TUM

Institute for
Human-Machine Communication
Activity Report 1997–2000

The Institute for Human-Machine Communication belongs to the Department of Electrical Engineering and Information Technology of the Technische Universität München. In the course of reorganizing the structure of the department in 1998 it became one of its five Institutes for Information and Communication Technology.

The Institute is situated in the southeastern part of the main campus of the Technische Universität München, close to the center of Munich. The institute can easily be reached by public transport (U-Bahn U2, stop Königsplatz). Please use the map on the back side of the cover to find your way.

Contents

Preface

The integration of computer and communication technologies offers a rapidly growing number of information services since the invention of Morse telegraphy. Networked information and communication systems for data, text, voice and image make high-quality multimedia services and multimodal dialog possible beyond geographical and political borders. Due to technological and system-oriented developments these systems are getting more efficient, more cost-effective, but on the other hand also more complex. Not only engineers and computer experts are concerned in practice with computer operated systems but also more and more users with most differing professional background and manyfold tasks. New perspectives of how to obtain information come true for everyone.

Nevertheless, the users of modern information technologies shall not be overstrained by the methods of handling the systems and by the flows of available information. For that reason a system operability which is adequate to the user is not only a research and development aim of high priority but also a decisive quality criterion in the competition for market success. Progress in developing user-friendly, co-operative interfaces depends essentially on an optimal adaptation of the human-machine interaction to the user's sensory, motory, and cognitive capabilities and limits. The crucial significance to invest in an adequate human-machine interaction is also emphasized by topical Government supported research and development projects since recently.

The fields of application for user-friendly interfaces are various: network systems in the office, expert work stations for teaching, training, production, quality control and maintenance, in mobile applications, in medical services, and, last but not least, in the home and entertainment sector.

The Institute for Human-Machine Communication is concerned with research and teaching in the above mentioned field. The present activity report gives a brief overview on our recent investigations and contributions to an user adequate human-machine communication for different areas of application. It follows our preceding activity report 1990-1996, and it supplements the current internet pages of our institute.

A consistent continuation and inclusion of usability engineering methods into our research topics, and the reinforcement in multimodal human-machine interaction combining tactile modi with natural language and speech, handwriting, vision, and gesture have been considerable challenges during the reporting period. We also continued our activities in technical acoustics, noise evaluation, and audiological acoustics. We adapted our usability laboratory, our anechoic chamber, and our reverberation chamber to the settings of new tasks. In order to guarantee reproducible investigations and interpretations of visual information in the context of multimodal human-machine dialogs we established a new computer vision laboratory. As future automotive systems are a most important area of application where very high requirements must be met, we installed a new navigation laboratory for usability studies in cars too.

We are very grateful and proud of experiencing remarkable response and approval to our research results from national and international conferences as well as from our co-operation partners in application oriented research and development laboratories. Details of our investigations are documented in a series of doctoral dissertations, in theses submitted for a diploma degree, and in a couple of student research projects. Persons, parties, and friends interested in more details of our teaching and research work are kindly invited to get in touch with us.

Manfred K. Lang

1 Staff

**Faculty Members
and Lecturers**

Manfred K. Lang, Univ.-Prof.
  Dr. rer. nat., head of the chair

Ernst Terhardt, Univ.-Prof. Dr.-Ing.,
  adjunct professor, retired 3/99

Hugo Fastl, Prof. Dr.-Ing.,
  academic director

Günther Ruske, Prof. Dr.-Ing.,
  academic director

Mathias Schneider-Hufschmidt, Dr.,
  lecturer (Siemens AG)

**Administrative
and Technical Staff**

Peter Brand, Dipl.-Inf. (FH),
  system administrator

Ernst Ertl, mechanic

Gertrud Günther,
  technical assistant

Heinrich Hundhammer,
  electronics technician

Claus von Rücker, Dr.-Ing.,
  software coordinator

Melitta Schubert, secretary

Scientific Staff

Frank Althoff, Dipl.-Inf.

Josef Chalupper, Dipl.-Ing.

Stephan Demmerer, Dipl.-Phys.

Robert Faltlhauser, Dipl.-Ing.

Michael Geiger, Dipl.-Ing.

Karla Geiss, Dipl.-Inf.

Marc Hofmann, Dipl.-Ing.

Jörg Hunsinger, Dipl.-Phys.

Gregor McGlaun, Dipl.-Math.

Fred Nentwich, Dipl.-Ing.

Ralf Nieschulz, Dipl.-Ing.

Christine Patsouras
  (formerly Huth), Dipl.-Ing.

Björn Schuller, Dipl.-Ing.

Bernhard Seeber, Dipl.-Ing.

Martin Zobl, Dipl.-Ing.

External Postgraduates

Sergey Astrov, Dipl.-Ing.
  (Siemens AG)

Josef Bauer, Dipl.-Ing.
  (Siemens AG)

Frank Forster, Dipl.-Inf.
  (Siemens AG)

Hans-Peter Grabsch, Dipl.-Ing.
  (Bosch)

Bernhard Niedermaier, Dipl.-Ing.
  (BMW AG)

Ronald Römer, Dipl.-Ing.
  (Infineon Technologies AG)

Wolfgang Schmid, Dipl.-Ing.
  (Gasteig GmbH)

Former Staff

Thomas Filippou, Dipl.-Ing.
  (until 12/99)

Gerd Gottschling, Dipl.-Ing.
  (until 6/97)

Frank Haferkorn, Dipl.-Phys.
  (until 3/98)

Dietmar Mass, Dipl.-Phys.
  (until 6/99)

Peter Morguet, Dr.-Ing. (until 1/00)

Johannes Müller, Dr.-Ing. (until 5/99)

Robert Neuss (formerly Grau),
  Dipl.-Ing. (until 9/00)

Thilo Pfau, Dr.-Ing. (until 10/00)

Christine Reischer, secretary
  (until 4/00)

Wolfgang Schmid, Dipl.-Ing.
  (until 4/98)

Holger Stahl, Dr.-Ing. (until 4/97)

Ingeborg Stemplinger, Dr.-Ing.
  (until 12/97)

Miriam Valenzuela, Dr.-Ing.
  (until 12/97)

Hans-Jürgen Winkler, Dr.-Ing.
  (until 1/97)

Lars Witta, Dipl.-Ing. (until 8/99)

Robert Zwickenpflug, Dr.-Ing.
  (until 7/97)

Former External Postgraduates

Udo Bub, Dr.-Ing.
  (Siemens AG)

Angela Engels (formerly Schreyer),
  Dr.-Ing. (Siemens AG)

Jochen Junkawitsch, Dr.-Ing.
  (Siemens AG)

Joachim Köhler, Dr.-Ing.
  (Siemens AG)

Christian Krapichler, Dr.-Ing.
  (GSF medis)

Henning Lenz, Dr.-Ing.
  (Siemens AG)

Helmut Spannheimer, Dr.-Ing.
  (BMW AG)

Christoph Wagner, Dr.-Ing.
  (Siemens AG)

2 Facilities

2.1 Usability Lab

The main goal of the research work at the institute is to improve human-machine interaction by developing new pattern recognition based input modalities and new intelligent dialog strategies entailing advanced interface concepts. As there is no experience of how people use and accept these new technologies, usability studies have to be made from an early stage of system development. Therefore the institute has recently installed a special laboratory for studying and evaluating new interaction methods and techniques.

In order to employ such investigations, it usually will be necessary to observe and record the behavior of experimental subjects. Thereby both, goal orientated actions for solving a problem and spontaneously occurring actions and reactions to stimuli brought in from the outside while handling technical systems, are of interest. For later analysis the actions of the experimental subject can be recorded by video and audio equipment and processed with computers.

The usability laboratory of our institute has a number of different purposes for our research:

- Studies on usability, user friendliness, and acceptance of user interfaces
- Gathering of practically relevant training data for new interface concepts, pattern recognition based input modalities and dialog strategies
- Gathering of empirical data for developing and optimizing interface concepts and dialog strategies also considering ergonomic viewpoints

The usability lab consists of three separate rooms (see fig. 1, left):

Observation room: It is equipped with several remote controlled video cameras and microphones for observation and recording, and with loudspeakers to provide any acoustical environment. In that room the subject is supposed to deal with the experimental setup.

Control room: It is provided with video and audio mixers, computers and various studio equipment (see fig. 1, right). The subject is additionally monitored through an one way mirror. From that room a scientist controls the whole experiment.

Recording room: A so-called sound proved booth is used for special examinations which require audio signals with a high signal-to-noise ratio.

Marc Hofmann, Ralf Nieschulz

Fig. 1: Outline (left) and control room (right) of the usability lab.
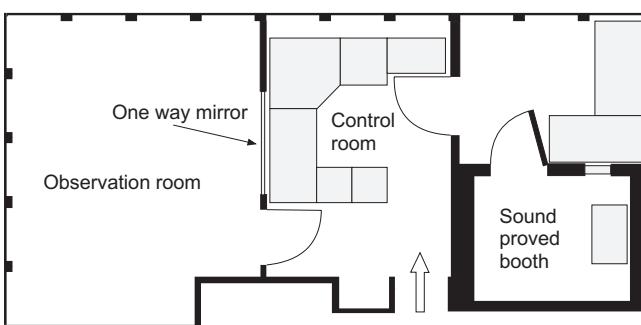
2.2 Navigation Lab for Usability Studies in Cars

This usability lab is used to test and optimize new approaches for multi-modal man-machine interfaces (MMI) in the environment of a car. It is called "navigation lab" since its primary use was to investigate the MMI of automotive navigation systems.

Because of large environmental influences on the test persons, MMIs have to be tested in a real environment. It obviously makes little sense to evaluate the usability of a car mobile phone – which actually has to be operated while driving – in a desktop scenario. Unfortunately, it is difficult to integrate the required technology into a car boot. Some test scenarios might even lead to dangerous situations in real traffic. To make a replica of real conditions, the navigation lab was built up. However, real conditions do not only have effect on the test-person, but also on the used technology.

Image processing suffers under changing illumination and background setups – especially in the car environment conditions for this are really rough. Speech recognition is affected by driving noise or additional passengers in the car. Drivers may switch, but the system still has to function correctly. These conditions call for robust algorithms and statistical methods for user and situational modelling. Naturally, operating the MMI is not the primary task of the driver. Thus a number of precautions (e.g., adaptation to user and situation, structure of dialog, guidance to the user) to support or even assist the driver have to be taken.

This can all be simulated and tested in the navigation lab.

Surely, not all the technologies have to be implemented before testing the usability of an MMI. We want to know exactly how the whole system has to function before wasting time inventing technologies not satisfying the user's needs.

Real car environment is simulated in a car with force-feedback steering, automatic transmission, hand brake, pedals, knobs, car-audio etc. (see fig. 2, right).

A lcd-monitor with a haptic input device displays the user interface. Several cameras and microphones are installed inside the car in order to observe the test person's activities throughout a usability test. The driving simulation is projected (5 m projection-diagonal) by a lcd-beamer onto the wall in front of the car (see fig. 3, right). Four speakers assure that acoustical events can be positioned in every direction.

The simulator is surrounded by molton walls to separate the proband from the lab environment.

A control room (see fig. 3, left) is also
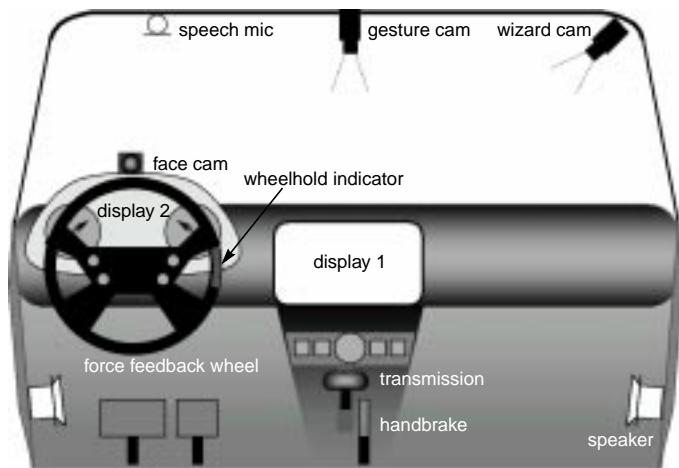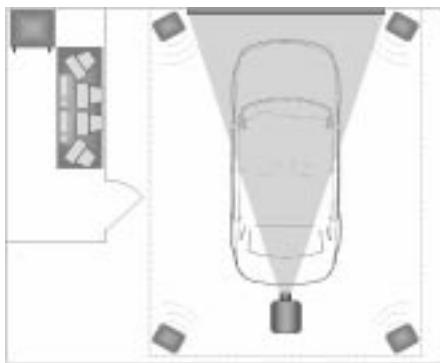
Fig. 2: Left: Outline of the navigation lab. Right: Car equipment.

Fig. 3: Navigation lab. Left: Control room. Right: Car and projection from behind.

located in the navigation lab, which can be fully separated (acoustically and visually) from the test-scenario. It is equipped with several monitors, video recorders, a video mixer, a scan-converter, audio compressors, audio mixer, audio amplifiers for car audio, monitoring and driving simulation, a MMI computer (hi-speed PC with two monitors) and the driving-sim computer (hi-speed PC with fast graphic engine). This equipment allows a maximum of flexibility for all conceivable setups and scenarios.

More computers are scheduled, e.g. for gesture recognition, adaptive components, emotion estimation, multimodal dialog concepts or speech recognition (cf. sections 4.2-4.5.3). They can easily be integrated into the setup. The navigation lab is connected to the data processing equipment of the institute through 100 Mb/s LAN, which allows the use of practically any desired computing performance.

Michael Geiger, Martin Zobl

2.3   Computer Vision Lab

The computer vision lab was established at the institute of Human-Machine Communication in the beginning of the year 2000. In this lab, both analytical investigations concerning digital image processing and image synthesis are taken up.

Currently, the technical equipment of the lab consists of two Silicon Graphics Workstations (Indigo2 Impact and Indy with special grabbing and video compression devices), three high-performance PCs (with standard frame grabber cards) and several high-quality video cameras. Moreover the laboratory is equipped with a video cassette recorder, a studio TV monitor and a sound audio-mixing device to do basic usability studies. To avoid interfering effects on picture sampling a flicker-free high-frequency lighting was installed.

On the image processing side, usually the following steps are carried out to digitize images. The scene is recorded by one or more cameras. The analog signal of the camera(s) is sent to the A/D-converter of the computer, which is called frame grabber card. The sequence of digitized pictures, i.e. the video frames, can be viewed on the monitor and optionally, single frames or a sequence of frames can be saved in a computer-compatible file format (e.g. common image formats like Jpeg, PNG, or video formats like AVI or MPEG). For further evaluation the frames can be improved by image preprocessing (e.g. segmentation or texture analysis). By overlaying a rectangular pattern on the image it is subdivided into single sections. This process is called screening. From these frames, some special (densitorical, geometrical etc.) features can be extracted. By applying special determination algorithms the single features can be classified. Motions can be detected by comparing the frames of a sequence. Thus a set of translation vectors can be defined and rated.

On the image synthesis side, some research is done concerning virtual 3D-scenarios that are modelled using the virtual reality modelling language (VRML). One of the current endeavors in this field is to build a prototypical virtual model of several facilities of the institute. In this context, a system is generated and tested which enables the user to navigate through the virtual room with the help of natural speech utterances. Moreover, with regard to a multi-modal handling, both usual haptic interfaces (keyboard and mouse) and a dynamic gesture-recognition module have been integrated.

Gregor McGlaun, Frank Althoff

Fig. 4: Left: Experimental setup for gesture recognition techniques. Right: VRML model of the lab.

2.4 Anechoic Chamber and Reverberation Chamber

Many physical-technical, electroacoustical, and psychoacoustical measurements rely on environments with defined room-acoustical parameters. Thus an anechoic chamber and a reverberation chamber, both available at the Institute for Human-Machine Communication, establish the standard equipment of acoustical measurement instrumentation.

Although both rooms define opposite acoustic environments, they both need to fulfill the following two requirements: noise, even infrasonic, must have a very low level in the chambers and noise transmitted from the chambers to the control room must not annoy or even endanger the operators. A combination of room and structural acoustics measures as the costly room-in-room construction ensures the fulfillment of these requirements.

The anechoic chamber should support the generation of a free, undisturbed sound field. The sound intensity reflected from the walls should therefore be minimal. For this reason all walls, the ceiling, and the floor are covered with an absorbing material. Nested side-by-side arranged wedges of mineral fibres achieve a high noise absorption coefficient over a wide frequency range through a continuous transition of the sound wave from the air to the absorber. The lower cutoff frequency of the anechoic chamber is defined as the lowest frequency at which the noise absorption coefficient under normal angle of incidence is at least 0.99. For the arrangement of the absorption wedges used in this anechoic chamber the lower cutoff frequency of the room is about 125 Hz and its wavelength corresponds to four-times the length of the absorption wedges. The usable volume of the room is $L*W*H = 7.5\,m * 4.2\,m * 2.8\,m = 88.2\,m^3$. Extensive measurements were carried out to optimize speaker- and microphone positions

Fig. 5: Acoustical measurements are carried out in the anechoic chamber.

against the background of standing waves at frequencies below the cutoff frequency of the anechoic chamber.

Sound fields in the reverberation chamber should be statistical, i.e. the temporal mean of the sound intensity should be equal for all room directions at all places in the room and for all measurement frequencies. Therefore the walls are built non-parallel and all surfaces are covered with sound-reflective material. Furthermore perspex reflectors are installed as sound-diffusors. The reverberation time of the room at low frequencies is longer than 10 s and can be reduced for experiments by mountable plate absorbers. Mounting bars allow the fast attachment of different materials whose acoustical properties are to be investigated. The size of the reverberation chamber is $L*W*H = 5.5\,m * 4.9\,m * 3.9\,m = 106\,m^3$.

In a control room adjacent to both chambers the experimenter finds systems to generate, analyze, control and document measurements and experiments. Besides different electroacoustical transducers, systems for generating and analyzing measurement signals and procedures are available.

Bernhard Seeber

3   Lectures and Seminars

Lectures	Lecturer	Credit hrs	Subject (in English)
Signaldarstellung (undergraduate)	Lang	4	Signal Representations
Mensch-Maschine-Kommunikation 1	Lang	3	Human-Machine Communication 1
Mensch-Maschine-Kommunikation 2	Lang	3	Human-Machine Communication 2
Audiokommunikation	Fastl	3	Audio Communication
Technische Akustik und Lärmbekämpfung	Fastl	2	Technical Acoustics and Noise Abatement
Musikalische Akustik	Fastl	2	Musical Acoustics
Automatische Mustererkennung in der Sprachverarbeitung	Ruske	2	Automatic Pattern Recognition in Speech Processing
Datenanalyse und Informationsreduktion	Ruske	2	Data Analysis and Information Reduction
Digitale Verarbeitung von Sprachsignalen	Ruske	2	Digital Processing of Speech Signals
Gestaltung ergonomischer Benutzungsoberflächen	Schneider-Hufschmidt	2	Design of Ergonomic User Interfaces

Laboratories	Lecturer	Credit hrs	Subject (in English)
Praktikum Mensch-Maschine-Kommunikation	Lang	4	Laboratory "Human-Machine Communication"
Praktikum "Praxis der Mensch-Maschine-Kommunikation"	Lang, Fastl, Ruske	4	Laboratory "Practical Aspects of Human-Machine Communication"
Praktikum System- und Schaltungstechnik 1, 2	Lang, Ruske, and others	2	Laboratory "Systems and Circuitry"

Regular Seminars	Lecturer	Subject (in English)
Hauptseminar "Aktuelle Fragen der Mensch-Maschine-Kommunikation"	Lang, Fastl, Ruske	Advanced Seminar "Current Topics in Human-Machine Communication"
Oberseminar Mensch-Maschine-Kommunikation	Lang, Fastl, Ruske	Seminar Human-Machine Communication
Oberseminar "Laufende Arbeiten zur Mensch-Maschine-Kommunikation"	Lang, Fastl, Ruske	Seminar "Ongoing Work in Human-Machine Communication"
Oberseminar "Interdisziplinäre Grundlagen der Mensch-Maschine-Kommunikation"	Lang, Fastl, Ruske, and others	Seminar "Interdisciplinary Foundations of Human-Machine Communication"
Kolloquium Informationstechnik	Lang and others	Seminar Information Technology
Seminar "Umweltaspekte der Verkehrstechnik"	Fastl	Seminar "Environmental Aspects of Transportation Technology"

Teaching Scripts

[U-F1]   Fastl, H.: Umdrucke zur Vorlesung Technische Akustik und Lärmbekämpfung (1999).
[U-F2]   Fastl, H.: Umdrucke zur Vorlesung Musikalische Akustik (2000).
[U-F4]   Fastl, H.: Umdrucke zur Vorlesung Audiokommunikation (2000).
[U-L1]   Lang, M.: Skript zur Vorlesung Signaldarstellung (1994).
[U-L2]   Lang, M.; Stahl, H.; Winkler, H.-J.: Übungskatalog zur Signaldarstellung (1994).
[U-L3]   Lang, M.: Skript zur Vorlesung Mensch-Maschine-Kommunikation 1 (1999).
[U-L4]   Lang, M.; Müller, J.; Winkler, H.-J.: Übungs- und Prüfungsaufgaben Mensch-Maschine-Kommunikation 1 (1994).
[U-L5]   Lang, M.; Morguet, P.: Versuchsmanuskript zum Praktikum Mensch-Maschine-Kommunikation (1995).
[U-L6]   Lang, M.: Kurzmanuskript zur Vorlesung Mensch-Maschine-Kommunikation 2 (1996).
[U-L7]   Lang, M., Althoff, F.; Geiger, M.: Ergänzungen und Übungen zur Vorlesung Mensch-Maschine-Kommunikation (1999).
[U-L8]   Lang, M.: Ringpraktikum Elektrotechnik und Informationstechnik (1996).
[U-R1]   Ruske, G.: Skriptum zur Vorlesung Datenanalyse und Informationsreduktion (1999).
[U-R2]   Ruske, G.: Skriptum zur Vorlesung Automatische Mustererkennung in der Sprachverarbeitung (2000).
[U-R3]   Ruske, G.: Skriptum zur Vorlesung Digitale Verarbeitung von Sprachsignalen (2000).

