in vocal reactivity and further exploration of the link between acoustic changes under stress or emotion and verification results

[147] I. Shahin and N. Botros. Modeling and analyzing the vocal tract under normal and stressful. In *Proceedings IEEE SoutheastCon 2001*, pages 213–220, 2001.

Abstract: In this research, we model and analyze the vocal tract under normal and stressful talking conditions. This research answers the question of the degradation in the recognition performance of text-dependent speaker identification under stressful talking conditions. This research can be used (for future research) to improve the recognition performance under stressful talking conditions.

[148] R. P. Sloan, J. B. Korten, and M. M. Myers. Components of heart rate reactivity during mental arithmetic with and without speaking. *Physiology & Behavior*, 50(5):1039–1045, November 1991.

Abstract: Reactivity to psychological stressors has been hypothesized to be related to the development of cardiovascular disease. Because mental arithmetic (MA) has been shown to produce to produce significant increases in heart rate and blood pressure, it is one of the most commonly utilized laboratory psychological stressors. However, the use of MA to assess hemodynamic reactivity raises two issues: 1) increases in heart rate can be produced by vagal withdrawal, sympathetic activation, or a combination of the two, and these mechanisms may differ in their pathogenic implications; and 2) in most MA studies, subjects are instructed to perform the task aloud, thus raising the possibility that speaking may interfere with respiratory patterns which in turn can influence hemodynamic outcomes. To address these two issues, we studied heart rate responses of 10 subjects to 2 different versions of MA and a control condition in which vocalization of answers was manipulated. Heart rate (HR), heart rate variability (HRV) in the low frequency (LB) and respiratory frequency (RSA) bands, and respiratory rate were measured. Results indicate that, although the two task conditions produced similar heart rate increases, RSA decreased only in the nonspeaking condition. Overall, the findings suggest that HR changes during MA are attributable to vagal withdrawal but that vocalization of answers during the task interferes with analysis of HRV, thus obscuring the mechanisms responsible for these changes.

[149] R.E. Slyh, W.T. Nelson, and E.G. Hansen. Analysis of mrate, shimmer, jitter, and f0 contour features across stress and speaking style in the susas database. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 2091–2094, Phoenix, AZ USA, March 15-19 1999.

Abstract: This paper highlights the results of an investigation of several features cross the style classes of the 'simulated' portion of the SUSAS database. The features considered here include a previously-introduced measure of speaking rate called mrate, measures of shimmer, measures of jitter, and features derived from fundamental frequency (F0) contours. The F0 contour features are the means of F0 and ΔF0 over the first, middle, and last thirds of the ordered set of voiced frames for each word. The mrate exhibits differences between the fast, neutral, and slow styles and between the loud, neutral, and soft styles. Shimmer and jitter exhibit differences that are similar to those of mrate; however, the shimmer and jitter differences are less consistent than the mrate differences across the speakers in the database. Several F0 contour features exhibit differences between the angry, loud, Lombard, and question styles and most of the other styles

[150] H.J.M. Steeneken and J.H.L. Hansen. Speech under stress conditions: overview of the effect on speech production and on system performance. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 2079–2082, Phoenix, AZ USA, March 15-19 1999.

Abstract: The NATO research study group on 'Speech and Language Technology' recently completed a three year project on the effect of 'stress' on speech production and system performance. For this purpose various speech databases were collected. A definition of various states of stress and the corresponding type of stressor is proposed. Results are reported from analysis and assessment studies performed with the databases collected for this project

[151] FJ Tolkmitt and KR Scherer. Effect of experimentally induced stress on vocal parameters. *J Exp Psychol Hum Percept Perform*, 12(3):302–13, August 1986.

Abstract: In a factorially designed experiment the factors Mode of Stress (cognitive vs. emotional) and Degree of Stress (low vs. high) were studied in their effect on phonatory and articulatory processes during speech production in groups of male and female students (factor Sex) selected to represent three types of personality (factor Coping Style: low anxiety, high anxiety, anxiety denying). Using digital speech analysis procedures, mean F0, F0 floor, formant location, and spectral energy distribution were extracted from the experimental speech samples. Although mean F0 did not covary with stress manipulation, F0 floor of high-anxious and anxiety-denying subjects increased with stress, probably due to physiologically based changes in muscle tension. For articulatory processes, as measured through formant location and spectral composition, significant changes were found for females. For anxiety-denying female subjects, precision of articulation increased under cognitive stress and decreased under emotional stress. The results are interpreted in terms of differential susceptibility to stress

[152] N. Tosa and R. Nakatsu. Life-like communication agent-emotion sensing character MIC and feeling session character MUSE. In *Third IEEE International Conference on Multimedia Computing and Systems*, pages 12–19, Hiroshima, Japan, 1996.