4 Research Topics

Modern systems for communication and information processing are prerequisites for a high quality interpersonal communication and information exchange. Moreover, those systems enable us to interact with all kinds of computers and computer controlled machines, e.g. to operate entertainment electronics, to access the internet, to use information services, or even to navigate a car. With ongoing technological progress, these systems do not only become more capable and efficient, but also more complex. Nowadays, these systems have become a part of everyday life, in contrast to former times, when only engineers and experts had to operate them. For this reason, an adequate and efficient user interface is a major goal of research and development to enable everyone to participate in modern communication infrastructure and technology.

The research topics at the Institute for Human-Machine Communication deal with the fundamentals of a widely intuitive, natural, and therefore multimodal interaction between humans and complex information processing systems. All forms of interaction, i.e. modalities, that are available to humans, are to be investigated for this purpose. Both the machine's representation of information and the interaction technique is to be considered in this context, like

- text and speech,
- sound and music,
- haptics,
- graphics and vision,
- gesture and mimics,
- and emotions.

The improvement of a single method of interaction is important, but not our main goal of research. The *coaction* of different modalities is most promising to enhance the efficiency of Human-Machine Communication. Even a combination of only two modalities, e.g. speech and haptics, can be more efficient and less error-prone than a single way of interaction.

To ensure the end-user's acceptance of those new ways of Human-Machine Communication, it is of prime interest to investigate the usability of new interfaces in an early stage of development. This usability engineering process yields numerous hints to enable developers to produce a more efficient and less cryptic dialog between humans and machines.

Since the user's knowledge about a system and its properties changes over time and differs for each individual, it is advisable to create adaptive user interfaces. Therefore, our research also investigates the foundations of adaptivity and learning systems to enable us to develop man-machine interfaces that really take the user's variable skills into account.

Human-Machine Communication is an interdisciplinary field of research. Therefore many different subjects are involved to reach the long-term research objective of a natural, intuitive way of interaction with "machines". This chapter gives an overview of the current research topics at the Institute for Human-Machine Communication, but cannot be complete. Please see over the list of scientific publications in section 8.2 for an exhaustive view of our research work in the four-year period of this report.

Claus von Rücker

4.1 Usability Engineering and Multimodal Human-Machine Communication

4.1.1 System Architecture for Multimodal User Interfaces

Human beings process several interfering perceptions at a high level of abstraction so that they can meet the demands of the prevailing situation. Most of today's technical systems are incapable of emulating this ability yet. Another problem of current systems is that due to their growing functionalism their interfaces are often complex to handle and require adaptation by the user to a high degree.

However, those interfaces would be particularly desirable whose handling can be learned in a short time and that can be worked with quickly, easily and, above all, intuitively. Therefore we propagate the design of multimodal system interfaces as they provide the user with greater naturalness, expressive power and flexibility. Moreover multimodal operating systems probably function more robustly than their unimodal counterparts because they integrate redundant information shared between the individual input modalities.

We are specially interested in the design of a generic multimodal system architecture which can easily be adopted to various conditions and applications and thus serve as a basis for multimodal interaction systems. In contrast to most of the existing architectures our design philosophy is to merge several competing modalities (i.e. two or more speech modules, dynamic and static gesture modules, etc.) instead of using a specially designed set of combined modalities.

To integrate the various information contents of the individual input modalities we are following a new approach. By employing biological motivated evolutionary strategies the core algorithm compromises a population of individual solutions to the problem at hand and a set of operators defined over the population itself. According to evolutionary theories, only the most suited elements in a population are likely to survive and generate offspring, transmitting their biological heredity to new generations and thus lead to stable and robust solutions.

The concepts of the developed multimodal system architecture are validated in various scenarios, both research and industry-sponsored projects. One special application for example enables the user to navigate in arbitrary virtual worlds by freely combining natural and command speech, dynamic hand gestures and conventional graphical interfaces

Frank Althoff, Gregor McGlaun

4.1.2   Usability Engineering for Adaptive Dialog Procedures in the Automobile (ADVIA Project)

Due to the multiplicity of coexisting electronic devices in modern luxury and upper class automobiles, the limit of usability has been reached for the standard user. Some examples of these devices are navigation and telematics systems, audio and video components like cd-changer, radio and television, mobile phone, car computer, air condition and any additional conceivable unit such as internet applications. As the devices of different manufacturers not only look different, but also have company-specific user interfaces and control elements, these can hardly be operated by a standard user – particularly while driving.

For some years now, car manufacturers have been trying to solve this problem by developing multi-functional integrated man-machine interfaces (MMI), which employ a common graphical display and a reduced number of haptic control elements.

Reducing control elements for example increases the complexity of menu structures. The user still has to read lengthy manuals to find his way through the menus of the resulting user-interface. In order to make operating the MMI more intuitive, alternative basic approaches and studies are essential.

We consider the introduction of familiar human communication modalities like natural speech and gestures a precondition for an intuitive dialog with machines, which also applies to car MMIs. Additionally, the user should get further support by an adaptive, help (cf. section 4.3.2) and assistance system (cf. section 4.3.1). However, further problems appear.

How does such an integrated user interface have to be structured? What kind of modalities can or have to be used for which function? How and when should acoustical or visual feedback be given? How and when can the system sensibly adapt to the user? How can the user interface be designed user-friendly, intuitively usable, assisting (but not dominating) the user, all without drawing the drivers attention away from traffic?

Our research in this domain mainly takes place in our navigation-lab (see 2.2), a specially automotive-fitted usability-lab (see 2.1) which permits us to do usability tests in an automotive environment. The MMI is simulated by a computer. With the "Wizard-of-Oz methodology" we test new concepts in a kind of rapid prototyping, with a "wizard" observing the test person and controlling the MMI. The test person gets the impression of controlling the MMI with gestures, speech or haptics. Promising concepts are then transferred and implemented into reality.

This way we have developed, tested and implemented haptic, gesture, speech and multimodal controlled MMIs with adaptive, incremental help systems and adaptive assistance systems.

Michael Geiger, Martin Zobl

4.1.3   A Multimodal Mathematical Formula Editor for Audio-Visual Natural Interaction

Introduction

Electronic acquisition of mathematical formulas via conventional tools is a time consuming and complicated task. Therefore, a soft-decision solution for online handwritten formula recognition was successfully demonstrated in a former project [97win2]. Recently we developed a novel approach towards a multimodal analysis of natural, especially speech and handwriting interaction, both being the fastest and most intuitive channels for entering mathematical expressions into a computer [00hun1]. We utilize an integrated, multilevel probabilistic architecture with a joint semantic and two distinct syntactic models describing speech and script properties, respectively. Basic arithmetic operations, roots, indexed sums, integrals, trigonometric functions, logarithms, convolutions, fourier transforms, exponentiations, and indexing (among others) are supported. Compared to classical multistage solutions our single-stage strategy benefits from an implicit transfer of higher level contextual information into the lower level segmentation and pattern recognition processes involved. For visualization and postprocessing purposes, a transformation into Adobe™ FrameMaker™ documents is performed.

Methods

The syntactic-semantic attributes of spoken and handwritten mathematical formulas are represented by the parameters of a so-called Multimodal Probabilistic Grammar. It combines properties of context-free phrase structure grammars with those of graph grammars by allowing for word-type, symbol-type, and position-type terminals.

The grammar is implemented into a single stage semantic decoder by means of a compact semantic representation called Semantic Structure S [99mue1]. It is given by a hierarchically structured combination out of a predefined inventory of Semuns s (semantic units) with corresponding types, values, and successor attributes, every unit referring to a certain mathematical operator or operand. On the syntactic level every Semun of a given semantic hypothesis is assigned to a so-called Syntactic Module (SM). It consists of an advanced transition network which enables two distinct stochastic processes: 1) transitions from one node to another, 2) emissions of spoken words, handwritten symbols, or local offsets between associated symbols or symbol groups [00hun2]. Transitions are responsible for modelling speaking and writing order, whereas emissions account for varying word, symbol, or position choice, respectively. All the necessary transition and emission probabilities as well as semantic type, value, and successor probabilities were estimated from training corpora obtained from separate speech and handwriting usability tests.

An extended Earley-type top-down chart parser performs a MAP (maximum a-posteriori) classification across all abstraction levels. On the signal near levels, the preprocessed speech or handwriting input sequences are rated using 30-dimensional phoneme based semi-continuous HMMs (Hidden Markov Models) or 7-dimensional DTW (Dynamic Time Warping) matching, respectively. In a one pass search algorithm all possible semantic hypotheses are successively tested until the best overall semantic representation of a given input is found. Due to a breadth-first search strategy, inline first-last processing is enabled.

The most significant advantage of a single stage semantic decoding architecture as it is used in this work results from the simultaneous evaluation of knowledge belonging to all the involved abstraction levels: Apart from self-focussing effects achieved by restricting the search process to locally consistent sub hypotheses only, semantically corrupt recognition results are prohibited due to the integrated expectation driven classification scheme.

The overall system architecture is sketched in fig. 6.

Results and Conclusions

For evaluation purposes we performed independent test classifications in either modality. Fully spoken or handwritten realistic formulas were examined, yielding a structural recognition accuracy of 61.1 % for
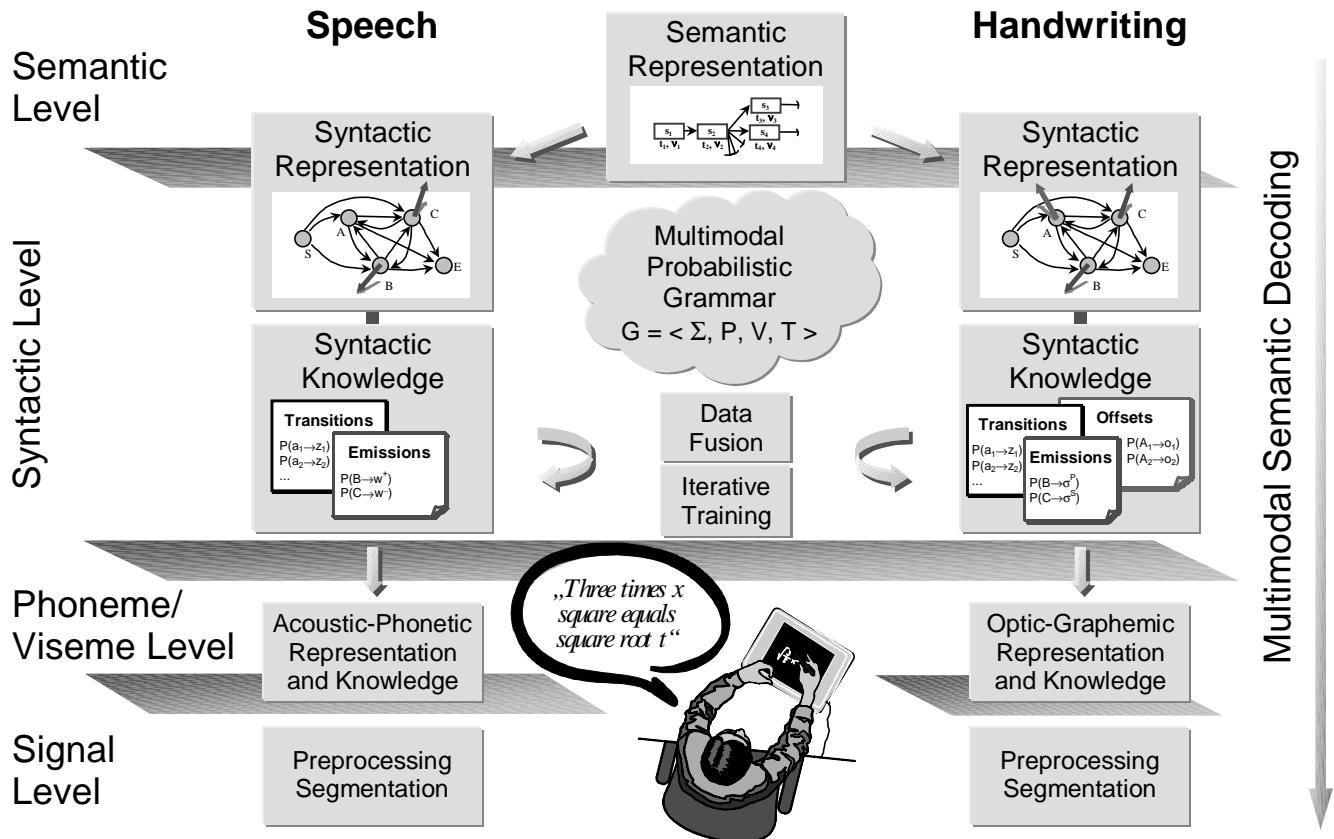
Fig. 6: System overview of a multimodal mathematical formula editor.

speech and 83.3 % for handwriting (note: these numbers refer to complete formula correctness). For the future we wish to support freely interfering speech and handwriting interactions including mutual coreferencing due to deictic wording and pen gesturing. To this end, the use of speech will presumably be focussed to subterm input and error corrections so that we anticipate a robust and approximately real-time forthcoming system performance. Via a back and forth transformation to FrameMaker™'s formula editor natural speech and handwriting interaction may also be complemented by conventional input modes in the future.

Jörg Hunsinger

4.1.4 FERMUS (Fehlerrobuste Multimodale Sprachdialoge – Error Robust Multimodal Speech Dialogs)

The FERMUS project was started in March 2000 in cooperation with the industry partners BMW AG, DaimlerChrysler AG, Siemens AG and Mannesmann VDO AG, with the primary intention of localizing and evaluating various strategies to analyze errors in information systems by using various modalities which are mainly recognition-based.

In this regard, an effective error-management module is considered to play a very important role. Ad hoc, three potential approaches can be identified:

• the extensive use of multimodal information sources
• generation of new pieces of information by using adaptive dialog structures
• context-sensitive interpretation of the available information.

A special goal is to investigate how the robustness of technical systems can be enhanced by using multimodal information already on the recognition level. By developing determined dialogue-techniques and special strategies of system adaptation to the current situation and to the intention of the user we expect an additional increase of system functionality because of a dynamic specification. Moreover, the influence of emotions, especially with regard to stress-situations, is intensively researched to enable a reliable separation and – in some cases – transformation of unusable to usable information.

The primary test domain is the handling of diverse communication facilities in an upper-class automobile (such as radio/cd, telephone, inter-net etc.) in connection with the variety of potential error sources by internal and external troublesome side-effects. In this context we are specially interested in the modality specific impacts on the performance of the overall system. For the examinations a simplified man-machine interface which facilitates the operation of basic information- and communication-devices is used.

The project bases on information of already completed projects with various industry partners (mainly the BMW AG and the Siemens AG), particularly this holds for the examination results concerning adaption mechanisms as well as usability-tests with regard to multimodal communication.

Frank Althoff, Gregor McGlaun

4.2   Visual Human-Machine Interaction Utilizing Image Interpretation

Introduction

Gestures and facial expressions are important components of interpersonal communication. Using these visual modalities, human-machine dialog too can be provided with a more natural and intuitive form. Here we will exclusively consider vision-based methods in order to achieve non-intrusive gesture and facial expression recognition. The major topics in this field of research are specification, implementation and evaluation of the human-machine interface and development of image-

based methods for automatic gesture recognition. It has turned out that the user's gestures can be massively influenced by the graphical design of the visual interface, therefore it is very important to coordinate the menu-driven handling and the visual feedback in order to create a gesture optimized application. During the research and development process frequent usability-tests have been carried out, for which we principally use the "Wizard-of-Oz methodology" (c.f. section 4.1.2). Most human gestures are move-

ments; therefore, a lot of information is transferred by motion.

System Overview

The process of automatic gesture recognition is described in the following, fig. 7 shows the system overview. The whole system does not need any model of the target object. Thus the system can be easily transferred to other objects or domains. The object in question (e.g. hand or face) is separated from the background by spatial segmentation. This is done by color-segmentation
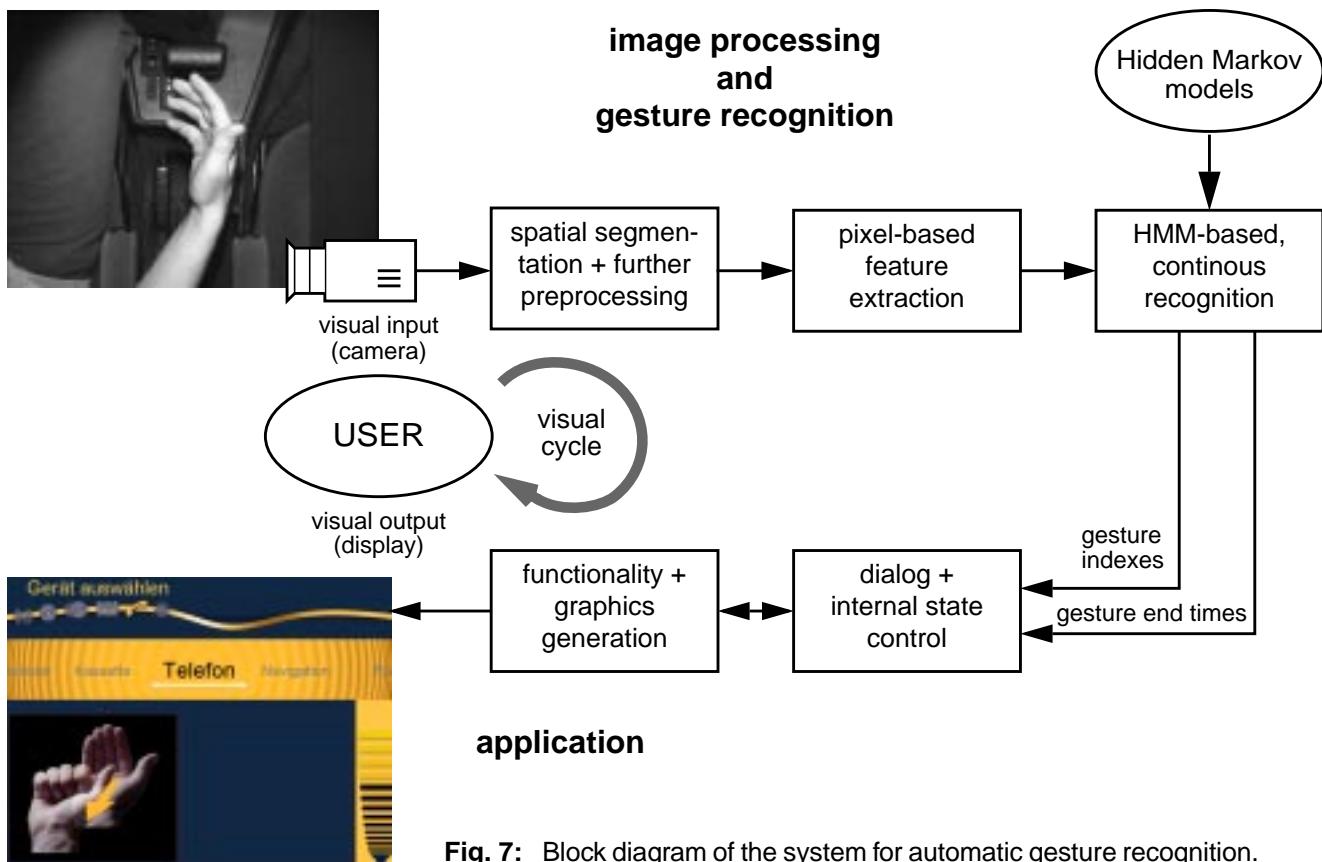

Fig. 7: Block diagram of the system for automatic gesture recognition.
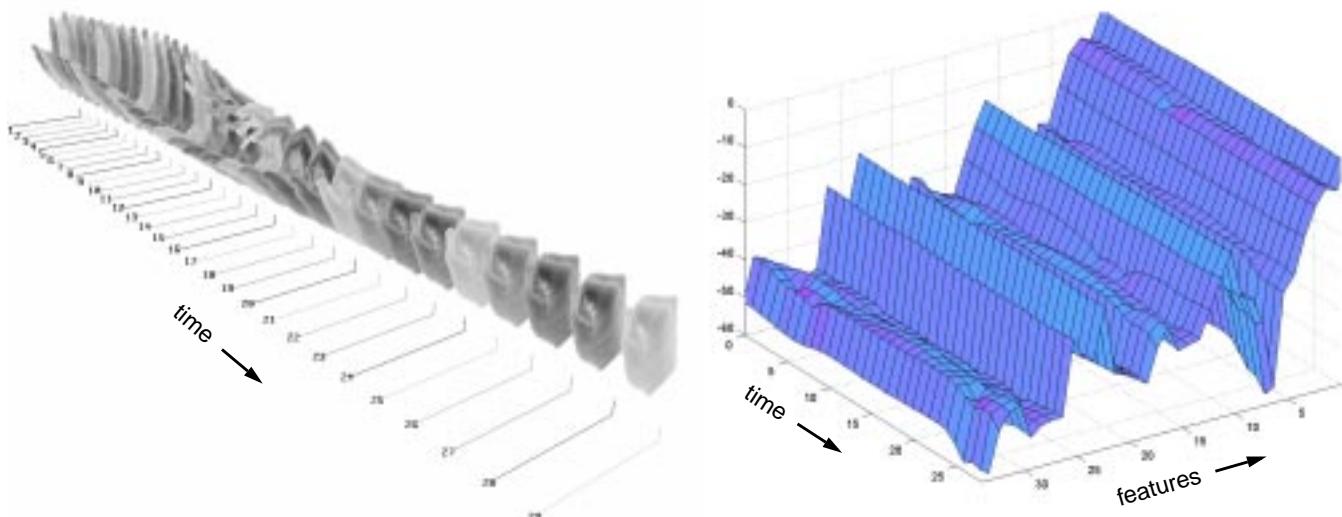
Fig. 8: Temporal image sequence of grabbed gesture (left) and corresponding set of features (right).

in an uncluttered environment with defined lighting and a combination of low-level image-processing algorithms with object-tracking in cluttered scenes. The movement of the segmented object is then classified with stochastic models. Hidden Markov Models (HMMs) are used for this, which can reproduce non stationary temporal processes. The central problem with this model is to find suitable features for transforming the spatio-temporal image sequence (fig. 8, left) into a time sequence of feature vectors (fig. 8, right). Different features, extracted from object-area or object-contour, have been implemented and tested. A further problem is to separate the continuous video stream of the observation camera into meaningful sections and non-meaningful sections, like coincidental movements or pauses. Temporal segmentation is done with a fast, however little robust, two-level approach and a robust, single-level approach with large computational cost (HMM-based spotting). The recognition

methods developed in this project allow – depending on the used features – segmentation and classification for each type of movement. Thereby feature extraction methods have been found, which are suitable for recognizing dynamic human gestures and mimic. A real-time demonstration system has been implemented, the complex functionality of which can be controlled exclusively with dynamic gestures [98mor1, 98mor2, 98mor3, 99mor1, P-L13].

Applications

Gesture recognition has been broadened to other domains like controlling electronic car devices (see ADVIA project, section 4.1.2) and navigation in VRML-worlds (see section 4.1.1) with specifically adapted, usability-tested visual interfaces and gesture vocabulary. In order to permit a more intuitive and human fitted handling of an application, the gesture recognition system has to be modified and improved. Currently, a gesture is defined as a hand movement with a defined start

and end position. The response of an application takes place after the recognition of such a completed sequence. Analyzing everyday situations show that, in most cases this form of indirect manipulation is used with gestures. But in some cases, it is reasonable to analyze and process the visual input directly while detecting a directional hand movement, for example setting analogue quantities or moving objects to a precise point on the screen. The main problem here is to make the recognition system distinguish automatically between direct and indirect control modes. Further information sources in gestures are speed, amplitude and repetition frequency, which are not especially analyzed so far by the system. In addition to carrying important information about the amount the user wants to change a parameter by, these features carry information about the user's habits and emotional situation.

Martin Zobl, Michael Geiger

4.3 Adaptation and Learning Strategies

4.3.1 Probabilistic Modelling of the User Intention for Adaptive Human-Machine Interaction

The information sources which need to be exploited for user adaption strongly depend on the goal of adaption and the application itself. Tutoring systems for example provide some kind of guidance or training on a special topic, the user's goal is obvious as he uses the tutoring system. Other applications allow a variety of different intentions and the complexity of these applications doesn't allow direct inference of the user intention. However, knowing the user intention is the key for an appropriate adaptive dialog and system modelling. To calculate an estimation of the user's goal, we developed two different approaches, both based on probabilistic networks, which provide methods of dealing with uncertain and incomplete information.

A plan recognizer is used to infer the user intention by considering recent user actions. Depending on the application, the user's plans differ often very much from optimal plans, i.e. users behave suboptimal. Our plan-based approach has been optimized for such applications, considering that even suboptimal behavior may support plan/intentions-hypotheses to a certain extent. Trying to determine the user intention merely from a small number of actions obviously entails a risk to fail, i.e. not to estimate the real user intention. To reduce the risk of the system not to offer help or assistance for the real user intention, a help system was developed that is capable of creating help texts according to a number of nearly equally likely plans without overstraining the user's cognitive capabilities.

The second approach has been developed for scenarios/applications with varying situations entailing changes of the user's preferences and intentions. In a car the driver's intentions and preferences strongly depend on location, weather, speed, traffic and inmates, for example. Therefore a number of different probabilistic expert systems based on probabilistic networks have been created which are capable of learning typical user behavior and intentions according to the current situation. Real-time online training of the influence of each situation parameter on the user behavior allows the expert system to infer the user intention in similar situations. As a result, reasoning about a certain situation doesn't require the expert system to be trained to that exact situation.

Having an estimation of the user intention offers a lot of potential for enhancing the human-machine dialog. For example knowing the user intention helps to reduce system requests and to formulate system requests in regard of the situation and the user's goal. Additionally we use the estimation of the user intention not only for user adaption, but also for a user and situation specific evaluation of pattern recognition-based inputs to improve recognition rates.

Marc Hofmann

4.3.2 Neuro-Fuzzy based Architecture for Help and Tutoring Systems

In order to facilitate the interaction with modern software systems, an adaptation to the user is to be achieved. Therefore the system must be capable of gaining and valuing information about the user. Fundamental methods and approaches to the adaptation of a software system to a user can be developed and demonstrated best within the area of „intelligent help and tutoring systems". Central viewpoint is the modelling of the knowledge status of the user. We examine both statistical and rule-based methods for learning user models, which are to enable an independent, user-adequate adaptation of the help or tutoring system. Essentially the architecture underlying the system consists of a combination of neural networks and fuzzy logic, one speaks also from „neuro-fuzzy and soft computing".

In the first phase of the system creation (training) we use the information won from user actions and afflicted with a factor of uncertainty (preprocessing) in order to generate and train an easily modified neural network. Depending upon application a generalized regression network, a probabilistic neural network, a competitive neural network or a self organizing map is taken as a basis. In order to model different characteristics of the users optimally, also a combination out of several networks can be used. In the phase of use it is now possible with this network to classify a user in different categories to give him appropriate assistance embedded in the context. With the user actions evaluated in this phase we re-train the network online. Thus an by and by increasing adaptation of the system to the individual user is achieved.

We develop and test the fundamental architecture in different projects. As a part of the ADVIA project (see 4.1.2) an adaptive help system is created as supporting introduction to the gesture controlled operation of an MMI in the vehicle. In order to infer the neediness and the knowledge status of the driver, in particular the execution quality of the gestures (confidence measure of the gesture recognition), the execution duration of the gestures, as well as the number of assistance request, in each case as a function of the context are analyzed here. Our goal is to support the user with an automatic, but unobtrusive help to avoid or minimize the need of help requests. Beyond that an expansion of the help system is conceivable on the entire, multimodal operation of the MMI. Furthermore we develop a tutoring system for an introduction in creating HTML-pages.

Ralf Nieschulz

4.4 Emotional Aspects in Human-Machine Communication

Nowadays, it is well known that emotions play a fundamental role in perception, attention, reasoning, learning, memory, decision making and other human abilities and mechanisms we generally associate with rational and intelligent behavior. In addition to verbal communication, nonverbal communication represents an important factor to enable a reasonable, purposeful and efficient interaction between people. Beside these communicative functions, nonverbal behavior and especially facial expression can provide information about the current affective state of a person. Correspondingly, we can't afford to neglect emotions anymore, if we want to develop future computer systems capable of solving complex problems or to interact with humans in an intelligent way, that means to create human-like behavior in machines.

Numerous questions to be dealt with are raised in the context of emotions in human-machine communication: which emotions occur when humans interact with computer systems, how can they be recognized, which stimuli cause these emotions, in which way should a computer system respond to emotional reactions of the user, etc. Once these problems get solved, the computer system capable of emotions would register the signals sent out by the user, recognize the inherent patterns and assimilate this data in a model of the user's emotional reactions. Then the system would be able to transmit useful information about the user, for example to applications that can use such data.

The probably largest application area might be future generations of human-machine interfaces, which will be able to recognize the emotional states of the users and react to them adequately. For example, if an user becomes frustrated or annoyed while interacting with an application, the system might respond to these emotional states of the user, preferably in such a way, that the user would sense it as intuitive. This way, following the interaction between humans, it might be possible to achieve a more natural human-machine interaction.

By integrating emotions as an additional modality to haptics, speech, gesture, etc. it might be possible to improve the error robustness of the control of technical systems. Within the framework of the FERMUS project (as described in section 4.1.4) research on this topic will be done. Thereby, the operation of communication devices in automobiles like phone, radio/cd, internet will serve as application area. Emotional states of the user will be used as a supplementary control element to achieve a better adaptation of the system to the current situation of the user. Within the studies it is of particular interest to detect in which case system errors lead to negative emotional reactions of the user and how it is possible to avoid them or at least to weaken them.

Karla Geiss

4.5 Automatic Speech Recognition and Speech Understanding

In many applications of modern human-machine interaction it seems to be necessary to allow continuous speech as input and output of the system. The typical task of continuous speech recognition consists in evaluating the phonetic information of an utterance (e.g. a sentence) and to represent the result as a chain of words which are defined in a lexicon. This task usually is carried out on the basis of single procedures or "modules" which perform the preprocessing of the acoustic speech signal, the application of phonetic units (phoneme models), classification of words, utilizing syntactic constraints, recognizing the complete sentence, and analyzing the semantic meaning of the spoken utterance with respect to a given, well-defined task.

The research tasks carried out at the Institute for Human-Machine Communication are:

- Preprocessing of the speech signal
- Analysis and extraction of acoustic-phonetic features
- Defining proper acoustic-phonetic decision units
- Modeling of acoustic-phonetic units by Hidden Markov Models (HMM)
- Application of Neural Nets (NN)
- Pronunciation rules
- Stochastic language models
- Algorithms for fast search
- Training procedures
- Adaptation methods (e.g. speaker adaptation, channel normalization)
- Stochastic methods for speech understanding
- Development of complete systems

Since speech sounds are characterized especially by their spectral properties, a preprocessing step calculates short-time spectra in time intervals of about 10 ms. Beyond that, it is advantageous to take into account the properties of the auditory system. For this purpose the Bark- or mel-scale should be chosen instead of a linear frequency axis. A basic model of human auditory frequency analysis can provide so-called "loudness spectra" which are especially suited for speech recognition.

A fundamental problem are the coarticulation effects which cause a strong influence between neighboring speech sounds. As a result, the feature vectors can be strongly dependent of the phonetic context, so that a phoneme unit cannot be described by a single pattern. In our work we tried to use syllables or parts of syllables as decision units, these are syllabic consonant clusters and syllabic nuclei (vowels and vowel clusters). By use of these clusters the main coarticulation effects are contained within the units. As an alternative, so-called triphones can be introduced which constitute each phoneme together with a specific left and right phoneme context.

The classification of phonemes usually is based on stochastic modeling by means of "Hidden Markov Models" (HMM). These models consist of states within a left-to-right Markov graph, whereby the states contain the distributions (pdf) of the feature vectors observed in these states. During the training step these distributions are determined from a large training set. The research work nowadays is concentrated on fast estimation methods and the adaptation of the HMMs to new speakers and new environments.

It seems rather attractive to combine the HMM approach with Neural Nets (NN) giving hybrid models. The pdfs within the HMMs can be properly calculated by means of NNs. In this way it is possible to get a discriminant representation of the feature vector distributions. These NNs can also favorably be used for determining the syllable nuclei and for measuring the speaking rate.

It is important to allow for pronunciation variants, which may be dependent on the speaking style and the speaking rate. These variants can be determined by inspection of large speech corpora or on the basis of rules.

Speech understanding can be logically divided into a speech recognition phase and an interpretation phase. It is important to evaluate both tasks in common. This means, the search for the chain of spoken words and the corresponding semantic structure is carried out simultaneously. This is facilitated if the semantic structure is described by stochastic modeling techniques, too. In our work the semantic structure is built up by semantic units which contain the single parts of the semantic meaning. The semantic understanding module describes the syntactic and semantic structure within a

probabilistic framework. Thus, recognition and understanding are able to yield a common optimal result.

Günther Ruske

4.5.1 Robustness to Speaking Rate in Automatic Speech Recognition

The processing of spontaneously spoken human-to-human dialogues is a special challenge for automatic speech recognition systems. Compared to read speech, where the speaking mode is well defined, in spontaneous speech several sources of variability contribute to changes in the speech signal. Among others, the speaking rate is an important factor influencing the speech signal. Unlike human listeners, which can cope with a large range of speaking rates without any problem, state of the art automatic speech recognition systems show severe degradations in recognition performance when the speaking rate is higher than normal. Therefore several approaches to improve robustness of hidden Markov model (HMM) based automatic speech recognition systems towards speaking rate were evaluated.

The implications of the speaking rate on the speech signal can be divided into three categories. There are timing effects, acoustic-phonetic and phonological effects resulting from varying speech rates. Whereas HMM-based recognizers can handle different lengths of phonetic units without any problem, the speech rate specific changes in the spectral domain as well as the use of variants differing from the standard pronunciation in general lead to higher confusions in the pattern recognition process. To be able to capture these effects, the corresponding knowledge sources of the recognizer have to be adapted. These are the acoustic models which represent the spectral characteristics of the phonetic units and the pronunciation dictionary which includes the pronunciations of the vocabulary of the recognition task. In comparison to speaker adaptation this adaptation problem is more difficult to handle, as the speaking rate can change even within one sentence, whereas the speaker's identity remains constant.

Firstly, in an explicit adaptation strategy comprising a rule and feature based estimation of the speaking rate [98pfa1] was combined with a switching between several speech rate specific recognition systems. A retraining of the acoustic models as well as changes in the pronunciation dictionary were performed to build the speech rate specific recognizers. Special emphasis was put on robust reestimation procedures for HMM-parameters such as maximum aposteriori (MAP) training [98pfa2]. This is a very important issue as the speech rate specific speech material for parameter reestimation is generally very limited.

Secondly, different types of normalization procedures were carried out to reduce the variations which have to be captured by the parameters of the HMMs. It proved to be useful either to reduce speech rate specific or speaker specific variations [99pfa1] as well as to capture pronunciation variants [99pfa1] to increase robustness towards speaking rate. Finally, it was also shown that the improvements of the different normalization procedures are almost cumulative [00pfa1].

Thilo Pfau

4.5.2 Robustness to Adverse Conditions: Speaker and Speaking Mode

In the past years there has been tremendous progress in the performance of speech recognition systems. Nevertheless there is still a variety of non-solved problems, such as different speakers, speaking style (fast, slow), noise.

It is well known, that speaker-dependent (=specifically trained for a speaker) systems outperform their speaker-independent counterparts. So in the recent years there has been a turnaround in the design of speech recognizers: most of them now inhere some kind of speaker-dependency.

Especially for closed-speaker systems, i.e. systems with a limited number of users, a combination of speaker identification and speaker-adaptation is very effective. The identification part allows the system to determine which speaker is currently using the recognizer, enabling the system to successively collect enrolment data for this particular speaker. By this way the system can be adapted by-and-by to the individual users, asymptotically reaching the performance of a speaker dependent system.

The key part of an identification system are the speaker models. Commonly, either Hidden Markov Models or some kind of vector quantized (VQ) codebooks can be used. A main advantage of the latter models is the fact, that they do not rely on any phonetic segmentation of the speech signal, a fact that is particularly advantageous for on-line application. So these models can be trained on whole utterances without the need of a phonetic segmentation. Such codebooks are usually built of K centroids, which can be trained using clustering techniques or sophisticated discriminative training algorithms.

With an increasing number of speakers the computational load becomes more excessive in the search process since all models have to be calculated in parallel. In this case efficient pruning strategies for the discarding of improbable speakers have to be applied. Another effective strategy lies in the application of tree-based speaker clusters, which can be pre-computed on the training data.

By collecting more and more enrolment data, the actual Hidden Markov Models can be adapted towards the speakers and their mode of speaking. The adaptation itself can be performed using training algorithms such as MLLR (Maximum Likelihood Linear Regression) in case the enrolment data is still sparse. If more data is available, Bayesian techniques like MAP (Maximum Aposteriori) or discriminative algorithms like MCE (Minimum Classification Error) or MMI (Maximum Mutual Information) can effectively be applied.

Robert Faltlhauser

4.5.3 Understanding Natural Speech

Human-machine interfaces may become much more intuitive and efficient if we enhance them by established modalities of interpersonal communication.

Speech is considered as the most common and therefore natural communication medium between human beings and offers advantages like hand-freeness and the ability of working in the dark when being used as an input medium. State-of-the-art speech recognizers in general allow only single-word commands or phrases containing a certain keyword. The research work carried out at the Institute for Human-Machine Communication led to a different approach which encourages the user to speak spontaneously in a most natural manner without the requirement of any learning process [P-L4, P-L5, 97mue, 98mue3, 97sta]. The only restriction of our system is the limitation to a single domain and a single language at a time. An average recognition rate of the right user's intention of 90% could be reached in different domains [98mue1].

The core of the system is realized in a top-down architecture of a one-stage maximum a-posteriori seman-tic decoder and a signal preprocessor. The stochastic semantic decoder utilizes pre-trained probabilistic knowledge on the semantic, syntactic, phonetic and acoustic levels. An integrated chart parser makes use of the Viterbi algorithm, calculating semantic, syntactic, and acoustic probabilities on the basis of Hidden-Markov-Models (HMM) and similar network structures. Its output, a semantic structure, is the input for a rule-based intention decoder which is placed "on top" of the system core. This decoder communicates bi-directional with an external application and allows for online control, provided that the application owns a well-defined command interface. One of the basic advantages of such an approach is the portability to different domains. Front-ends for several applications have already been successfully realized following this principle, e.g. for a graphics editor, for a service robot, for medical image visualization, for scheduling dialogues and for a mathematical formula editor. It will also be applied in the ongoing projects "speech in the automotive environment" (c.f. 4.1.2) and "navigation in VRML worlds" (c.f. 4.1.1).