Abstract: Why do people, regardless of age or gender, have an affinity for objects manifested in the human form? From the earthen figures of ancient times to mechanical dolls, teddy bears and robots, is it not true that man has conceived such objects in his imagination, then formed attachments and transferred emotions to them? In this paper, we address the issues of communication and esthetics of artificial life that possess this 'human form' in modern society, both from artistic and engineering standpoints. An example is presented in which emotions are interpreted from human voices and emotional responses are triggered within the interactive setting of 'MIC & MUSE.' MIC is a cute male child who is an emotion communication character. MUSE is a goddess; she has the capability of musical communication

[153] Hartmut Traunmüller. Evidence for demodulation in speech perception. In *Proceedings of the 6th ICSLP*, volume 3, pages 790–793, Beijing, China, October 16-20 2000.

Abstract: According to the Modulation Theory, speakers modulate their voice with linguistic gestures, and listeners demodulate the signal in order to separate the linguistic from the expressive and organic information. Listeners tune in to the carrier (the voice) on the basis of an analysis of a stretch of speech and they evaluate its modulation. This is reflected in many perceptual experiments that involved manipulated introductory phrases, blocked vs. randomized speakers, and other non-linguistic variables.

[154] Terence J. Turner and Andrew Ortony. Basic emotions: Can conflicting criteria converge? *Psychological Review*, 99(3):556–571, 1992.

[155] Arie M. van Roon, Lambertus J.M. Mulder, Monika Althaus, and Gijsbertus Mulder. Introducing a baroreflex model for studying cardiovascular effects of mental workload. *Psychophysiology*, 41(6):961–981, 2004.

Abstract: A quantitative baroreflex control model is presented aimed at estimating differences in autonomic activation due to mental task performance. The model, which builds on earlier work of Wesseling and colleagues, is strongly supported by well-established knowledge of physiological control processes. Spectral measures of heart rate and blood pressure variability provide the information to estimate autonomic gain and tone parameters. The article gives a detailed model description as well as an evaluation in terms of spectral variability distributions, respiratory sinus arrhythmia, and vagal control mechanisms. The estimation procedure is outlined while presenting two studies that describe the effects of mental workload and vagal blockade, respectively. It is concluded that, using the model approach, cardiovascular effects of mental task performance can be interpreted in terms of specific changes in autonomic state. The model is implemented in a Matlab/Simulink environment and is available for other researchers in the field.

[156] J. A. Veltman. A comparative study of psychophysiological reactions during simulator and real flight. *International Journal of Aviation Psychology*, 12(1):33–48, 2002.

Abstract: During selection tests in a flight simulator and a real aircraft, physiological workload measures were evaluated. The selection context guaranteed high motivation in the participant to exert additional effort during difficult flight tasks. The aim of the study was to obtain information about the sensitivity to mental effort of several physiological measures and to explore the applicability of these measures in real flight. The following measures were used: heart rate, heart rate variability, respiration (frequency and amplitude), blood pressure, eye blinks (frequency, duration, and amplitude), and saliva cortisol. Blood pressure was the only measure that was difficult to obtain during real flights

[157] J. A. Veltman and A. W. Gaillard. Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, 41(5):656–69, May 1998.

Abstract: The sensitivity of physiological measures to mental workload was investigated in a flight simulator. Twelve pilots had to fly through a tunnel with varying levels of difficulty. Additionally, they had to perform a memory task with four levels of difficulty. The easiest memory task was combined with the easiest tunnel task and the most difficult memory task with the most difficult tunnel task. Between the tunnel tasks, subjects had to fly a pursuit task in which a target jet had to be followed. Rest periods before and after the experiment were used as a baseline for the physiological measures. Mental workload was measured with heart period, continuous blood pressure, respiration and eye blinks. Several respiratory parameters, heart rate variability, blood pressure variability and the gain between systolic blood pressure and heart period (modulus) were scored. All measures showed differences between rest and flight, and between the pursuit and the tunnel task. Only heart period was sensitive to difficulty levels in the tunnel task. Heart rate variability increased when respiratory activity around 0.10 Hz increased, which occurred often. The modulus was hardly influenced by respiration and therefore appears to be a better measure than heart rate variability. Among the respiratory parameters, the duration of a respiratory cycle was the most sensitive to changes in workload. The time in between two successive eye blinks (blink interval) increased and the blink duration decreased as more visual information had to be processed. Increasing the difficulty of the memory task led to a decrement in blink interval, probably caused by subvocal activity during rehearsal of target letters. The data show that physiological measures are sensitive to mental effort, whereas rating scales are sensitive to both mental effort and task difficulty.