Even the difficult task of automatic translation was successfully examined with this approach [99mue1]. The speech understanding system plus a language production module are capable of translating German phrases of a special domain into English, French or other languages. This module contains a word chain generator with the syntactic model and a linguistic postprocessor including grammar rules and the inflection model of the target language. Another add-on is an automatic HMM based speech detection module. It allows the user to speak at any time without pushing any button to indicate the beginning or end of dialogue instances.

Future research will deal with unknown words, confidence measures and the idea of integrating the intention decoder into the semantic decoder. The latter idea promises higher recognition performance due to the more abstract and less ambiguous nature of the intention compared with the semantic structure. We also aim to exploit the detailed contextual knowledge obtained from the current dialogue state to integrate constraints for even more robustness.

Björn Schuller
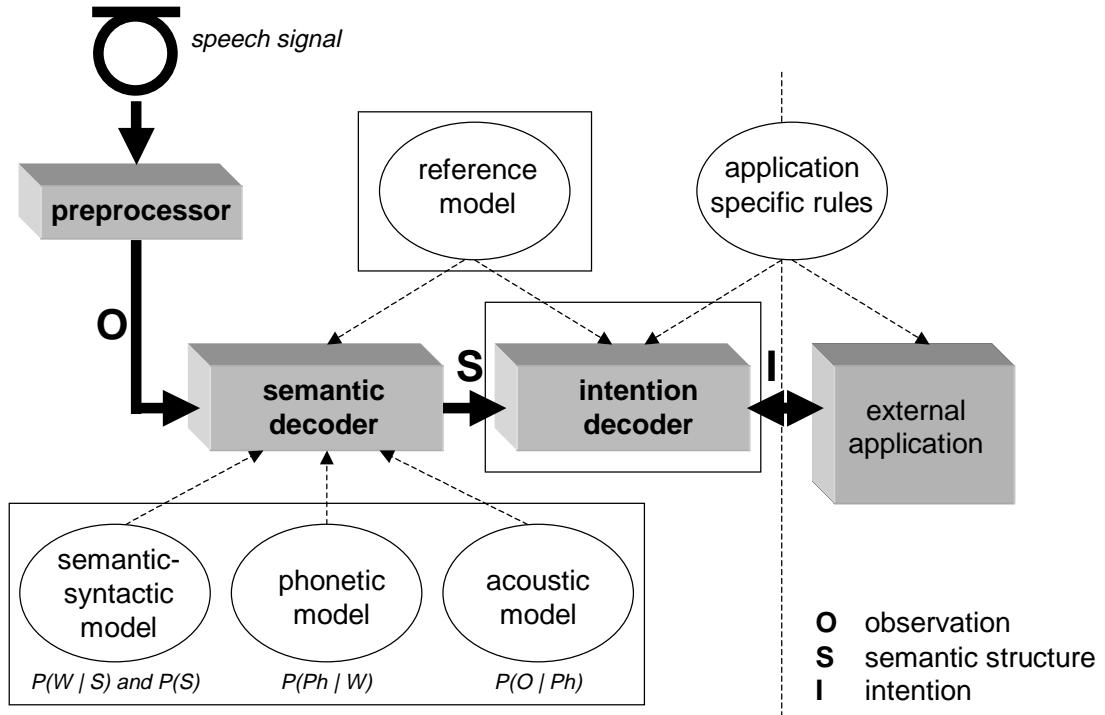
Fig. 9: Overview of a speech understanding system. The frames indicate parts that have to be adjusted for new domains.

4.6 Technical Acoustics and Noise Evaluation

4.6.1 Advanced Numerical Techniques in the Field of Vibro-Acoustics

Noise and vibration in the passenger compartment of a vehicle contribute substantially to the overall impression of a cars quality and therefore have a great influence on the buying decision of a customer. Because the time to market is constantly decreasing while the pressure to save costs is increasing, the developing engineers seek to define and optimize a vehicles vibrational and acoustic comfort characteristics as early as possible – even before the first prototype has been built. Due to the advances in the field of numerical structural acoustics and in computer technology, the finite element models used for this type of calculation have increased in size continuously, nearly overcompensating the fast growing performance of modern computers. Therefore one major deficiency of the FE-method are the very long computing times, which are in the range of several hours for a simple frequency response function up to several days for an optimization task even on a state of the art super-computer.

This fact, together with the batch-oriented computing style and the unwieldy and non intuitive user interfaces of commercial FE-calculation programs prevent a flexible and creative employment of FE-methods and a detailed understanding of the structural-acoustic coupling phenomena.

These problems have been successfully addressed in a cooperation with BMW during the last three years.

Using a new approach for the description of the vibro-acoustic equations of motion of model variants the computing time for modifications and optimization of coupled fluid-structure systems could be drastically reduced. The so-called modal correction technique allows to implement the entire optimization loop using generalized coordinates for the description of the dynamic state of the system, thus reducing the problem size roughly by a factor of 1000. Additional decrease of problem size and computation time was achieved

applying an adaptive mode reduction technique during the solution stage of the modal equations and by streamlining the dataflow and workflow.

The improvements render it possible to carry out vibro-acoustic calculations on a workstation with FE-models, which up to now could only be handled by super-computers. As a consequence, it is now possible to run the FE-calculations and optimizations in a quasi on-line, interactive way. This is still between 5 up to 120 times faster than the conventional approach on a super-computer.

To demonstrate the benefits of this new approach a software called VAO (Vibro-Acoustic Optimization) was developed, which offers a full palette of tools for the visualization, investigation, modification and optimization of coupled structure-fluid systems.

Lars Witta

4.6.2 A new Transfer Path Technique for the Analysis and Optimization of Vehicle Exterior Noise Characteristics

In most countries, the noise level of vehicle exterior noise is regulated by legislative limits. These restrictions have been intensified over the last two decades and further reductions of exterior noise level will be ratified. This trend requires much more acoustical optimization of cars. On the other hand, development periods in the automotive industry decrease, and a modification in the pre-production state is hardly possible. To meet this challenge, the department develops a new method for analyzing the exterior noise in cooperation with the BMW Group Munich. The method provides an easy way for calculating the exterior noise of a vehicle in the early development stage.

The ISO R 362 regulation states the required measuring procedure of the vehicle exterior noise. The car has to accelerate on a 20 meters track, where a microphone measures the noise 7.5 meters beside the lane. According to this pass-by test procedure, the maximum noise level is restricted to 74 dB(A) [1]. For acoustical analysis and developments, the BMW Group set up a pass-by test chamber, where the pass-by test can be simulated. In principle, it is the same situation as in ISO R 362, but with standing car and moving microphone [2]. All necessary measurements will be performed in this test chamber.

The exterior noise is being generated by a couple of noise sources like muffler, orifice, engine, etc. In order to reduce the exterior noise level effectively, the noisiest source has to be reduced. Therefore it is necessary to know the noise ranking of all components. A lot of effort has to be spent to determine this noise ranking [3]. The new method offers a nice and quick solution for this task [4]. It is divided into 3 main steps:

Step 1: Measurement of operation state

The noise field will be detected in the far field and near field by about 100 microphones p_k^{op}. For this setup, the car is in an operational state.

Step 2: Measurement of Transfer-Functions

The noise sources will be replaced by loudspeakers (see fig. 10). Each loudspeaker is driven with white noise and the transfer function from each speaker to every microphone will be detected. The loudspeaker parameter for the transfer functions will be the membrane acceleration a_j:

$$T_{kj} = p_k^{LS}/a_j.$$

Step 3: Synthesis of operational sound field

The principle idea is to replicate the operational sound field by loudspeakers. In order to synthesize the measured microphone sound pressure (step 1), the necessary speaker adjustments have to be calculated. The sound pressure synthesis will be virtually only. Therefore, in step two the transfer function of each loudspeaker is measured and in step three all loudspeakers will be adjusted and activated simultaneously. In this way, each speaker sounds like the component which it has replaced.

With this result, the noise ranking of all noise sources during the pass-by test can be determined. Furthermore, the calculation of the pass-by noise of a car with modified sources is possible. The only adjustments to be done are the loudspeaker sounds, according to the modified components.
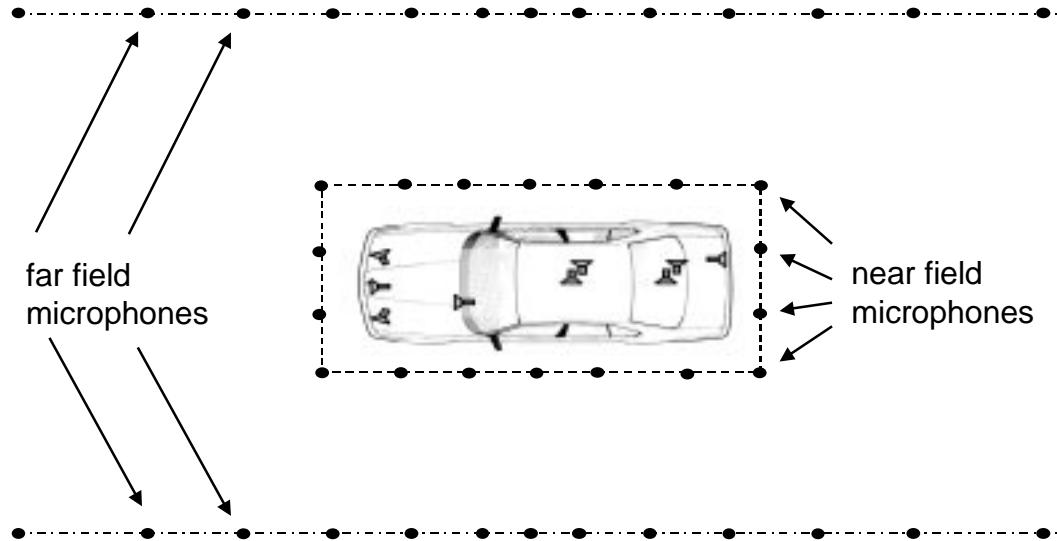
Fig. 10: Setup of step 2. Far field and near field microphones placed around the car. The noise sources are replaced by loud-speakers.

far field microphones

near field microphones

References

[1] Kevin, J.: Measuring Vehicle Pass-by Noise. Automotive Engineering, March 1995, pp. 28-32.

[2] Eilker, R. et al.: Die neuen BMW-Versuchseinrichtungen für Akustik. Automobiltechnische Zeitschrift ATZ 88 (1986), Heft 4, pp. 219-230.

[3] Buschmann, J.: Vorbeifahrt-geräuschmessung mit simultaner Erfassung von Drehmoment und Schlupf an einem geräuschiso-lierten Fahrzeug. VDI-Berichte Nr. 1224 (1995).

[4] Freymann, R.; Stryczek, R.; Riess, M. and Demmerer, S.: A new CAT-Technique for the Analysis and Optimization of Vehicle Exterior Noise Characteristics. ImechE Conf. Trans. „Vehicle Noise and Vibration 2000", London 2000.

Stephan Demmerer

4.6.3   Noise Evaluation

The concept that physical noise evaluation has to be based on features of the human hearing system was further put forward and described in detail in invited review papers [96fas2, 97fas1, 97fas3, 97fas4, 99fas1, 00cha1]. The description of noise emissions by loudness as standardized in DIN 45 631 is nowadays commonplace in most acoustic labs worldwide.

On the other hand, a firm psychoacoustic basis had to be established with respect to the physical measurement of noise immissions. Numerous studies were performed on industrial noise [97ste2] and traffic noise [98got2, 99got1, 00kuw1]. In particular, for the first time, the "railway bonus" could be confirmed also in laboratory studies [98fas1, 00kuw1]. In addition, the noise immission from leisure noise was studied for the example of tennis noise

[98ste1, 99fil1].

As a global result it turned out that the percentile loudness N_5, i.e. the loudness which is reached or exceeded in 5% of the measurement time, is a good indicator for the impact of noise immissions [00fas10]. Even the effects of "railway bonus" and "aircraft malus" can be predicted on the basis of N_5 [00fas7].

Despite the fact that loudness constitutes a dominant feature in the rating of sound quality, in particular for sounds with similar loudness, other hearing sensations like sharpness, fluctuation strength, or roughness may play an important role [00fas11].

For physical evaluation of sounds, new metrics were proposed which are described in an overview paper [98fas4]. In particular with respect to temporal aspects, signal processing algorithms in loudness analysis sys-

tems have to mimic in great detail features of the human hearing system [98wid1, 98wid2, 98wid3].

With respect to loudness of stationary sounds, data of different analysis systems available on the market are in good agreement with deviations of less than 5 % [97fas2]. In contrast, with respect to temporal processing, huge differences may occur for instruments of different manufacturers according to the degree of sophistication of the algorithms implemented [98fas3].

As concerns the physical measurement of noise immissions, statistical procedures were developed and implemented, which predict the accuracy of measurement, which can be achieved in a given measurement time [97ste4].

Hugo Fastl

4.7 Psychoacoustics and Audiological Acoustics

4.7.1 Psychoacoustics

A new, updated and extended edition of the book "Psychoacoustics – Facts and Models" was published. Some older material was re-arranged, text was adapted to current terminology, and new results were added, in particular in the chapters on pitch, fluctuation strength, roughness, and practical application (Zwicker and Fastl [99zwi1]).

The hearing sensation "pitch strength" was studied in great detail and the results are compiled in an overview paper (Schmid [99sch1]).

Among other things it could be shown that modulations enhance the pitch strength of low pass noise [98sch3], but reduce the pitch strength of pure tones [97sch3]. For complex tones, the interaction of "pointing tones" and pitch strength was studied in detail [97cha, 98sch1, 98sch2]. Correlations between pitch strength and frequency discrimination [98fas6] as well as effects of vibrato on the pitch strength were established [98hut1].

With respect to the Zwicker-Tone it could be demonstrated that also combinations of pure tone plus lopes noise, pure tone plus band pass noise, as well as pure tone plus band stop noise can produce a Zwicker-Tone [00fas11]. Current psychoacoustic models of the Zwicker-Tone could be confirmed. In addition, neurologically based models of the Zwicker-Tone nicely account for the psychoacoustical facts (paper in preparation).

Hugo Fastl

4.7.2 Sound Quality Design for High Speed Train Interior Noise

In cooperation with Müller BBM and with the Deutsche Bahn AG investigations concerning the topic "Sound quality design for high speed train interior noise" were carried out.

For the specification of the sound quality inside future high speed trains, tonal components produced by the motors or corrugated rails can play an important role. Therefore, in psychoacoustic experiments, the dominance of tonal components at 630 Hz or 1250 Hz was assessed [99hut1]. For an increase of the corresponding 1/3-octave band by 20 dB, a clear tonal character is audible which is only half as pronounced for an increase of 12.5 dB at 630 Hz or 10 dB at 1250 Hz. In line with the expectation, no tonal quality is perceived, if the 1/3-octave band in question is not enhanced. However, a decrease of sound energy in an 1/3-octave band by 20 dB can also produce a faint tonal sensation with a magnitude of about 1/10 of the tonal sensation produced by an increase of 20 dB. The results obtained with stimuli simulating the sound quality inside high speed trains are in good agreement with data from basic psychoacoustic experiments. Therefore, it is expected that sound quality evaluation of high speed train indoor noise can profit from a multitude of psychoacoustic data available.

Another aspect investigated is the disturbance of privacy in high speed trains which has become more evident in recent years due to successfully performed noise reduction measures [00pat1]. In psychoacoustic investigations, the contradictory requirements of unwanted speech intelligibility disturbing privacy on the one hand, and the desired sound quality on the other hand, were assessed. Early results show that in order to ensure privacy in large cabins of high speed trains across the tiers and at the same time not to reduce sound quality much, intensive shielding measures would be necessary. By means of basic psychoacoustic tools quantitative results are received by which a reasonable cost-benefit calculation could be established.

Christine Patsouras

4.7.3 Audiological Acoustics

The method of "line length" which has proven very successful for the evaluation of noise immissions was adapted for its use in audiology [98got1]. In comparison to the presently used categorical scaling [97bau], advantages for clinical applications were verified.

For the Ukrainian language, a speech test was developed, realized and recorded on CD [98cha1]. Moreover, the intelligibility of monosyllables in background noise was tested for the languages German, Hungarian, and Slovene [97ste1].

For patients with Cochlea-Implants, their ability to understand speech was tested both in quiet and in background noise [98fas2]. While in quiet surroundings, speech perception of Cochlea-Implant-Patients can be restored nearly to normal, in noisy environments they experience extreme problems and need in comparison to normal hearing persons more than 15 dB better signal to noise ratio [98fas1].

Hugo Fastl

4.7.4 Perceptual Consequences of Hearing Impairment: Modeling, Simulation and Rehabilitation

Nowadays about 12 % of the population in Germany suffer from hearing disorders. In order to develop signal processing algorithms and fitting procedures for hearing aids, detailed knowledge about perceptual consequences of hearing impairment is very important. However, this knowledge still is far from being complete.

Therefore, based on results of psychoacoustic experiments, Zwicker's model of loudness [99zwi1] was modified such that also the loudness for hearing impaired listeners can be predicted. This is achieved by fitting the loudness function to a specific hearing loss. Since the specific loudness time pattern can be regarded as an aurally adequate representation of sound, various other psychoacoustic hearing sensations like "fluctuation strength" can be modeled on the basis of this pattern. Results from hearing experiments concerning "loudness fluctuation" –

which is nearly the same as "fluctuation strength" – can be accounted for by the same model for normal and hearing impaired listeners, if loudness fluctuation is calculated from the modified specific loudness time pattern [99cha1, 00cha2].

From the specific loudness functions of normal and hearing impaired listeners, input-output functions for a signal processing system, which simulates hearing impairment, are deduced [00cha1]. Time-frequency analysis and synthesis within this system is done by the Fourier Time Transformation and its inverse [1, 98mum1]. Simulation of hearing impairment is helpful in evaluating hearing aid algorithms and fitting procedures for certain hearing losses as well as for providing normal hearing listeners with realistic demonstrations of auditory consequences of hearing impairment.

In cooperation with "GEERS Hörakustik" it was investigated, how the models of loudness and loudness fluctuation can be applied in an interactive psychoacoustic fitting procedure ("A-Life").

Moreover, in order to improve speech intelligibility in noise for hearing impaired listeners, the influence of so called "psychoacoustic processors" on speech intelligibility was assessed [00cha3] and – in cooperation with "Gasteig München" – magnetic induction loops, which are commonly used for speech transmission in churches and concert halls, were optimized [00cha4].

Supplementary reference

[1] Terhardt, E.: Fourier transformation of time signals: Conceptual revision. Acustica 57, 242-256 (1985).

Sepp Chalupper

4.8 Acoustical Communication

The main subject of research in the field of Acoustical Communication is the relation between the properties of acoustical signals on one hand, and the sensory analysis techniques of the human ear on the other hand. The book "Akustische Kommunikation", published by Professor Ernst Terhardt in 1998, treats almost any aspect of this interdisciplinary field of research and includes an audio-CD with demonstrations of various phenomena of auditory perception [98ter1].

In the four-year period of this report, research focused on the foundations of ear-adapted spectral representations, their relation to hearing perception, and their technical application. The following paragraphs give an overview of the research topics.

Linear Model of Peripheral-Ear Transduction

A system of linear filters was designed for modelling auditory preprocessing of sound [97ter]. It consists of a filter that accounts for the first two resonances of the ear-canal, and a filterbank that accounts for the spectral analysis of the inner ear. The design of the cochlear filter, which forms an element of this filterbank, takes advantage of knowledge about properties of the ear, in particular the threshold of hearing and the characteristics of tuning curves. Since the signal processing of the system allows easy digital computation, it is especially suited to be used as a front-end for models of more complex features of auditory perception.

Speech Coding

Speech coding algorithms were developed and investigated that are based on contours of auditory spectrograms, which are computed with an advancement of the Fourier Time Transformation (FTT) [1, 98mum1]. These contours are defined as "ridges" in the 3D-representation of the magnitude spectrogram. Contours bearing relevant information correspond, among other, to part tones perceptible by the ear. Starting off from a known representation, additional ridges ("time contours") and a new signal reconstruction procedure are introduced. A classification of ridges allows to separate tonal from noise-like signal components. Applying these foundations to data reduction algorithms, speech codecs with data rates down to 4 kbps were presented and evaluated.

Sound Quality of Piano Tones

Models and algorithms were developed to extract those sound signal parameters that are characteristic for the specific sound of a piano tone and its quality [98val1, 98val2, 00val1]. Just as the speech coding application mentioned before, the algorithms rely on the significance of contours of FTT spectrograms. Listening experiments were carried out to investigate the perceived dissimilarity and the quality of the sound of different piano tones. The results were utilized to design a model for the sound quality of piano tones and to validate the predictions. The new methods for the measurement of the discrimination criteria can suit the purpose of enhancing the quality of electronic and acoustic pianos and could be used for automatic quality control in musical instrument manufacturing.

Pitch Determination

A system for determination of the virtual and spectral pitches of non-stationary sound signals was developed [98rue1, 00rue1, 00rue2]. It allows for those essential properties of the human ear that can be observed in psychoacoustical experiments concerning pitch perception. Besides the elementary feature of frequency selectivity, mainly spectro-temporal contrast effects are to mention in this regard. These effects are not being addressed by previously known systems. The developed system is able to model both the existence and the prominence of perceived pitches in time-variant sound signals. Also this system uses contours of an auditory FTT spectrogram as a basis.

Graphic Sound Modification

A method to resynthesize CD-quality sounds from their auditory FTT spectrogram images was found, suggesting sophisticated sound modification via image processing [98hor1]. In particular, resynthesis of specifically selected spectrogram areas has confirmed the validity of strong audio-visual gestalt analogies.

Supplementary reference

[1] see section 4.7.4.

Claus von Rücker

5 Cooperative Research Projects

5.1 Industry Partners

BMW AG, München

Several projects were carried out in co-operation with BMW AG, München. Some of them are still in progress.

Speech Dialog Design. We analyzed and developed models and procedures for speech dialogs for the BMW navigation system. The main focus was upon the use of natural speech and the adaptation to specific situations [D97.14, D99.16].

ADVIA. The goal of the ADVIA project is to investigate the prerequisites of an intuitive, adaptive human-machine dialog in automobiles. The project is described in detail in section 4.1.2.

Active Resonator. In co-operation with BMW AG we developed and realized an active resonator in order to reduce low frequency noise in luxury cars [94spa]. Applying the active resonator, the sound quality of the interior noise of luxury cars can be improved [00fas8]. In addition, after long hours of driving, even in high quality cars, a feeling of "dizziness" can occur, which is substantially reduced if low frequencies are removed from the spectrum.

HPC-VAO. Within the scope of the EU-project "High Performance Computational Environment for Vibro-Acoustic Optimization" a simulation tool was developed, which allows the fast optimization of the vibro-acoustic properties of cars (see section 4.6.1). For that purpose new optimization methods were developed that drastically reduce the computation time of changes to existing models [P-L14, P-L16, 98wit1, 98mas1].

Exterior Noise Optimization. A new transfer path technique for the analysis and optimization of vehicle exterior noise characteristics is developed in co-operation with BMW AG. Please read section 4.6.2 for a detailed description.

Siemens AG

Speech Recognition. In co-operation with Siemens ZT, application-specific methods for in-service adaptation of Hidden-Markov Models in automatic speech recognition systems were developed [P-L10, 97bub1, 98bub1]. In another co-operation, a statistically modelled multi-lingual phoneme inventory for speech recognition systems was created [P-L9, 98koe1, 00koe1].

Document Analysis. A technique for attention-based localization and interpretation of information in paper documents was developed [P-L12, 99sch2].

Traffic Flow Control. Models for traffic flow were investigated in co-operation with Siemens AG. Nonlinear, discrete controllers were designed to reduce inhomogenities in traffic flow models [P-L7, P-L8, 97len1-4, 97wag1-2, 98wag1].

Consortium, consisting of BMW AG, DaimlerChrysler AG, Siemens AG and VDO Mannesmann

The above mentioned companies co-operate with the Institute for Human-Machine Communication in the FERMUS project, which deals with the investigation of error-robust, multimodal speech dialogs. Please read section 4.1.4 for a detailed description of the goals of this joint project.

DaimlerChrysler AG, Ulm

In co-operation with FAU Erlangen, we carried out a study for Daimler-Chrysler about the role of emotions in human-machine interaction. Please read section 4.4 for further information about this topic.

Audi AG, Ingolstadt

In co-operation with Audi AG we contributed to an integrated, user-adequate man-machine interface for automotive applications [D98.14].

Cellway GmbH, Hallbergmoos

Noise suppression methods for mobile phones used in the automotive environment were investigated in co-operation with Cellway [D-97.7].

Deutsche Bahn AG

In co-operation with Deutsche Bahn AG, the effects of noise reduction for freight trains, e.g. by noise barriers were studied. For practical purposes, two modes have to be distinguished, namely „to listen" or „to hear". As expected, difference limens in the mode „to hear" are substantially larger than in the mode „to listen" [97jae].

GSF medis, Neuherberg

In co-operation with GSF medis, we worked on a new, user-friendly man-machine interface for a system for the visualization and analysis of medical 3D image data [D98.12, P-L11, 97kra, 97kra2, 98kra1, 98mue2, 99kra1, 99kra2].

Institut für Rundfunktechnik (IRT), München

In co-operation with the Institute for Broadcasting Research (IRT), it could be demonstrated that the conventional definition of directivity patterns of electroacoustical transducers needs further clarification. Encouraging results are obtained, when the directivity pattern is divided in a vertical and a horizontal part [97wol].

Leicher GmbH, Unterföhring

In co-operation with Leicher, we worked on the analysis, implementation and evaluation of biometric identification systems for high-security areas [D98.19].

Mannesmann VDO Car Communication GmbH

A study about the state of the art of automatic speech recognition was conducted for Mannesmann/VDO.

In the SOMMIA-project, we develop and evaluate a concept for a speech-based man-machine interface (MMI) for several facilities in an automobile. Since a large variety of functions is comprised, the MMI has to be both efficient and intuitive to handle.

Müller-BBM, Planegg

In co-operation with Müller-BBM, active noise control in ducts was studied both theoretically and experimentally. With a Swinbanks array, tonal components of noises propagating in ducts can be reduced substantially [97sch1].

MVP München

In co-operation with MVP München, the noise produced by the German Magnetic Levitation Train (TRANSRAPID) was studied in psychoacoustic experiments. With respect to loudness, approximately the same noise immission is obtained for a TRANSRAPID at 400 km/h in comparison to an ICE with 250 km/h. However, if the TRANSRAPID runs also at 250 km/h, the advantage of the Magnetic Levitation System with respect to noise immission is about 15 dB(A). Therefore, as concerns noise immissions, the TRANSRAPID would be well suited to connect an airport with the center of a city [97got].

5.2 Collaborative Research Projects

Sonderforschungsbereich 204 "Nachrichtenaufnahme und -verarbeitung im Hörsystem von Vertebraten"

Collaborative Reseach Center "Reception and processing of information in the auditory system of vertebrates"

Sponsored by the Deutsche Forschungsgemeinschaft (DFG) from 1983-1997, three long-term projects within this research center were carried out at the institute. These projects dealt mainly with hearing sensations as a basis of acoustic information transfer, and complex perceptual processes on speech and music. The results of this proliferous research center were published in innumerable papers and are summarized in a substantial final report [00man1].

BMBF-Project "Verbmobil"

Verbmobil is a speaker-independent and bi-directional spech-to-speech translation system for spontaneous dialogs in mobile applications, and was running from 1993-2000. This multilingual system is able to handle German, English and Japanese spoken utterances. In this project 33 research groups were working together. The speech processing group of the Institute for Human-Machine Communication especially developed methods for robust speech recognition, for adaptation to new speakers and for tolerating various speaking styles [00hai1].

Graduiertenkolleg „Sprache, Mimik und Gestik im Kontext technischer Informationssysteme"

Graduate School "Speech, Mimic and Gestures in the context of technical information systems"

This graduate school funded by the DFG aims at the improvement of the interaction between man and machine in complex information systems. This goal is approached conjointly with an interdisciplinary team of scientists from the Ludwig-Maximilians-Universität and the Technische Universität München. More information is available at
`http://www.phonetik.uni-muenchen.de/GradKoll/`

Graduiertenkolleg „Interaktion sensorischer Systeme"

Graduate School "Interaction of Sensory Systems"

The group "Technical Acoustics" (Professor H. Fastl) of the Institute for Human-Machine Communication is part of the Graduate School "Interaction of Sensory Systems", sponsored by the Deutsche Forschungsgemeinschaft (DFG). In particular, interactions of acoustic and visual as well as acoustic and soma-to-sensory inputs are studied. Visit
`http://www.nefo.med.uni-muenchen.de/kolleg/` for more information.

Forschergruppe „Hörobjekte"

Research Group „Auditory Objects"

Within the Research Group "Auditory Objects" sponsored by the Deutsche Forschungsgemeinschaft (DFG), project 6 (Fastl) studies questions of basic psychoacoustics. In particular, new evidence about the Zwicker-Tone is compiled and implemented in psychoacoustic models [00fas11]. In addition, in co-operation with colleagues from project 7 (van Hemmen), neurophysiologically based models of the Zwicker-Tone are put forward (Fastl, Patsouras, Franosch, van Hemmen NL). Please visit
`http://www.hoerobjekte.vo.tu-muenchen.de/` for more information.

Klinikum Großhadern, LMU München

In an extended co-operation with the ENT Department of the Klinikum Großhadern, centered on the loudness perception of patients with different audiological diseases, new methods were developed which improve the understanding of loudness deficits in patients with different pathologies [97bau, 98got1].

University of Cambridge, UK

In co-operation with the lab of Roy Patterson at the University of Cambridge, UK, pitch strength of regular interval noise (RIN) was studied. Results described by a neuro-physiologically based model draw heavily on the autocorrelation function [99hir1, 99wie1].

Osaka University, Japan

The long standing, very effective co-operation with Professor Namba and Professor Kuwano of Osaka University, Japan was continued and extended. Topics of common interest are in particular evaluation of noise immissions, sound quality rating, and cultural differences in the interpretation of warning signals [99kuw1, 99kuw2].

6 Student Research Work

At the Institute for Human-Machine Communication, we incorporate our students into research work as early as possible. The students' contributions are reflected in diploma theses and research projects, which are listed in the following sections.

Please visit our website `http://www.mmk.ei.tum.de` if you are interested in an up-to-date list of available topics for student research work.

Note: After the subject of the work (in German) the name of the supervisor is given in parentheses.

6.1 Diploma Theses

1997

[D-97.1]   Philip Mackensen: Messung der Außenohrübertragungsfunktion eines Kunstkopfes bei verschiedenen Entfernungen (Haferkorn).

[D-97.2]   Thomas Filippou: Beurteilung der globalen Lautheit von Freizeitgeräuschen (Stemplinger).

[D-97.3]   Jochen Vallen: Klassifikationsverfahren zur Unterscheidung von Schallquellen in mechanisch belasteten Materialien (Ruske, Wolfertstetter).

[D-97.4]   Thomas Bachun: Implementierung und Erprobung von Verfahren zur Glättung von Spektralfunktionen als Hilfsmittel der gehörgerechten Konturierung (von Rücker).

[D-97.5]   Roland Trißl: Einfluß des Klangspektrums auf die subjektive Beurteilung von Klavierklängen (Valenzuela).

[D-97.6]   Tibor Fabian: Implementierung von Methoden der Sprecheradaption in einem automatischen Spracherkennungssystem (Ruske, Pfau).

[D-97.7]   Ramon Trombetta: Beiträge zur Grundlagenentwicklung für digitale Sprachfiltersysteme innerhalb einer Freisprecheinrichtung (Grau).

[D-97.8]   Dirk Vollmerhaus: Implementierung und Evaluierung von 3D-Applikationen für den gestikbasierten Mensch-Maschine-Dialog (Morguet).

[D-97.9]   Wolfgang Zimprich: Akustisch-mechanische Meßverfahren für die Komponenten eines implantierbaren Hörgeräts (Terhardt).

[D-97.10]   Klaus Menzl: Extraktion und Untersuchung gehörrelevanter Schallsignalparameter von Flügelklängen (Valenzuela).

[D-97.11]   Matthias Puchner: Visualisierung und numerische Auswertung dynamischer Mithörschwellenmuster (Schmid).

[D-97.12]   Matthias Bayerlein: MATLAB-basiertes Rechenschema zur Bestimmung der Ausgeprägtheit der Tonhöhe grenzfrequenzmodulierter Tiefpaßrauschen (Schmid).

[D-97.13]   Christine Huth: Die Ausgeprägtheit der Tonhöhe als Funktion psychoakustischer Empfindungsgrößen (Schmid).

[D-97.14]   Max Kreilein: Anpassung einer Bedienoberfläche für das Navigationssystem CARIN an eine Sprachsteuerung (Grau).

1998

[D-98.1]   Christina Niehaus: Untersuchungen zur erweiterten linearen Diskriminanzanalyse für die Spracherkennung (Ruske, Pfau).

[D-98.2]   Thorsten Ansahl: Untersuchungen eines Echokompensationsalgorithmus (Ruske).

[D-98.3]   Lam Song Nguyen: Intergration natürlichsprachlicher Interaktionen in eine virtuelle Umgebung für die medizinische Bildverarbeitung (Müller).

[D-98.4]   Ralf Nieschulz: Verarbeitung von Zeitstrukturen in FTT-Spektrogrammen (von Rücker).

[D-98.5]   Pierre Eric Gerard: Physical and subjective evaluation of speech quality for Hands-Free-Terminals (Fastl).

[D-98.6]   Dimitrios Haratsis: Portierung eines sprachübersetzenden Systems auf die Domäne gesprochener Terminvereinbarungen (Müller).

[D-98.7]   Lars Eric Fiedler: Programm zur Visualisierung mehrdimensionaler Meßdaten (Schmid).

[D-98.8]   Konstantin Spasokukotskij: Entwicklung eines Sprachtests für Ukrainisch (Fastl).

[D-98.9]   Gerald Stöckl: Merkmalsextraktionsverfahren für die Bildsequenzerkennung (Morguet).

[D-98.10] Ralf Schiffert: Einfluß zeitlicher Hüllkurvenverzerrungen auf die Sprachverständlichkeit (Chalupper).

[D-98.11] Selim Ben Yedder: Labelung, zeitliche Segmentierung und Klassifikation zusammenhängender Gesten (Morguet).

[D-98.12] Thomas Volk: Entwicklung und Integration einer menübasierten 3D-Bedienoberfläche für die medizinische Bildanalyse in virtuellen Umgebungen (Müller, Hunsinger).

[D-98.13] Dimitrios Patsouras: Zur Ausgeprägtheit der Tonhöhe und ihrer Abhängigkeit von anderen Empfindungsgrößen (Schmid).

[D-98.14] Michael Geiger: Grundlagen für ein integriertes Bedienkonzept im Fahrzeug (Grau).

[D-98.15] Stefan Fissek: Informationstheoretische Untersuchung menschlichen Fahrverhaltens (Grau).

[D-98.16] Martin Zobl: Sprachverständlichkeitsvorhersage in einem perzeptiven Modell (Chalupper).

[D-98.17] Tarek Said: Entwicklung und Test eines Verfahrens zur Aktivierung eines Einzelworterkennungssystems durch ein spezielles Kommandowort (Ruske).

[D-98.18] Stefan Hirsch: Experimental Investigation and Modelling of the Pitch Strength of Regular Interval Sounds (Fastl).

[D-98.19] Meinrad Kranewitter: Analyse, Implementierung und Evaluierung biometrischer Identifikationssysteme im Hochsicherheitsbereich (Morguet).

1999

[D-99.1]   Christian Hansen: Entwicklung von Lösungsansätzen zur Nutzung des Internet im Fahrzeug (Grau).

[D-99.2]   Markus Fruhmann: Stationäres Lautheitsmodell für Innenohrschwerhörige (Chalupper).

[D-99.3]   Bernhard Seeber: Zur Frequenzselektivität des Gehörs: Zusammenhänge zwischen Mithörschwellen und Tuningkurven (Terhardt).

[D-99.4]   Lutz Hönisch: Kontextabhängige Modellierung von sprachlichen Einheiten in der automatischen Spracherkennung mit Hilfe von Polyphonen (Faltlhauser, Pfau, Ruske).

[D-99.5]	Matthias Thomae: Optimierung der erweiterten linearen Diskriminanzanalyse durch Modelltransformation (Ruske, Pfau).
[D-99.6]	Zeljko Miletic: Einkanalige Geräuschreduktion für Mobiltelefon-Freisprechanlagen im Auto (Ruske).
[D-99.7]	Jörg Franke: Untersuchungen zur Halligkeit von zeitvarianten Schallen (Chalupper).
[D-99.8]	Markus Szymkowiak: Computergestützte Erkennung der Position und der Abmessung des menschlichen Gesichtes im digitalen Bild für die automatische Bildverarbeitung (Morguet).
[D-99.9]	Thilo Hinz: An interaktive support environment for figuring, programming, and using the PLC Siemens SIMATIC S7-2000 (Hofmann).
[D-99.10]	Jasmin Jin Qian: Implementierung eines Clustering-Verfahrens zur Inferenz in Bayes'schen Netzen (Hofmann).
[D-99.11]	Andrés Ucke: Entwicklung eines Algorithmus zur automatischen Generierung einer Planbibliothek (Hofmann).
[D-99.12]	Günther Steinmaßl: Evaluierung und Optimierung einer stochastischen Gestikerkennung für den visuellen Dialog (Morguet).
[D-99.13]	Björn Schuller: Automatisches Verstehen gesprochener mathematischer Formeln (Hunsinger).
[D-99.14]	Jürgen Straßer: Einsatz des HTK-Toolkits für die kontextabhängige akustische Modellierung in der automatischen Spracherkennung (Ruske, Pfau, Faltlhauser).
[D-99.15]	Richard Stenzel: Natürlichsprachliche Eingabe für das Automobil (Hunsinger, Grau).
[D-99.16]	Bernhard Niedermaier: Entwicklung und Bewertung eines nutzerorientierten Dialogkonzepts zur Sprachbedienung eines Autotelefons (Grau).
[D-99.17]	Mark Schützendorf: Integration eines DTW-Musterbewertungsverfahrens in einen statistischen semantischen Decoder zur Erkennung handgeschriebener mathematischer Formeln (Hunsinger).

2000

[D-00.1]	Christian Lorenz: Elektro- und psychoakustische Untersuchungen an einem Oberwellengenerator (Chalupper).
[D-00.2]	Stephan Bayer: Subjektive Beurteilung von Reifen-/Fahrbahngeräuschen (Fastl, Patsouras).
[D-00.3]	Philipp Detemple: Natürlichsprachliche Bedienung im Kraftfahrzeug – Erweiterung des Dialogkonzeptes um eine Komponente zur Steuerung von Mensch-Maschine-Dialogen (Neuss).
[D-00.4]	Martin Rosnitschek: Multimodale Mensch-Maschine-Kommunikation im Fahrzeug (Neuss).
[D-00.5]	Rainer Schiller: Entwicklung und Evaluierung eines adaptiven Benutzerassistenzsystems (Hofmann).
[D-00.6]	Thomas Filippou: Audio-visuelle Interaktion beim Lautheitsurteil (Patsouras).
[D-00.7]	Oliver Thost: Resynthese von Getriebegeräuschen aus dem Betragsspektrum (Fastl, Chalupper).
[D-00.8]	Michael Kellnberger: FTT-basierte Simulation und Rehabilitation von Schwerhörigkeit (Chalupper).
[D-00.9]	Robert Lieb: Einstufig-probabilistische semantische Decodierung handgeschriebener mathematischer Formeln (Hunsinger).
[D-00.10]	Derik Schröter: An Integral Approach to Real-time Segmentation and Tracking of Objects without explicit Modeling (Zobl, Althoff).
[D-00.11]	Hubert Wimmer: Sprachübertragung mit Induktionsschleifen: Messtechnische und psychoakustische Untersuchungen (Chalupper).

6.2   Student Research Projects

1997 - 2000

[S-97.1]   Michael Geiger: Erstellen eines Programmes zur Berechnung des Vertrauensbereiches von Lautheitsperzentilen (Stemplinger).