[158] J. A. Veltman and A. W. K Gaillard. Physiological indices of workload in a simulated flight task. *Biological Psychology*, 42(3):323–342, 1996.

Abstract: The sensitivity of physiological measures to evaluate workload was investigated in a simulated flight task. Heart rate, blood pressure (from beat to beat), respiration and eye blinks were recorded in 14 subjects while they performed a complex task in a flight simulator. Workload was manipulated by introducing an additional task and by varying the task difficulty of segments of the flight scenarios. Heart rate and blood pressure were both affected by the different levels of task difficulty. Heart-rate variability was found to be confounded by respiration. Slow respiratory activity contributed considerably to heart rate variability, especially after periods of high workload (for example, after landing). The gain between blood-pressure and heart-rate variability (modulus) was sensitive to mental effort and was not influences by respiration. Eye blinks, in particular the duration, were specifically affected by the visual demands of the task and not by the workload in general. When subjects had to process visual information, the number and duration of blinks decreased

[159] Bruno Verschuere, Geert Crombez, Armand De Clercq, and Ernst H W Koster. Autonomic and behavioral responding to concealed information: differentiating orienting and defensive responses. *Psychophysiology*, 41(3):461–6, May 2004.

Abstract: A mock crime experiment was conducted to examine whether enhanced responding to concealed information during a polygraph examination is due to orienting or defensive responding. Thirty-six undergraduate students enacted one of two mock crimes. Pictures related to both crimes were presented while heart rate, magnitude of the skin conductance response, and reaction times to a secondary probe were measured. Compared to control pictures, participants showed heart rate deceleration and enhanced electrodermal responding to pictures of the crime they had committed. Probe reaction times did not differ significantly between crime and control pictures. The present findings support the idea that the orienting reflex accounts for the enhanced responding to concealed information. Theoretical and practical implications of the orienting account are discussed.

[160] Alex Vincent, Fergus I. M. Craik, and John J. Furedy. Relations among memory performance, mental workload and cardiovascular responses. *International Journal of Psychophysiology*, 23(3):181–198, October 1996.

Abstract: The levels of processing paradigm has been a powerful research framework in the study of memory for close to a quarter century. However, an objective index of depth of processing is still lacking. Two experiments using lists of words, presented to male subjects, were performed to compare the effects of depth of processing, rate of presentation, and task incentive on recognition memory performance, self-reported workload, and cardiovascular responding. Memory performance results from the two experiments demonstrated higher recognition levels associated with deeper processing and slower presentation rates. Deeply encoded items were associated with faster recognition latencies. Self-reported workload levels were higher for deeper processing and faster presentation rates. Cardiovascular responses were generally amplified with the addition of a task incentive. Increased blood pressure was associated with faster presentation rates. Increased heart rate and decreased T-wave amplitude (i.e., increased sympathetic activity) were uniquely associated with the deep encoding of information presented at the fastest rate. This particular encoding condition was associated with increased recognition levels. Deeply encoded items were associated with increased suppression of heart rate variability during recognition. This combination of behavioral and cardiovascular measures may provide the basis for an objective index of depth of processing.

[161] Jeff Waters, Steve Nunn, Brenda Gillcrist, and Eric VonColln. Detection of stress by voice: Analysis of the glottal pulse. Technical Report NRaD TR 1652, Space and Naval Warfare Systems Center San Diego, 1994.

[162] D.W. Watson. Physiological correlates of heart rate variability (hrv) and the subjective assessment of workload and fatigue in-flight crew: a practical study. In *Human Interfaces in Control Rooms, Cockpits and Command Centres, 2001. People in Control. The Second International Conference on*, pages 159–163, Manchester, UK, June 19-21 2001.

Abstract: Stress, fatigue and performance have historically featured as important physiological variables within aviation medicine. This study reviews the application of heart rate variability and its use as an effective somatic indicator of workload and fatigue in flight crew

[163] Jeffrey Whitmore and Stanley Fisher. Speech during sustained operations. *Speech Communication*, 20(1-2):55–70, 1996.

Abstract: Research was conducted to determine if alterations in the acoustical characteristics of voice occur over periods of sustained operations. Twelve male United States Air Force B-1B bomber aircrewmen participated in the study. The participants served in crews of four and performed three 36-hour experimental periods (missions) in a high-fidelity simulator. The missions were interspersed with 36-hour rest breaks. Data were lost from two members of the third team due to a communication malfunction. Speech, cognitive and subjective fatigue data were collected approximately every three hours for 11 trials per mission. Fundamental frequency and word duration were both found to vary significantly over trials (fundamental frequency $F(10,90) = 2.63$, $p = 0.0076$, word duration $F(10,90) = 2.5$, $p = 0.0106$). Speech duration results also showed a significant main effect of mission ($F(2,18) = 6.91$, $p = 0.0082$). The speech data follow the same trend as the data from the cognitive tests and subjective measures. A strong diurnal pattern is reflected in nearly all of the dependent measures. Overall, the results support the proposition that voice may be a valid indicator of a speaker's fatigue state

[164] Glenn F. Wilson. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *International Journal of Aviation Psychology*, 12(1):3–18, 2001.

Abstract: Piloting an aircraft is a complex task that places demands on several aspects of a pilot's cognitive capabilities. Because of the multifaceted nature of flying, several measures are required to identify the effects of these demands on the pilot. Several psychophysiological measures were recorded so that a wider understanding of the effects of these demands could be achieved. Heart rate, heart rate variability, eye blinks, electrodermal activity, topographically recorded electrical brain activity, and subjective estimates of mental workload were recorded. Ten pilots flew an approximately 90-min scenario containing both visual and instrument flight conditions. To determine the reliability of the psychophysiological measures, the pilots flew the same scenario twice. The responses during the 2 flights were essentially identical. Cardiac and electrodermal measures were highly correlated and exhibited changes in response to the various demands of the flights. Heart rate variability was less sensitive than heart rate. Alpha and delta bands of the brain activity showed significant changes to the varying demands of the scenarios. Blink rates decreased during the more highly visually demanding segments of the flights

[165] Glenn F Wilson and Chris A Russell. Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Human Factors*, 45(3):381–9, 2003.

Abstract: We studied 2 classifiers to determine their ability to discriminate among 4 levels of mental workload during a simulated air traffic control task using psychophysiological measures. Data from 7 air traffic controllers were used to train and test artificial neural network and stepwise discriminant classifiers. Very high levels of classification accuracy were achieved by both classifiers. When the 2 task difficulty manipulations were tested separately, the percentage correct classifications were between 84% and 88%. Feature reduction using saliency analysis for the artificial neural networks resulted in a mean of 90% correct classification accuracy. Considering the data as a 2-class problem, acceptable load versus overload, resulted in almost perfect classification accuracies, with mean percentage correct of 98%. In applied situations, the most important distinction among operator functional states would be to detect mental overload situations. These results suggest that psychophysiological data are capable of such discriminations with high levels of accuracy. Potential applications of this research include test and evaluation of new and modified systems and adaptive aiding

[166] Peter Wittels, Bernd Johannes, Robert Enne, Karl Kirsch, and Hanns-Christian Gunga. Voice monitoring to measure emotional load during short-term stress. *European Journal of Applied Physiology*, 87(3):278 – 282, July 2002.