[S-97.2]   Dimitrios Patsouras: Zur Ausgeprägtheit der Tonhöhe grenzfrequenzmodulierter Tiefpaßrauschen bei unterschiedlicher Filter-Flankensteilheit (Schmid).

[S-97.3]   Bernhard Niedermaier: Drehtisch-Schrittmotorantrieb zur Positionierung elektroakustischer Wandler (Schmid).

[S-98.1]   Tarek Said: Differenztonverfahren - Realisierung mit "System One" (Schmid).

[S-98.2]   Bernard Payer: Erzeugung eines variablen Bandpasses mit großer Flankensteilheit (Schmid).

[S-98.3]   Robert Lieb: Integration neuer Hidden-Markov-Modelle in ein sprachverstehendes System (Ruske, Hunsinger).

[S-00.1]   Andrea Nobbe: Aufnahme, Analyse und Resynthese von Instrumententönen (Patsouras).

7 Miscellaneous Events

7.1 Conferences and Symposia

7.1.1 ICASSP, April 21-24, 1997

The 22nd IEEE International Conference on Acoustics, Speech, and Signal Processing was held in Munich in 1997, for the first time in Germany. Staff members of the Institute for Human-Machine Communication took an active part in the conference committee: it was chaired by M. Lang, G. Ruske was in charge of publications, and H. Fastl took care of the registration procedure. More than 1000 authors presented their contributions to research in the fields of digital signal processing, speech, and acoustics; and some 1800 participants attended this top event.

Summaries of the most relevant new ideas were presented by 8 leading experts from each technical committee on the last day of the conference. These summary sessions were highly appreciated by the ICASSP 97 attendees.

Other highlights included the combined opening and awards ceremony in the Gasteig concert hall, where the 1996 IEEE Signal Processing Society awards were presented. The Bavarian Government invited ICASSP attendees to a State Reception in the "Kaisersaal" of the Munich Residence, emphasizing the significance of the event. A Bavarian evening was arranged in one of the Munich brewery halls to make known typical entertainment and folklore to the international audience. Proceedings in print and on CD-ROM were produced to summarize the results. As the positive response from the participants showed, ICASSP 97 was a success; and it paved the way for other conference venues outside of the US, e.g. Istanbul, host of the ICASSP 2000.

7.1.2 WE-Heraeus-Seminar "Speech Recognition and Speech Understanding", April 3-5, 2000

Sponsored by the WE-Heraeus-Stiftung, Professor M. Lang chaired a seminar about Speech Regognition and Speech Understanding at the Deutsche Physikalische Gesellschaft (DPG) in Bad Honnef. Professor G. Ruske and Professor H. Fastl were members of the organizing committee. 16 internationally renowned experts were invited to give lectures on the state of the art and current problems at issue in the field of speech processing. For instance, Professor G. Ruske contributed a lecture about "Robust Speech Recognition." Moreover, representatives of information technology companies gave an insight into industrial applications of speech technology and marketing aspects as well. It was generally agreed that the seminar gave all attendees the opportunity for fruitful discussions and mutual learning.

7.1.3   euro-noise 1998

The European Noise Conference euro-noise 1998 was held for the first time in Germany, and co-chaired by Dr. Joachim Scheuren from Müller-BBM and Professor Hugo Fastl from the Institute for Human-Machine Communication. The conference had the motto "Designing for Silence", and internationally renowned experts discussed questions of prediction, measurement and evaluation of noise and vibration. In addition, an exhibition showed the latest developments in materials for noise abatement as well as new algorithms for assessing questions of noise problems. The results are compiled in the proceedings edited by H. Fastl and J. Scheuren [98fas7].

7.1.4   Japan in Deutschland

Within the framework of "Japan in Deutschland" Professor Sonoko Kuwano from the Laboratory of Environmental Psychology of Osaka University, Japan and Professor Hugo Fastl from the Institute for Human-Machine Communication organized a joint symposium of the Acoustical Society of Japan (ASJ) and the German Acoustical Society (DEGA). Experts from both countries discussed about questions of noise measurement and evaluation. The results are available in the volume *Fortschritte der Akustik* – DAGA 2000, published by DEGA, Oldenburg. In addition, a publication of these papers in the Journal of the Acoustical Society of Japan, English Version (JASJ(E)) is planned.

7.1.5   Miscellaneaous Conferences and Symposia

Professor M. Lang was involved with the organization or was a member of the program committee of the following conferences:

D-CSCW 2000: "Verteiltes Arbeiten – Arbeit der Zukunft", 11.-13.9.2000, München.

Münchner Kreis: "Anwenderfreundliche Kommunikationssysteme", 16.-17. 6.1999, München.

VDE/ITG-Fachtagung: "Technik für den Menschen", 26.-27.10.1998, Eichstätt.

Workshops des Arbeitskreises "Physik, Informatik, Informationstechnik" der Trägergesellschaften DPG, GI, VDE/ITG:

- Optimierung in Physik und Informatik, 8.-11.9.1997, Rostock
- Quanten-Computing, Dynamische Systeme, Datenaquisition, Computeranwendungen, 15.3.1999, Heidelberg

- Optik in der Rechnertechnik, 7.10.1999, Jena
- Optik in der Rechnertecnik und Mikrooptik, 27.-29.9.2000, Hagen.

Professor G. Ruske was Head of the organizing committee of the "Verbmobil Akustik-Workshop", held in Munich, 25.-26.11.1999.

7.1.6   Invited Talks

Professor Manfred Lang gave the following invited talks:

- Gibt es den Computer, der sich auf den Nutzer einstellt? BMBF, Bonn, 25.2.1997.
- Entwicklungen und Visionen für eine benutzeradäquate Mensch-Maschine-Kommunikation. Siemens AG Schweiz, World Trade Center, Zürich, 24.9.1997.
- Beherrschung der Komplexität bei multimodaler Mensch-Maschine-Kommunikation (with Professor B. Radig). Symposium "Realität und Abstraktion", TU München, 2.3.1998.
- Digitale Entdeckungsreisen – Fernsehen als Datenstrom. Panel discussion at Systems 98, München, 23.10.1998.
- Aspekte einer natürlichen Mensch-Maschine-Kommunikation. Betaresearch, Andechs, 8.-9.4.1999.
- Mensch-Maschine-Kommunikation. Diehl-Stiftung, Lengenfeld, 16.-17.11.1999.
- Können Maschinen wie Menschen kommunizieren? Tag der Fakultät Elektrotechnik und Informationstechnik, TU München, 7.7.2000.
- Television interview by *Focus TV* and presentation of running research work in the institute's laboratories (20.7.1999).

Professor Hugo Fastl presented keynote lectures at the DAGA 1997 in Kiel, Germany, at the Sound Quality Symposium 1998 in Ypsilanti, USA, and at the WESTPRAC 2000 in Kumamoto, Japan.

Professor Günther Ruske gave an invited lecture about "Robust Speech Recognition" at the GuHT-Workshop in Munich, 20.10.2000, organized by the "Gesellschaft & High-Tech eV."

Miscellaneous Events

7.2   Retirement of Professor Ernst Terhardt

On the evening of 23 April 1999, some 90 invited guests attended the celebration to mark the occasion of the retirement of Professor Ernst Terhardt, Professor of Acoustical Communication and Musical Acoustics at the Technische Universität München.

At the momentous occasion, Professor Hugo Fastl, a long standing colleague, spoke about Professor Terhardt's scientific career in the field of acoustics.

Professor Terhardt earned his degree in Communications Engineering at the University of Stuttgart, where he also graduated. He was promoted to Professor at the Technische Universität München in 1970. His research experience is wide ranging: auditory roughness and pitch perception, auditory time and frequency resolution, musical con-

sonance, the physics of string instruments, and signal theory were major topics of his research work. His concept of "virtual pitch," established in the early 1970ies to explain the pitch perception of complex tones, is certainly his most influential work. In 1998, Professor Terhardt published the essence of his scientific achievements in his substantial book "Akus-

tische Kommunikation" [98ter1].

The music for the festive occasion was provided by Manfred Seewann, a former student and co-worker of Professor Terhardt, who is a professional musician. He played pieces of Chopin on the grand piano for the pleasure of all those present.

After the presentations, the evening was completed by a buffet supper, which gave all guests the opportunity to both utilize and talk about Acoustical Communication. The evening was an enourmous pleasure for all those present, and much appreciation goes to all who contributed to the organisation of the event. We all wish Professor Terhardt a very happy and positive retirement from University and – as a somewhat selfish wish – that he will continue his productivity.

Claus von Rücker

7.3   Honors and Awards

Professor M. Lang frequently reviews research projects for the following organizations:

- Bundesministerium für Bildung und Forschung – BMBF
- Deutsche Forschungsgemeinschaft – DFG
- several state governments
- VW-Stiftung
- Karl Heinz Beckurts-Stiftung

Professor H. Fastl was one of three scientists from Germany who were granted the Research Award of the Japan Society for the Promotion of Science (JSPS) in 1998. He also frequently reviews research projects for the DFG.

Professor G. Ruske is a permanent reviewer for the Australian Research Council (ARC) for large research grants.

Johannes Müller and Holger Stahl received the *Rohde & Schwarz*-prize in 1998 for their outstanding doctoral dissertations [P-L4, P-L5]. Their work about a single-stage stochastic approach to natural speech understanding was awarded therewith. Meanwhile, H. Stahl holds a professorship at the Fachhochschule Rosenheim.

8 Publications

8.1 Doctoral Dissertations

8.1.1 Supervised by Professor Lang

[P-L3]
Hans-Jürgen Winkler (1996):

Entwurf und Realisierung eines auf statistischen Ansätzen basierenden Systems zur Erkennung handgeschriebener mathematischer Formeln

Design and Realization of a System for the Recognition of Handwritten Mathematical Formulas Based on a Statistical Approach

In dieser Arbeit wird ein System zur Erkennung handgeschriebener mathematischer Formeln vorgestellt. Die Problemstellung bestehend aus Symbolsegmentierung, -erkennung und struktureller Analyse wird hierbei mittels eines statistischen Ansatzes beschrieben und unter Verwendung von wissensbasierten und stochastischen Verfahren bearbeitet. Im Gegensatz zu den bisher vorgestellten Analysemethoden können somit Entscheidungsalternativen innerhalb der einzelnen Verarbeitungsstufen toleriert und im weiteren Verlauf durch neu erworbenes Wissen automatisch aufgelöst werden. Die erzielten Erkennungsergebnisse demonstrieren die Leistungsfähigkeit des realisierten Systems.

[P-L4]
Johannes Müller (1997):

Die semantische Gliederung zur Repräsentation des Bedeutungsinhalts innerhalb sprachverstehender Systeme

The Semantic Structure for the Representation of Meaning in Speech Understanding Systems

Die semantische Gliederung wird als eine neuartige Repräsentation des Bedeutungsinhaltes einer gesprochenen Äußerung aus einer vorgegebenen Domäne innerhalb eines sprachverstehenden Systems vorgestellt. Da sie eine probabilistische Aussage über die ihr zugrundeliegende Wortkette erlaubt, wird die unmittelbare Decodierung einer Sprachsignal-Merkmalsvektorenfolge in eine solche semantische Gliederung durch einen rein stochastischen Algorithmus ermöglicht. Als Beispielapplikation wurde ein „sprachverstehender Grafikeditor" implementiert, mit dem dreidimensionale Objekte auf dem Bildschirm mit natürlichsprachlichen Kommandos erzeugt, verändert oder gelöscht werden können. Durch Übertragung der Algorithmen in einen „sprachverstehenden Serviceroboter" wurde der anschauliche Nachweis der System-Portabilität erbracht. Darüber hinaus ermöglicht die semantische Gliederung als Interlingua-Ebene die automatische Übersetzung von natürlicher, gesprochener oder geschriebener Sprache.

[P-L5]
Holger Stahl (1997):

Konsistente Integration stochastischer Wissensquellen zur semantischen Decodierung gesprochener Äußerungen

Consistent Integration of Stochastic Knowledge for Semantic Decoding of Spoken Utterances

Diese Arbeit beschreibt die Entwicklung eines Systems zum Verstehen natürlicher, fließend gesprochener Sprache. Den Kern des Systems bildet ein semantischer Decoder, der das Sprachsignal einer Äußerung auf den zugehörigen Bedeutungsinhalt abbildet. Dazu wird eine Maximum-a-posteriori-Klassifikation durchgeführt, d.h. auf der Basis stochastischen Wissens wird der wahrscheinlichste Bedeutungsinhalt zum gegebenen Sprachsignal ermittelt. Die Einführung der semantischen Gliederung zur Repräsentation des Bedeutungsinhaltes und die konsistente, nahtlose Verknüpfung der stochastischen Wissensquellen ermöglichten eine äußerst effiziente Implementierung des semantischen Decoders mit hoher Treffsicherheit.

[P-L6]
Robert Zwickenpflug (1997):

Entwurf und Realisierung eines Systems zur Erstellung von verteilten Anwendungen für kontinuierliche Medien

Design and Implementation of a System for the Creation of Distributed Applications for Continous Media

Zur Erstellung von modularen verteilten Anwendungen für kontinuierliche Medien wird ein Client-Server-System vorgestellt. Es erlaubt, Dienste in für den Endbenutzer einfach zu handhabender Art und Weise auf einem Rechnernetz zu verteilen und mehreren Benutzern zugänglich zu machen. Dienste können über definierte Anschlüsse miteinander kommunizieren. Jeder Benutzer kann neue Dienste in das Rechnernetz an einer von ihm frei wählbaren Stelle einbringen und diese untereinander und mit bereits vorhandenen Diensten verbinden. Er kann bei diesen Verbindungen auch Dienste mitverwenden, die von einem anderen Benutzer eingebracht worden sind.

[P-L7]
Christoph Wagner (1997):

Verkehrsflußmodelle unter Berücksichtigung eines internen Freiheitsgrades

Traffic Flow Models Allowing for an Inner Degree of Freedom

Ausgehend von einer kinetischen Verkehrsgleichung auf einem durch die Wunschgeschwindigkeit der Fahrer erweiterten Ort-Geschwindigkeit-Phasenraum wird durch Momentenbildung ein verbessertes makroskopisches Verkehrsflußmodell abgeleitet. Das Modell zeigt ein realistisches dynamisches Verhalten über den gesamten Dichtebereich und liefert neben der genauen Form und der funktionalen Abhängigkeit von bisher nur heuristisch eingeführten Termen der Modellgleichungen auch die dazugehörigen Transportkoeffizienten. Weiterhin erlaubt der zusätzlich eingeführte Freiheitsgrad und die davon abgeleiteten Größen nun eine direkte Modellierung von Regeleingriffen.

[P-L8]
Henning Lenz (1999):

Entwicklung nichtlinearer, diskreter Regler zum Abbau von Verkehrs-flußinhomogenitäten mithilfe makroskopischer Verkehrsflußmodelle

Design of Nonlinear, Discrete Controllers to Reduce Inhomogenities in Traffic Flow Models

Ein Schema für die Entwicklung nichtlinearer Regler wurde vorgestellt, mit dem Ziel, Inhomogenitäten im Straßenverkehr abzubauen. Die Anforderungen an einen derartigen Regler wurden modellunabhängig formuliert. Eine Datenanalyse zeigte, daß Geschwindigkeitsbegrenzungen so geschaltet werden können, daß sie diese Anforderungen erfüllen. Für einen effizienten Abbau von Stop-&-Go-Wellen bietet sich eine im Ort vorausschauende Strategie an. Weitere Anwendungen in der Verkehrstechnik wurden dargestellt.

[P-L9]
Joachim Köhler (1999):

Erstellung einer statistisch modellierten multilingualen Lautbibliothek

Creation of a Statistically Modelled Multi-lingual Phoneme Inventory for Speech Recognition

Die vorliegende Arbeit beschreibt die Entwicklung einer multilingualen Lautbibliothek für die statistische Spracherkennung. Dazu werden die akustisch-phonetischen Ähnlichkeiten zwischen verschiedenen Sprachen ausgenutzt. Basierend auf der HMM-Technologie werden Verfahren entwickelt, mit denen die sprachspezifischen Modelle in multilinguale Lautmodelle überführt werden. Dadurch läßt sich eine drastische Einsparung von Modellparametern erreichen, ohne daß ein signifikanter Abfall der Worterkennungsrate auftritt. Im zweiten Teil der Arbeit werden dann Methoden zur Portierung der multilingualen Sprachlaute in neue Sprachen entwickelt und beschrieben.

[P-L10]
Udo Bub (1998):

Anwendungsspezifische Online-Anpassung von Hidden-Markov-Modellen in automatischen Spracherkennungssytemen

Application-Specific In-Service Adaptation of Hidden-Markov Models in Automatic Speech Recognition Systems

Die Arbeit befaßt sich mit den Problemen, die in der automatischen Spracherkennung entstehen, wenn zwischen dem Trainings- und Testdatensatz eine Fehlanpassung vorliegt. Insbesondere Ungleichheiten bei den akustisch-phonetischen Lautkontexten führen zu einer verschlechterten Erkennung. Diesem Trend wird durch neuartige Lernalgorithmen entgegengewirkt, die während der Anwendungsphase online ablauffähig sind. Bei unüberwachtem Lernen kann bei 6000 Adaptionswörtern die Fehlerrate um 56 % gesenkt werden, bei überwachtem Lernen um 67 %. Dies entspricht der Erkennung eines Modells, das im Falle des Vorliegens geeigneter Sprachdatenbanken hätte trainiert werden können.

[P-L11]
Christian Krapichler (1999):

Eine neue Mensch-Maschine-Schnittstelle für die Analyse medizinischer 3D-Bilddaten in einer virtuellen Umgebung

A New Man-Machine Interface for the Analysis of Medical 3D-Image Data in Virtual Reality

Durch die Entwicklung neuer Verfahren der 3D-Visualisierung und der intuitiven räumlichen Interaktion entstand ein VR-System, mit dem alle Arbeitsschritte der digitalen medizinischen Bildanalyse durchgeführt werden können. Die neuen Interaktionsmethoden umfassen die Analyse von Handgestik, Sprachverstehen und den Einsatz von VR-Eingabegeräten ebenso wie innovative virtuelle Werkzeuge. Im Vergleich zur heute üblichen Darbietung unzähliger Schichtaufnahmen erleichtert der Einsatz des entwickelten VR-Systems das Erfassen und analysieren räumlicher Zusammenhänge und die weitere Verarbeitung der tomographischen Bilddaten. Durch die an die menschlichen Sinne und Fähigkeiten angepaßten Darstellungs- und Interaktionsformen ist es dem Mediziner möglich, den gesamten Arbeitsablauf in einer Zeitspanne zu bewältigen, die den Einsatz im klinischen Alltag erlaubt.

[P-L12]
Angela Engels (formerly Schreyer, 2000):

Aufmerksamkeitsbasierte Lokalisierung und Bewertung relevanter Information auf Papierdokumenten

Attention-Based Localization and Interpretation of Information in Paper Documents

Die Arbeit beschreibt in einem Sender-Empfänger-Modell eine neue, aufmerksamkeitsbasierte Sichtweise auf Dokumente: der Autor markiert relevante Informationen auf dem Dokument durch auffällige gestalterische Merkmale, die die Aufmerksamkeit eines Lesers auf den ersten Blick anziehen und ihm so eine effiziente Informationsextraktion ermöglichen. Die Arbeit setzt diesen Mechanismus in ein technisches Verfahren um, das ausschließlich auf dem Bild eines eingescannten Papierdokuments relevante Information findet und die Relevanz der einzelnen Informationen relativ zueinander beurteilt. Wichtige Schritte bei der Umsetzung sind die Formalisierung einer psychologischen Theorie zur Texturwahrnehmung und eine Befragung zur Wahrnehmung von gestalterischen Merkmalen.

[P-L13]
Peter Morguet (2000):

Stochastische Modellierung von Bildsequenzen zur Segmentierung und Erkennung dynamischer Gesten

Stochasic Modelling of Image Sequences for the Segmentation and Recognition of Dynamic Gestures

In der Arbeit wird die Entwicklung eines bildverarbeitungsgestützten Systems für den mit Handgesten gesteuerten Mensch-Maschine-Dialog vorgestellt. Mit zwei alternativen Ansätzen, die auf der stochastischen Modellierung mit teilweise erweiterten Hidden-Markov-Modellen beruhen, werden gestische Bewegungen im kontinuierlichen Videostrom zeitlich segmentiert und klassifiziert. Zur Anpassung der räumlich-zeitlichen Bildsequenzen an die serielle Verarbeitung werden mehrere Merkmalsextraktionsverfahren entwickelt und vergleichend untersucht. Als Beispielanwendung wird die Implementierung eines echtzeitfähigen dreidimensionalen Szenen-Editors beschrieben. Über das Konzept der indirekten Manipulation sind hierin auch komplexe Aktionen über Gesten intuitiv steuerbar.