Abstract: It is well known that there is a relationship between the voice the human emotional status. Previous studies have demonstrated that changes of fundamental frequency (f0), in particular, have a significant relationship with emotional load. The aim of the present study was to investigate how f0 changes in response to an unknown emotionally stressful task under real-life conditions. A further question was whether repetitions of this task lead to an adaptation of f0, indicating a lower emotional load. The participants of this study included 26 healthy males. f0 and heart rate (fc) were recorded for baseline testing (BLT) under relaxed laboratory conditions. Then the participants were asked to negotiate a natural obstacle by way of sliding down a rope hanging from a handlebar without any safety provisions, thus being exposed to the danger of a fall from a height of up to 12 m into shallow water (guerrilla slide I, GSI). The task was repeated after 30 min (GSII) and after 3 days of physical strain (GSIII). Immediately before starting the task the participants were asked to give a standardised speech sample, during which fc was recorded. The mode value of f0 (f0,mode) of the speech samples was used for further analysis. The mean (SD) value of f0,mode at BLT was 114.9 (14.8) Hz; this increased to 138.8 (19.6) Hz at GSI ($P<0.000$), decreased to 135.9 (19.6) Hz at GSII and to 130.0 (21.5) Hz at GSIII ($P=0.012$). The increase in fc was significantly different from BLT to GSI ($P<0.000$). The repetitions of the task did not produce significant changes in fc. It was shown that f0,mode is a sensitive parameter to describe changes in emotional load, at least in response to short-term psychoemotional stress, and seems to throw light upon the amount of adaptation caused by increased experience.

[167] Rogier Woltjer. Adaptive multi-modal interfaces – a composite research field. In *1st Nordic Symposium on Multimodal Communication*, pages 197–210, Copenhagen, Danmark, September 2003.

Abstract: This paper outlines the field of Adaptive Multi-Modal Interfaces (AMMI) by sketching its theoretical background from related fields of psychology, human factors, and artificial intelligence. These interfaces adapt to the situation of the operator of a system and of the environment, and let the

operator interact with the system through various sensory modalities. From this outline it becomes clear that several theories suggest that the diverse functionality of the adaptive multi-modal interface may be able to present information in a more comprehensible way than conventional interfaces. The advantages are greatest when an operator or user is working in highly stressful and complex contexts. Since this context can rapidly change in many complex dynamic operator contexts (such as controlling a transportation system or an industrial process), contextually-sensitive multi-modal adaptation of the interface may enable the operator to maintain better control and performance. Additional theoretic and applied research will be needed to guide the design of adaptive multi-modal interfaces so that their diverse functionality can be integrated in a natural way that facilitates human cognition

[168] B.D Womack and J.H.L. Hansen. Stressed speech recognition using multi-dimensional hidden markov models. In *Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding*, pages 404–411, Santa Barbara, CA USA, December 14–17 1997.

Abstract: Robust speech recognition systems must address variations due to perceptually induced stress in order to maintain acceptable levels of performance in adverse conditions. This study proposes a new approach which combines stress classification and speech recognition into one algorithm. This is accomplished by generalizing the one-dimensional hidden Markov model to a multi-dimensional hidden Markov model (N-D HMM) where each stressed speech style is allocated a dimension in the N-D HMM. It is shown that this formulation better integrates perceptually induced stress effects for stress independent recognition. This is due to the sub-phoneme (state level) stress classification that is implicitly performed by the algorithm. The proposed N-D HMM method is compared to neutral and multi-styled stress trained 1-D HMM recognizers. Average recognition rates are shown to improve by +15.72% over the 1-D stress dependent recognizer and 26.67% over the 1-D neutral trained recognizer

[169] B.D. Womack and J.H.L Hansen. N-channel hidden markov models for combined stressed speech classification and recognition. *IEEE Transactions on Speech and Audio Processing*, 7(6):668–677, 1999.

Abstract: Robust speech recognition systems must address variations due to perceptually induced stress in order to maintain acceptable levels of performance in adverse conditions. One approach for addressing these variations is to utilize front-end stress classification to direct a stress dependent recognition algorithm which separately models each speech production domain. This study proposes a new approach which combines stress classification and speech recognition functions into one algorithm. This is accomplished by generalizing the one-dimensional (1-D) hidden Markov model to an N-channel hidden Markov model (N-channel HMM). Here, each stressed speech production style under consideration is allocated a dimension in the N-channel HMM to model each perceptually induced stress condition. It is shown that this formulation better integrates perceptually induced stress effects for stress independent recognition. This is due to the sub-phoneme (state level) stress classification that is implicitly performed by the algorithm. The proposed N-channel stress independent HMM method is compared to a previously established one-channel stress dependent isolated word recognition system yielding a 73.8% reduction in error rate. In addition, an 82.7% reduction in error rate is observed compared to the common one-channel neutral trained recognition approach

[170] Brian D. Womack and John H. L. Hansen. Speech stress classification and feature analysis using a neural network classifier. *Journal of the Acoustical Society of America*, 96(5):3351–3351, 1994.

Abstract: Variability in speech production due to task induced stress contributes significantly to loss in speech processing performance. A new neural network algorithm is formulated which classifies speech under stress using the susas stress database. Five spectral parameter representations (Mel, Delta-Mel, Delta 2-Mel, Auto-Corr-Mel, and Cross-Corr-Mel cepstral) are considered as potential stressed speech relayers. Eleven stress conditions are considered which include Angry, Clear, Fast, Lombard, Loud, Normal, Question, Slow, and Soft speech. Stressed speech classification using these features is evaluated with respect to (i) pairwise class separability, (ii) an objective measure of pairwise and global feature separability, and (iii) analysis of articulatory vocal tract area functions. Perturbations in speech production under stress are reflected to varying degrees in the five speech feature representations. A neural network classifier over phoneme partitions can achieve good speaker dependent classification performance (80.6%) when stress conditions are combined into six groups of related stress domains. The implication of reliable stress classification will be discussed for applications such as monitoring speaker state, improving naturalness of speech coding, and increasing robustness of speech recognizers in adverse conditions

[171] Brian D. Womack and John H. L Hansen. Classification of speech under stress using target driven features. *Speech Communication*, 20(1-2):131–150, 1996.

Abstract: Speech production variations due to perceptually induced stress contribute significantly to reduced speech processing performance. One approach for assessment of production variations due to stress is to formulate an objective classification of speaker stress based upon the acoustic speech signal. This study proposes an algorithm for estimation of the probability of perceptually induced stress. It is suggested that the resulting stress score could be integrated into robust speech processing algorithms to improve robustness in adverse conditions. First, results from a previous stress classification study are employed to motivate selection of a targeted set of speech features on a per phoneme and stress group level. Analysis of articulatory, excitation and cepstral based features is conducted using a previously established stressed speech database (Speech Under Simulated and Actual Stress (SUSAS)). Stress sensitive targeted feature sets are then selected across ten stress conditions (including Apache helicopter cockpit, Angry, Clear, Lombard effect. Loud, etc.) and incorporated into a new targeted neural network stress classifier. Second, the targeted feature stress classification system is then evaluated and shown to achieve closed speaker, open token classification rates of 91.0%. Finally, the proposed stress classification algorithm is incorporated into a stress directed speech recognition system, where separate hidden Markov model recognizers are trained for each stress condition. An improvement of +10.1% and +15.4% over conventionally trained neutral and multi-style trained recognizers is demonstrated using the new stress directed recognition approach

[172] Hanako Yoshida, Linda B. Smith, Raedy M. Ping, and Elizabeth L. Davis. How are speech and gesture related? In *24th annual meeting of the Cognitive Science Society, (CogSci)*, 2002.

Abstract: People gesture when they speak. Despite considerable attention from a variety of disciplines, the precise nature of the relation between gesture, speech and thought has remained elusive. The research reported here considers two very different hypotheses about the fundamental relationship. By one account, gesture is a consequence of physiological arousal. By another account, gesture use reflects more cognitive processes and is strongly linked to mental representation. This paper seeks the mechanism underlying the link between gesture and speech by showing that both mental representation and physiological arousal are reflected in our gesture use

[173] Hans Zeier, Peter Brauchli, and Helen I Joller-Jemelka. Effects of work demands on immunoglobulin a and cortisol in air traffic controllers. *Biological Psychology*, 42(3):413–423, 1996.

Abstract: The professional activity of air traffic controllers (ATC) is often considered to be rather stressful. Certain characteristics of this job are likely to produce stress; for example an ATC can not predict when a situation becomes critical and he is not able to regulate the workload. In order to assess psychophysiological stress reactions in this working situation, saliva samples were taken from 158 male air traffic controllers before and after each of two working sessions. In contrast to the expected immunosuppressive effects, the working sessions caused a marked increase in the concentration and secretion rate of salivary immunoglobulin A (sIgA), as well as in the concentration of salivary cortisol. The increase in sIgA, however, was not correlated with the salivary cortisol response or with the amount of actual or perceived workload, whereas the cortisol response was correlated with both workload measures. It is suggested that positive emotional engagement is responsible for the observed sIgA increase and that measuring this physiological response may be a valuable tool for differentiating between positive and negative stress effects or between successful and unsuccessful adaptation or coping with situational demands

[174] Li Zhao, Wei Lu, Ye Jiang, and Zhenyang Wu. A study on emotional feature recognition in speech. In *Proceedings of the International Conference on Spoken Language Processing*, pages 961–964, 2000.