[P-L14]
Dietmar Mass
(work in progress):

Schnelle rechnerische Komfortoptimierung von Kraftfahrzeugen mittels modaler Korrektur

Fast Computational Optimization of Automotive Comfort by Means of Modal Correction

Ausgehend von einer Energieschreibweise der Bewegungsgleichungen eines gekoppelten Struktur-Hohlraum-Systems wird ein Verfahren entwickelt, das auf der Basis modaler Korrekturen eine Berechnung von Modifikationen in gegenüber konventioneller Finite-Elemente-Rechnung drastisch reduzierter Zeit ermöglicht. Es erlaubt auch das Einbinden experimenteller Methoden, um die Modellierunsgüte des Rechenmodells zu verbessern. In grundlegenden und praxisbezogenen Beispielen wird die Leistungsfähigkeit des Verfahrens demonstriert und seine Anwendungsmöglichkeiten zur Berechnung komfortrelevanter Fahrzeugeigenschaften, wie Schwingungsverhalten und Innenraumakustik, dargestellt.

[P-L15]
Robert Neuss (formerly Grau, work in progress):

Usability Engineering als Ansatz zum multimodalen Mensch-Maschine Dialog

Usability Engineering as Approach to Multimodal Human-Machine Interaction

Multimodale Mensch-Maschine-Kommunikation soll die Benutzung von Software durch freie Verwendung von Kommunikationskanälen wie Sprache, Gestik etc. erleichtern. Diese Arbeit untersucht zuerst Einzelmodalitäten, um die Eigenschaften eines multimodalen Systems postulieren zu können. Gemäß des Usability Engineerings wird dann ein Prototyp aufgebaut, um Benutzertests durchzuführen. Die Entwicklung dieses Systems, welches in einen Fahrsimulator integriert ist und die Bedienung von Komponenten wie Radio, Navigationssystem und Telefon ermöglicht, erfolgt zyklisch durch Tests und Verbesserungen. Die Resultate sind ein benutzeradäquates Design sowie Praxiserfahrungen mit den neuen Techniken.

[P-L16]
Lars Witta (work in progress):

Entwurf und Realisierung interaktiver modaler Berechnungs- und Optimierverfahren für gekoppelte Struktur-Fluid-Systeme

Design and Implementation of Interactive Modal Calculation and Optimization Methods for Coupled Structure-Fluid Systems

Mit Hilfe einer neuartigen Kopplungsbedingung wird die Bewegungsgleichung eines mit schallabsorbierendem Material ausgekleideten Struktur-Hohlraum-Systems aufgestellt und modal gelöst. Das modale Lösungsverfahren wird zum sogenannten „modalen Korrekturverfahren" erweitert, mit dem es gelingt, den Rechenzeitbedarf für die Berechnung von Modellvarianten und die automatische Optimierung solcher Systeme drastisch zu senken. Die durch die modalen Verfahren bedingten Näherungsfehler werden untersucht, und quantitativ erfaßt. Es wird die Realisierung eines interaktiven Programmsystems beschrieben, welches die Vorteile, die sich durch den Einsatz der entwickelten Methoden ergeben, demonstriert.

8.1.2 Supervised by Professor Terhardt

[P-T13]
Markus Mummert (1997):

Sprachcodierung durch Konturierung eines gehörangepaßten Spektrogramms und ihre Anwendung zur Datenreduktion

Speech Coding by Contourizing an Ear-adapted Spectrogram and its Application to Data Reduction

Konturen als Träger der relevanten Information entsprechen bei der Hörwahrnehmung unter anderem den hörbaren Teiltönen. Die Arbeit behandelt Audiorepräsentationen mit Konturen, die als 'Gratlinien' eines gehörangepaßten Spektrogramms definiert sind. Ausgehend von einer bekannten Repräsentation werden zusätzliche Gratlinien und eine neue Signalrekonstruktion eingeführt. Eine Klassifizierung der Linien trennt tonale und geräuschhafte Signalanteile. Damit werden Sprachcodierungen mit Datenraten bis hinab zu 4 kbit/s realisiert.

[P-T14]
Miriam Noemí Valenzuela (1998):

Untersuchungen und Berechnungsverfahren zur Klangqualität von Klaviertönen

Investigations and Calculation Methods Concerning the Sound Quality of Piano Tones

In dieser Arbeit wurden Modelle und Verfahren entwickelt, mit denen diejenigen Schallsignalparameter ermittelt werden können, die für den spezifischen Klang eines Klaviertons und dessen Qualität charakteristisch sind. Mit Hörversuchen wurde untersucht, worin die hörbare Unähnlichkeit im Klang verschiedener Klaviertöne besteht. Die erarbeiteten Verfahren für die meßtechnische Erfassung der Unterscheidungskriterien ermöglicht die gezielte Verbesserung sowohl elektronischer als auch akustischer Klaviere. Das entwickelte Modell für die Berechnung der Klangqualität von Klaviertönen könnte als automatische Klangqualitätskontrolle sowohl für Einzeltöne wie auch für Instrumente eingesetzt werden.

[P-T15]
Claus von Rücker (1999):

Ein Verfahren zur Tonhöhenanalyse unter Berücksichtigung zeitlich-spektraler Kontrasteffekte

A Pitch Determination Algorithm Allowing for Temporal and Spectral Contrast Effects

Die Arbeit beschreibt ein Verfahren zur Tonhöhenanalyse nichtstationärer Schallsignale. Es zeichnet sich durch die Berücksichtigung derjenigen wesentlichen Gehöreigenschaften aus, die in psychoakustischen Experimenten zur Tonhöhenwahrnehmung beobachtet werden können. Neben den elementaren Eigenschaften der Frequenzanalyse des Gehörs gehören dazu insbesondere zeitlich-spektrale Kontrasteffekte, die von bisherigen Verfahren nicht erfaßt werden. Das Verfahren ist in der Lage, sowohl die Tonhöhen, als auch den Zeitverlauf ihrer Prominenz bei zeitvarianten Schallen nachzubilden.

8.1.3   Supervised by Professor Fastl

[P-F4]
Helmut Spannheimer (1997):

Geräuschminderung im Kraftfahrzeug mit aktiven Resonatoren

Noise Abatement in Cars Utilizing Active Resonators

Zur Geräuschminderung in Kraftfahrzeugen wurde ein aktiver Resonator entwickelt, der in einem Frequenzbereich von 50 Hz bis 200 Hz die Eigenschaften eines Helmholtzresonators bei seiner Resonanzfrequenz nachbildet. Das System wurde mit einem digitalen Regler realisiert, der über ein Mikrofon als Sensor einen Lautsprecher ansteuert. Die Möglichkeiten zur Schalldruckreduktion, die optimale Anordnung und die effektivste Auslegung des Resonators wurden mit einer modalen Schalldruckberechnung bestimmt und an einem Modellhohlraum überprüft. Schließlich wurde das System für verschiedene Anwendungsfälle in Fahrzeuge integriert, im Fahrbetrieb erprobt und die Wirksamkeit subjektiv und objektiv beurteilt.

[P-F5]
Ingeborg Stemplinger (1999):

Beurteilung, Messung und Prognose der Globalen Lautheit von Geräuschimmissionen

Rating, Measurement und Prediction of the Global Loudness of Noise Immissions

Die Analyse der subjektiv empfundenen Globalen Lautheit von Geräuschimmissionen als Maß für die Lärmbelastung und deren meßtechnische Nachbildung bilden die zentrale Fragestellung dieser Arbeit. Die durch die psychoakustischen Experimente gewonnenen Daten lassen sich durch die Messung der Lautheit nach DIN 45631 und anschließende Perzentilwertberechnung gehörrichtig nachbilden. Ein statistisches Verfahren zur Berechnung des Vertrauensbereiches von Lautheitsperzentilen aus der Lautheits-Zeitfunktion ermöglicht erstmals deren qualitätsgesicherte Messung. Durch ein neu entwickeltes Prognoseverfahren kann die Globale Lautheit in Abhängigkeit der Lärmvorbelastung des Gebietes abgeschätzt werden.

8.1.4   Supervised by Professor Ruske

[P-R4]
Franz Wolfertstetter (1996):

Verallgemeinerte stochastische Modellierung für die automatische Spracherkennung

Generalized Stochastic Modeling for Automatic Speech Recognition

Die Arbeit behandelt am Beispiel natürlicher Sprache die Probleme und Lösungen bei der stochastischen Modellierung und Klassifikation von Signalen, die stark von Zufallsprozessen bestimmt sind. Der Schwerpunkt liegt in der Nachbildung des Signalverlaufs mit neuartigen stochastischen Markov-Graphen, die als ein sich verzweigendes und wieder rekombinierendes Pfadsystem mit Zuständen variabler Streuung interpretiert werden können. Zum Training der Modellparameter werden das sogenannte „Maximum-Likelihood"- und diskriminative Verfahren gegenübergestellt. Für die Verarbeitung fließender Sprache wird ein System zum Training und zur Erkennung mit beliebig strukturierten stochastischen Modellen entwickelt.

[P-R5]
Jochen Junkawitsch (2000):

Detektion von Schlüssel-wörtern in fließender Sprache

Keyword-Spotting in Fluent Speech

Der Gegenstand der vorliegenden Arbeit ist die Entwicklung eines neuartigen Verfahrens für Keyword-Spotting, das auf die speziellen Anforderungen derSchlüsselwortdetektion ausgerichtet ist und auf der direkten Optimierung eines Konfidenzmaßes beruht. Es werden vier verschiedene Möglichkeiten zur Definition von Konfidenzmaßen hergeleitet und zwei alternative Suchalgorithmen entwickelt, die eine Optimierung dieser Konfidenzmaße gewährleisten. Ausführliche Experimente bestätigen die Effektivität des vorgestellten Verfahrens, indem die Figure-Of-Merit von 81.5 % auf 87.9 % gesteigert wird.

[P-R6]
Thilo Pfau (2000):

Methoden zur Erhöhung der Robustheit automatischer Spracherkennungssysteme gegenüber Variationen der Sprechgeschwindigkeit

Methods for Improving the Robustness of Automatic Speech Recognition Systems against Variations of Speech Rate

In dieser Arbeit werden verschiedene Ansätze zur Erhöhung der Robustheit automatischer Spracherkennungssysteme gegenüber Variationen der Sprechgeschwindigkeit untersucht. Die Basis bilden Hidden Markov Modelle (HMMs). Im Rahmen einer Reduktion von Intramodell-Variationen wird eine Sprechgeschwindigkeitsnormierung durch Interpolation, ein Verfahren zur Sprechernormierung und zwei Methoden zur Modellierung von Aussprachevarianten vorgestellt. Für eine Anpassung des Systems an unterschiedliche Sprechgeschwindigkeiten wird das Maximum Aposteriori Training zur Schätzung von HMM-Parametern ausführlich diskutiert und ein neuartiges merkmals- und regelbasiertes Verfahren zur Bestimmung der Sprechgeschwindigkeit präsentiert.

8.2 Scientific Publications

Publications *(vertical, left margin)*

Abbreviations

DAGA	Fortschritte der Akustik – Jahrestagung der Deutschen Gesellschaft für Akustik e.V. (DEGA)
DFG	Deutsche Forschungsgemeinschaft
EEAA	East European Acoustical Association
EIS	International Symposium on Engineering of Intelligent Systems
Euro-Noise	European Conference on Noise Control
Eurospeech	European Conference on Speech Communication and Technology
ICASSP	IEEE International Conference on Acoustics, Speech, and Signal Processing
ICIP	IEEE International Conference on Image Processing
ICSLP	IEEE International Conference on Spoken Language Processing
IEEE	Institute of Electrical and Electronics Engineers, Inc.
IFAC	International Federation of Automatic Control
ITG	Informationstechnische Gesellschaft
JASA	Journal of the Acoustical Society of America
SPECOM	International Workshop "Speech and Computer"
VDE	Verband der Elektrotechnik Elektronik Informationstechnik e.V.

1997

[97bau] U. Baumann, I. Stemplinger, B. Arnold, K. Schorn: Bezugskurven für die Hörflachenskalierung in der klinischen Anwendung. Laryngo-Rhino-Otologie, Heft 8, 1997, S. 458-465.

[97bub1] U. Bub, J. Köhler, B. Imperl: In-Service Adaptation of Multilingual Hidden-Markov-Models. Tagungsband ICASSP 97 (München, 21.-24.4.1997), Vol. 2, S. 1451-1454.

[97cha] J. Chalupper, W. Schmid: Akzentuierung und Ausgeprägtheit von Spektraltonhöhen bei harmonischen Komplexen Tönen. Tagungsband DAGA 97 (Kiel, 1997), S. 357-358.

[97fas1] H. Fastl: Gehörgerechte Geräuschbeurteilung. Tagungsband DAGA 97 (Kiel, 1997), S. 57-64.

[97fas2] H. Fastl, W. Schmid: Comparison of Loudness Analysis Systems. Tagungsband inter-noise 97 (Budapest, Ungarn, 1997), Band II, S. 981-986.

[97fas3] H. Fastl: The Psychoacoustics of Sound-Quality Evaluation. Acustica/Acta Acustica, Band 83 (1997), S. 754-764.

[97fas4] H. Fastl: Psychoacoustic Noise Evaluation. Tagungsband Acoustics - High Tatras'97 (Vysoké Tatry, Slowakei, 1997) S. 21-26.

[97got] G. Gottschling, H. Fastl: Akustische Simulation von 6-Sektionen-Fahrzeugen des Transrapid. Tagungsband DAGA 97 (Kiel, 1997), S. 254-255.

[97hau1] M. Haubner, C. Krapichler, A. Lösch et al.: Virtual reality in medicine-computer graphics and interaction techniques. IEEE Transactions on Information Technology in Biomedicine, März 1997, 1(1) S. 61-72.

[97hoj] E. Hojan, I. Stemplinger, H. Fastl: Zur Verständlichkeit deutscher Sprache im Störgeräusch nach Fastl durch polnische Hörer mit verschiedenen Deutschkenntnissen. Audiologische Akustik - Audiological Acoustics, Band 36 (1997), Heft 1, S. 32-37.

[97hol] M. Holzapfel, G. Ruske, H. Höge: Failure Simulation for a Phoneme HMM Based Keyword Spotter. Tagungsband ICASSP 97 (München, 1997), S. 911-914.

[97jae]   K. Jäger, H. Fastl, F. Schöpf, G. Gottschling, U. Möhler: Wahrnehmung von Pegeldifferenzen bei Vorbeifahrten von Güterzügen. Tagungsband DAGA 97 (Kiel, 1997), S. 228-229.

[97jun]   J. Junkawitsch, G. Ruske, H. Höge: Efficient Methods for Detecting Keywords in Continuous Speech. Tagungsband Eurospeech '97, (Rhodos, Griechenland, 1997), S. 259-262.

[97kra]   C. Krapichler, M. Haubner, A. Lösch, M. Lang, K.-H. Englmeier: Human-Machine Interface for a VR-based Medical Imaging Environment. Tagungsband SPIE Medical Imaging '97 (Newport Beach, USA, 1997), S. 527-534.

[97kra2] C. Krapichler, M. Haubner, A. Lösch, K.-H. Englmeier: A human-machine interface for medical image analysis and visualization in virtual environments. Tagungsband ICASSP 97 (München, 21.-24.4.1997), Vol. 4, S. 2613-2616.

[97kuw]   S. Kuwano, S. Namba, H. Fastl, A. Schick: Evaluation of the Impression of Danger Signals - Comparison between Japanese and German Subjects. In A. Schick (Hrsg.): 7. Oldenburger Symposium, BIS Oldenburg, 1997, S. 115-118.

[97len1] H. Lenz, D. Obradovic: Effectivity of the Stabilization of Higher Period Orbits of Chaotic Maps. Proc. of Int. Conf. on Control of Oscillations and Chaos (COC'97, St. Petersburg, Rußland, 27.-29.08.1997), Vol. 1, S. 152-155.

[97len2] H. Lenz, R. Berstecher: Sliding-Mode Control of Chaotic Pendulum: Stabilization and Targeting of an Unstable Periodic Orbit. Proc. of Int. Conf. on Control of Oscillations and Chaos (COC'97, St. Petersburg, Rußland, 27.-29.08.1997), Vol. 3, S. 586-589.

[97len3] H. Lenz, D. Obradovic: Global Control of Lorenz Chaos. Proc. of IEEE Conf. on Decision and Control (CDC 97, San Diego, USA, 10.-12.12.1997), Vol. 2, S. 1486-1487.

[97len4] H. Lenz, D. Obradovic: Robust Control of the Chaotic Lorenz System. Int. Journal of Bifurcation and Chaos, 1997, Vol. 7, Nr. 12, S. 2847-2854.

[97mor1] P. Morguet, M. Lang: Feature Extraction Methods for Consistent Spatio-Temporal Image Sequence Classification Using Hidden-Markov-Models. Tagungsband ICASSP 97 (München, 1997), S. 2893-2896.

[97mor2] P. Morguet, M. Lang: A Universal HMM-Based Approach to Image Sequence Classification. Tagungsband ICIP 97 (Santa Barbara, USA, 1997), S. III/146-III/149.

[97mue]   J. Müller, H. Stahl: The Semantic Structure in Comparison with Other Semantic Representations. Tagungsband SPECOM 97 (Cluj-Napoca, Rumänien, 1997), S. 7-12.

[97obr1] D. Obradovic, H. Lenz: When is OGY Control more than just Pole Placement. Int. Journal of Bifurcation and Chaos, 1997, Vol. 7, Nr. 3, S. 691-699.

[97pfa]   T. Pfau, M. Beham, W. Reichl, G. Ruske: Creating Large Subword Units for Speech Recognition. Tagungsband Eurospeech '97, (Rhodos, Griechenland, 1997), S. 1191-1194.

[97rue]   C. von Rücker: Berechnung von Erregungsverteilungen aus FTT-Spektren. Tagungsband DAGA 97 (Kiel, 1997), S. 484-485.

[97sch1] R. Schirmacher, P. Maier, H. Fastl, J. Scheuren: Aktive Geräuschminderung an einem Lüftungskanal, S. 199-200. Tagungsband DAGA 97 (Kiel, 1997).

[97sch2] W. Schmid, J. Chalupper: Die Ausgeprägtheit der Tonhöhe als psychoakustisches Kriterium zur Qualitätsbeurteilung elektroakustischer Komponenten - warum ist der Phasengang des Differenztons 2. Ordnung so wichtig? 19. Tonmeistertagung - International Convention on Sound Design (Karlsruhe, 1996), Bildungswerk des Verbandes Deutscher Tonmeister (VDT), Berlin, 1997, S. 861-874.

[97sch3]   W. Schmid: Zur Ausgeprägtheit der Tonhöhe gedrosselter und amplitudenmodulierter Sinustöne. Tagungsband DAGA 97 (Kiel, 1997), S. 355-356.

[97sta]   H. Stahl, J. Müller, M. Lang: Controlling Limited-Domain Applications by Probabilistic Semantic Decoding of Natural Speech. Tagungsband ICASSP 97 (München, 1997), S. 1163-1166.

[97ste1]   I. Stemplinger, M. Schiele, B. Meglic, H. Fastl: Einsilberverständlichkeit in unterschiedlichen Störgeräuschen für Deutsch, Ungarisch und Slowenisch. Tagungsband DAGA 97 (Kiel, 1997), S. 77-78.

[97ste2]   I. Stemplinger: Beurteilung der Globalen Lautheit bei Kombination von Verkehrsgeräuschen mit simulierten Industriegeräuschen. Tagungsband DAGA 97 (Kiel, 1997), S. 353-354.

[97ste3]   I. Stemplinger, G. Gottschling: Auswirkungen der Bündelung von Verkehrswegen auf die Beurteilung der Globalen Lautheit. Tagungsband DAGA 97 (Kiel, 1997), S. 401-402.