Abstract: This paper analysis the feature of the time amplitude pitch and formant construction involved such four emotions as happiness anger surprise and sorrow. Through comparison with non-emotional speech signal, we sum up the distribution law of emotional feature including different emotional speech. Nine emotional features were extracted from emotional speech for recognizing emotion. We introduce three emotional recognition methods based on principal component analysis and the results show that these method can provide an effective solution to emotional recognition

[175] Guojun Zhou, J.H.L. Hansen, and J.F. Kaiser. Classification of speech under stress based on features derived from the nonlinear teager energy operator. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 549–552, Seattle, WA USA, May 12-15 1998.

Abstract: Studies have shown that the distortion introduced by stress or emotion can severely reduce speech recognition accuracy. Techniques for detecting or assessing the presence of stress could help neutralize stressed speech and improve the robustness of speech recognition systems. Although some acoustic variables derived from linear speech production theory have been investigated as indicators of stress, they are not consistent. Three new features derived from the nonlinear Teager (1990) energy operator (TEO) are investigated for stress assessment and classification. It is believed that TEO based features are better able to reflect the nonlinear airflow structure of speech production under adverse stressful conditions. The proposed features outperform stress classification using traditional pitch by +22.5% for the normalized TEO autocorrelation envelope area feature (TEO-Auto-Env), and by +28.8% for TEO based pitch feature (TEO-Pitch). Overall neutral/stress classification rates are more consistent for TEO based features (TEO-Auto-Env: σ=5.15, TEO-pitch: σ=7.83) vs. (pitch: σ=23.40). Also, evaluation results using actual emergency aircraft cockpit stressed speech from NATO show that TEO-Auto-Env works best for stress assessment

[176] Guojun Zhou, J.H.L. Hansen, and J.F Kaiser. Nonlinear feature based classification of speech under stress. *IEEE Transactions on Speech and Audio Processing*, 9(3):201–216, 2001.

Abstract: Studies have shown that variability introduced by stress or emotion can severely reduce speech recognition accuracy. Techniques for detecting or assessing the presence of stress could help improve the robustness of speech recognition systems. Although some acoustic variables derived from linear speech production theory have been investigated as indicators of stress, they are not always consistent. Three new features derived from the nonlinear Teager (1980) energy operator (TEO) are investigated for stress classification. It is believed that the TEO based features are better able to reflect the nonlinear airflow structure of speech production under adverse stressful conditions. The features proposed include TEO-decomposed FM variation (TEO-FM-Var), normalized TEO autocorrelation envelope area (TEO-Auto-Env), and critical band based TEO autocorrelation envelope area (TEO-CB-Auto-Env). The proposed features are evaluated for the task of stress classification using simulated and actual stressed speech and it is shown that the TEO-CB-Auto-Env feature outperforms traditional pitch and mel-frequency cepstrum coefficients (MFCC) substantially. Performance for TEO based features are maintained in both text-dependent and text-independent models, while performance of traditional features degrades in text-independent models. Overall neutral versus stress classification rates are also shown to be more consistent across different stress styles

[177] Guojun Zhou, James F. Kaiser, and John H. L. Hansen. Methods for stress classification: Nonlinear teo and linear speech based features. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 2087–2090, Phoenix, AZ USA, March 15-19 1999.

Abstract: Speech production variations due to perceptually induced stress contribute significantly to reduced speech processing performance. One approach that can improve the robustness of speech processing (e.g., recognition) algorithms against stress is to formulate an objective classification of speaker stress based upon the acoustic speech signal. An overview of methods for stress classification is presented. First, we review traditional pitch-based methods for stress detection and classification. Second, neural network based stress classifiers with cepstral-based features, as well as wavelet-based classification algorithms are considered. The effect of stress on linear speech features is discussed, followed by the application of linear features and the Teager (1990) energy operator (TEO) based nonlinear features for effective stress classification. A new evaluation for stress classification and assessment is presented using a critical band frequency partition based the TEO feature and the combination of several linear features. Results using NATO databases of actual speech under stress are presented. Finally, we discuss issues relating to stress classification across known and unknown speakers and suggest areas for further research