[97ste4]   I. Stemplinger, H. Fastl: Accuracy of Loudness Percentile Versus Measurement Time. Tagungsband inter-noise 97 (Budapest, Ungarn, 1997), Band III, S. 1347-1350.

[97ter]   E. Terhardt: Lineares Modell der peripheren Schallübertragung im Gehör. Tagungsband DAGA 97 (Kiel, 1997), S. 367-368.

[97val]   M.N. Valenzuela: Extraktion gehörrelevanter Schallsignalparameter aus Flügelklängen. Tagungsband DAGA 97 (Kiel, 1997), S. 321-322.

[97wag1]   C. Wagner: Successive deceleration in Boltzmann-like traffic equations. Physical Review E, 1997, Vol. 55, S. 6969.

[97wag2]   C. Wagner: A Navier-Stokes-like traffic model. Physica A., 1997, Vol. 245, S. 124.

[97win1]   H.-J. Winkler, M. Lang: On-line Symbol Segmentation and Recognition in Handwritten Mathematical Expressions. Tagungsband ICASSP 97 (München, 1997), S. 3377-3380.

[97win2]   H.-J. Winkler, M. Lang: Symbol Segmentation and Recognition for Understanding Handwritten Mathematical Expressions. In: A.C. Downton, S. Impedovo (Hrsg.): „Progress in Handwriting Recognition", Tagungsband 5th International Workshop on Frontiers in Handwriting Recognition IWFHR-5 (Essex, England, 1996), World Scientific, 1997, S. 407-412.

[97wol]   H. Wollherr, S. Goossens, K.D. Ruth, H. Fastl: Horizontal / Vertikal differenziertes Bündelungsmaß. Tagungsband DAGA 97 (Kiel, 1997), S.123-124.

1998

[98bub1]   U. Bub, H. Höge: Boosting Long-Term Adaptation of Hidden-Markov-Models: Incremental Splitting of Probability Density Functions. Proc. of ICASSP 98 (Seattle, USA, 12.-15.5.1998), Vol. 1, S. 429-432.

[98cha1]   J. Chalupper, K. Spasokukotskij, I. Stemplinger, H. Fastl: Ein Zweisilber-Sprachtest für Ukrainisch. Tagungsband DAGA 98 (Zürich, Schweiz, 1998), S. 310-311.

[98eng1]   K.-H. Englmeier, C. Krapichler, M. Haubner et al.: Virtual reality and multimedia human-computer interaction in medicine. Proc. of IEEE Workshop on Multimedia Signal Processing (MMSP '1998, Los Angeles, USA), S. 193-202.

[98eng2]   K.-H. Englmeier, C. Krapichler et al.: A New Hybrid Renderer for Virtual Bronchoscopy. J.D. Westwood et al. (Eds.): Studies in Health Technology and Informatics: Medicine Meets Virtual Reality, IOS Press Amsterdam, 1999, Band 62, S. 109-115.

<div style="writing-mode: vertical-rl">Publications</div>

[98fal1]   R. Faltlhauser, G. Ruske: Automatische Topologiegenerierung für kontinuierliche Hidden-Markov-Modelle. Tagungsband ITG-Fachtagung „Sprachkommunikation" (Dresden, 1998), S.33-36.

[98fas1]   H. Fastl, Th. Filippou, W. Schmid, S. Kuwano, S. Namba: Psychoakustische Beurteilung der Lautheit von Geräuschimmissionen verschiedener Verkehrsträger. Tagungsband DAGA 98 (Zürich, Schweiz, 1998), S. 70-71.

[98fas2]   H. Fastl, H. Oberdanner, W. Schmid, I. Stemplinger, I. Hochmair-Desoyer, E. Hochmair: Zum Sprachverständnis von Cochlea-Implantat-Patienten bei Störgeräuschen. Tagungsband DAGA 98 (Zürich, Schweiz, 1998), S. 358-359.

[98fas3]   H. Fastl, W. Schmid: Vergleich von Lautheits-Zeitmustern verschiedener Lautheits-Analysesysteme. Tagungsband DAGA 98 (Zürich, Schweiz, 1998), S. 466-467.

[98fas4]   H. Fastl: Psychoacoustics and Sound Quality Metrics. Proceedings of the 1998 Sound Quality Symposium (Ypsilanti, Michigan, USA, 1998) S. 3-10.

[98fas5]   H. Fastl: Intelligibility of Speech in Noise by Cochlea-Implant Patients. Proceedings Fall Meeting, Acoustical Society of Japan (Yonezawa, Japan, 1998) Vol. 1, S. 359-360.

[98fas6]   H. Fastl: Pitch Strength and Frequency Discrimination for Noise Bands or Complex Tones. "Psychophysical and Physiological Advances in Hearing", Whurr Publishers Ltd. London, England, S. 238-245.

[98fas7]   H. Fastl, J. Scheuren (Eds.): euro-noise 98, Designing for Silence. Proceedings of Euro-Noise 98 (München 1998).

[98got1]   G. Gottschling, W. Schmid, H. Fastl: Vergleich psychoakustischer Methoden zur Skalierung der Lautstärke: I. Grundlagen. Tagungsband DAGA 98 (Zürich, Schweiz, 1998), S. 476-477.

[98got2]   G. Gottschling, H. Fastl: Prognose der globalen Lautheit von Geräuschimmisionen anhand der Lautheit von Einzelereignissen. Tagungsband DAGA 98 (Zürich, Schweiz, 1998), S. 478-479.

[98hau1]   M. Haubner, C. Krapichler et al.: Neue Techniken der virtuellen Realität für den Einsatz in der Medizin. EDV-Jahrbuch 1998, Hüthig Fachverlage, Heidelberg, 1998, Computer in Zahnarztpraxis und Dentallabor, S. 90-93.

[98hor1]   T. Horn: Image Processing of Speech with Auditory Magnitude Spectrograms. Acustica/Acta Acustica, Band 84 (1998), S. 175-177.

[98hut1]   Ch. Huth, W. Schmid: Psychoakustische Untersuchungen zur Ausgeprägtheit der Tonhöhe bei musikalischen Klängen mit und ohne Vibrato. Tagungsband DAGA 98 (Zürich, Schweiz 1998), S. 448-449.

[98koe1]   J. Köhler: Language Adaptation of Multilingual Phoneme Models for Vocabulary Independent Speech Recognition Tasks. Proc. of ICASSP 98 (Seattle, USA, 12.-15.5.1998), Vol. 1, S. 417-420.

[98kra1]   C. Krapichler, M. Haubner et al.: VR interaction techniques for medical imaging applications. Computer Methods and Programs in Biomedicine, April 1998, 56(1), S. 65-74.

[98len1]   H. Lenz, R. Berstecher, M. Lang: Adaptive Sliding-Mode Control of the Absolute Gain. Proc. of 4th IFAC Nonlinear Control System Design Symposium (NOLCOS 98, Enschede, Niederlande, 01.-03.07.1998), Vol. 3, S. 667-672.

[98mas1]   D. Mass, M. Lang: Effects of Porous Absorbing Material on the Coupling of Mode Shapes via the Modal Damping Matrix of a Fluid. Tagungsband Euro Noise 98, Vol. 2 (München, 1998), S. 883-888.

[98mor1]   P. Morguet, M. Lang: An Integral Stochastic Approach to Image Sequence Segmentation and Classification. Tagungsband ICASSP 98 (Seattle, USA, 1998), S. 2705-2708.

[98mor2]   P. Morguet, M. Lang: Spotting Dynamic Hand Gestures in Video Image Sequences Using Hidden Markov Models. Tagungsband ICIP 98 (Chicago, USA, 1998), Vol. 3, S. 193-197.

[98mor3]   P. Morguet, M. Lang: Dynamische Gesten und indirekte Manipulation als Grundlage für eine intuitive Mensch-Maschine-Kommunikation. Tagungsband der ITG-Fachtagung: Technik für den Menschen - Gestaltung und Einsatz benutzerfreundlicher Produkte, 1988 (Eichstätt, 1998), S. 131-139.

[98mue1]   J. Müller, H. Stahl: Speech Understanding and Speech Translation in Various Domains by Maximum a-posteriori Semantic Decoding. Speech Understanding and Speech Translation in Various Domains by Maximum a-posteriori Semantic Decoding. Tagungsband EIS 98 (La Laguna, Spanien, 1998), Band 2 „Neural Networks", S. 256-267.

[98mue2]   J. Müller, C. Krapichler, L.S. Nguyen, K.H. Englmeier, M. Lang: Speech Interaction in Virtual Reality. Tagungsband ICASSP 98 (Seattle, USA, 1998), S. 3757-3760.

[98mue3]   J. Müller, M. Lang: Verstehen natürlicher Sprache für den Mensch-Maschine-Dialog. In H. Bubb (Hrsg.): „Bay-FORERGO - Plädoyer für einen bayerischen Forschungsverbund Ergonomie", Herbert Utz Verlag, München, 1998, S. 91-113.

[98mum1]   M. Mummert: Sprachcodierung durch Konturierung eines gehörangepaßten Spektogramms und ihre Anwendung zur Datenreduktion. Düsseldorf: VDI Verlag, 1998, 240 S., 42 Abb., 11 Tab. (VDI Reihe 10: Informatik/Kommunikationstechnik, Nr. 522)

[98pfa1]   T. Pfau, G. Ruske: Estimating the Speaking Rate by Vowel Detection. Tagungsband ICASSP 98 (Seattle, USA, 1998), S. 945-948.

[98pfa2]   T. Pfau, G. Ruske: Creating Hidden Markov Models for Fast Speech. Tagungsband ICSLP 98 (Sydney, Australien 1998), Paper no. 255.

[98rue1]   C. von Rücker: Spektraltonhöhenanalyse unter Berücksichtigung von Akzentuierung. Tagungsband DAGA 98 (Zürich, Schweiz, 1998), S. 498-499.

[98rus1]   G. Ruske, R. Faltlhauser, T. Pfau: Extended Linear Discriminant Analysis (ELDA) for Speech Recognition. Tagungsband ICSLP 98 (Sydney, Australien 1998), Paper no. 100.

[98sch1]   W. Schmid, J. Chalupper: Spektraltonhöhen Komplexer Töne: Psychoakustische Experimente und Berechnung der Ausgeprägtheit der Tonhöhe. Tagungsband DAGA 98 (Zürich, Schweiz, 1998), S. 480-481.

[98sch2]   W. Schmid: Die akzentuierende Wirkung auf Spektraltonhöhen Komplexer Töne. Tagungsband DAGA 98 (Zürich, Schweiz, 1998), S. 468-469.

[98sch3]   W. Schmid: Zur Ausgeprägtheit der Tonhöhe von Rauschen mit zeitvarianter Bandbegrenzung. Tagungsband DAGA 98 (Zürich, Schweiz, 1998), S. 470-471.

[98sch4]   A. Schreyer, G. Suda, G. Maderlechner: Font Style Detection Using Textons. Proc. of Document Analysis Systems Workshop (DAS 98, Nagano, Japan, 4.-6.11.1998), S. 99-108.

[98sch5]   A. Schreyer, G. Suda, G. Maderlechner: The Idea of Attention-Based Document Analysis. Proc. of Document Analysis Systems Workshop (DAS 98, Nagano, Japan, 4.-6.11.1998), S. 214-217.

[98ste1]   I. Stemplinger, Th. Filippou: Psychoakustische Untersuchungen zur Lautheit und zur Lästigkeit von Tennislärm. Tagungsband DAGA 98 (Zürich, Schweiz, 1998), S. 66-67.

[98ter1]   E. Terhardt: Akustische Kommunikation. Springer, Berlin/Heidelberg, 1998, 505 S., 221 Abb. 15 Tab., 1 DDD Audio-CD.

[98val1]   M.N. Valenzuela: Bewertung der Gehörrelevanz von Partialtonzeitstrukturen in Klaviertönen. Tagungsband DAGA 98 (Zürich, Schweiz, 1998), S. 632-633.

[98val2]   M.N. Valenzuela: Untersuchungen und Berechnungsverfahren zur Klangqualität von Klaviertönen. München: Herbert Utz Verlag Wissenschaft, 1998, 154 S.

[98val3]   J. Vallen, G. Ruske: Acoustic Emission Source Discrimination by Analyzing Short-term Frequency Spectra. Intern. Symposium on Acoustic Emission: Standards and Technology Update, American Standards for Testing Materials ASTM STP 1353 (Plantation, Florida, USA, 1998), Paper ID #4604.

[98wag1]   C. Wagner: Traffic flow models considering an internal degree of freedom. Journal of Statistical Physics, 1998, Vol. 90, S. 1251.

[98wid1]   U. Widmann, R. Lippold, H. Fastl: Ein Computerprogramm zur Simulation der Nachverdeckung für Anwendungen in Akustischen Meßsystemen. Tagungsband DAGA 98 (Zürich, Schweiz, 1998), S. 96-97.

[98wid2]   U. Widmann, R. Lippold, H. Fastl: A Computer Program Simulating Post-Masking for Applications in Sound Analysing Systems. Proceedings of NOISE-CON '98 (Ypsilanti, Michigan, USA, 1998), S. 451-456.

[98wid3]   U. Widmann, H. Fastl: Calculating Roughness using Time-Varying Specific Loudness Spectra. Proceedings of the 1998 Sound Quality Symposium (Ypsilanti, Michigan, USA, 1998), S. 55-60.

[98wit1]   L. Witta, M. Lang: A Finite Element for Bulk Reacting Porous Sound Absorbers. Tagungsband Euro Noise 98, Vol. 2 (München, 1998), S. 943-948.

1999

[99cha1]   Chalupper, J.: Loudness fluctuation and temporal masking in normal and hearing-impaired listeners. JASA Vol. 105, No. 2, Pt. 2 (1999), S. 1023.

[99fal1]   Faltlhauser, R.; Pfau, T.; Ruske, G.: Hidden Markov Models for Fast Speech by Optimized Clustering. Proc. of the 6th European Conf. on Speech Communication and Technology, Budapest, Ungarn, 5.-9.9.1999. Hrsg.: G. Olaszy et al. Bonn: ESCA, 1999, S. 407-410. (EUROSPEECH'99; Vol. 1)

[99fas1]   Fastl, H.: Analysis Systems for Psychoacoustic Magnitudes. 8th Oldenburg Symposium on Psychological Acoustics, Bad Zwischenahn, 29.8.-1.9.1999. Eds.: A. Schick, M. Meis, C. Reckhardt. Oldenburg: BIS-Verlag, 2000, S. 85-101. (Contributions to Psychological Acoustics)

[99fas2]   Fastl, H.: Psychoacoustic evaluation of noise emissions. JASA Vol. 105, No. 2, Pt. 2 (1999), S. 1082-1083.

[99fil1]   Filippou, Th.: Comparison of Subjective and Physical Evaluation of Tennis-Noise. Proceedings of the 28th Int. Congress on Noise Control Engineering, Fort Lauderdale, Florida, USA, 6.-8.12.1999. Eds.: J. Cushieri, St. Glegg, Y. Yong. Washington DC: Inst. of Noise Control Engineering, 1999. S. 1881-1886. (INTER-NOISE 99; Vol. 3)

[99fil2]   Filippou, Th.; Fastl, H.: Estimates of the instantaneous versus overall loudness of noise emissions. JASA Vol. 105, No. 2, Pt. 2 (1999), S. 1298.

[99got1]   Gottschling, G.: On the relations of instantaneous and overall loudness. Acustica/acta acustica, 1999, Vol. 85, Nr. 3, S. 427-429.

[99hir1]   Hirsch, H.S.; Wiegrebe, L.; Patterson, R.D.; Fastl, H.: Temporal dynamics of pitch strength; frequency effects and auditory modelling. British J. Audiology, 33.2 (1999), S. 111.

[99hut1]   Huth, Ch.; Fastl, H.; Hölzl, G.; Widmann, U.: Sound quality design for high-speed trains' indoor noise: Psychoacoustic evaluation of tonal components. JASA Vol. 105, No. 2, Pt. 2 (1999), S. 1280.

[99kra1]   Krapichler, Ch.: Eine neue Mensch-Maschine-Schnittstelle für die Analyse medizinischer 3D-Bilddaten in einer virtuellen Umgebung. München: Herbert Utz Verlag Wissenschaft, 1999, 130 S.

[99kra2]   Krapichler, Ch.; Haubner, M. et al.: Physicians in virtual environments: Multimodal human-computer interaction. Interacting with Computers, 1999, Vol. 11, 4, S. 427-452.

[99kuw1]   Kuwano, S.; Namba, S.; Florentine, M.; Zheng, D.R.; Fastl, H.; Schick, A.: A cross-cultural study of the factors of sound quality of environmental noise. JASA Vol. 105, No. 2, Pt. 2 (1999), S. 1081.

[99kuw2]   Kuwano, S; Namba, S.; Florentine, M.; Zheng, D.R.; Fastl, H.; Schick, A.; Weber, R.; Höge, H.: A cross-cultural study of the factors of sound quality of environmental noise. ASA/DEGA/EAA Conference FORUM ACUSTICUM, Berlin, 14.-19.3.1999. Hrsg.: DEGA, Oldenburg, als CD-ROM (Collected Papers from the Joint Meeting „Berlin 99").

[99kuw3]   Kuwano, S.; Fastl, H.; Namba, S.: Loudness, Annoyance and Unpleasantness of Amplitude Modulated Sounds. Proceedings of the 28th Int. Congress on Noise Control Engineering, Fort Lauderdale, Florida, USA, 6-8.12.1999. Eds.: J. Cushieri, St. Glegg, Y. Yong. Washington DC: Inst. of Noise Control Engineering, 1999. S. 1195-1200. (INTER-NOISE 99; Vol. 2)

[99len1]   Lenz, H.; Obradovic, D.: Stabilizing Higher Periodic Orbits of Chaotic Discrete-Time Maps. Int. Journal of Bifurcation and Chaos, 1999, Vol. 9, Nr. 1, S. 251-266.

[99len2]   Lenz, H.; Sollacher, R.: Nonlinear Speed-Control for a Continuum Theory of Traffic Flow. Proc. of the 14th World Congress of Int. Federation of Automatic Control, IFAC, Peking, China, 05.-09.01.1999, Vol. Q, S. 67-72.

[99len3]   Lenz, H.; Wagner, C.; Sollacher, R.: Multi-Anticipative Car-Following Model. European Physical Journal B, Vol. 7, S. 331-335.

[99mad1]   Maderlechner, G.; Schreyer, A.; Suda, P.: Information Extraction from Document Images Using Attention Based Layout Segmentation. Proc. of Document Layout Interpretation and its Applications Workshop (DLIA 99, Bangalore, Indien, 18.09.1999) Online-Proceedings, Beitrag Ib: http://www.science.uva.nl/events/dlia99/.

[99mor1]   Morguet, P.; Lang, M.: Comparison of Approaches to Continuous Hand Gesture Recognition for a Visual Dialog System. Proceedings ICASSP 99 (Phoenix, Arizona, USA, 15.-19.3.1999), IEEE, Vol. 6, S. 3549-3552.

[99mue1]   Müller, J.; Stahl, H.: Speech Understanding and Speech Translation by Maximum a-Posteriori Semantic Decoding. Artificial Intelligence in Engineering, Vol. 13 (1999), No. 4, Elsevier-Verlag, S. 373-384.

[99pfa1]   Pfau, T.; Faltlhauser, R.; Ruske, G.: Speaker Normalization and Pronunciation Variant Modeling: Helpful Methods for Improving Recognition of Fast Speech. Proc. of the 6th European Conf. on Speech Communication and Technology, Budapest, Ungarn, 5.-9.9.1999. Hrsg.: G. Olaszy et al. Bonn: ESCA, 1999, S.299-302. (EUROSPEECH '99; Vol. 1)

[99rus1]   Ruske, G.; Lee, K.Y.: Speech recognition and enhancement by a nonstationary AR HMM with gain adaptation under unknown noise. Proceedings ICASSP 99 (Phoenix, Arizona, USA, 15.-19.3.1999). IEEE, 1999, Vol. 1, S. 441-444.

[99sch1]   Schmid, W.: Zur Ausgeprägtheit der Tonhöhe: Konzepte und neuere Ergebnisse psychoakustischer Experimente. Tagungsband der 20. Tonmeistertagung - International Convention on Sound Design, Karlsruhe, 20.-23.11.1998. Hrsg.: Bildungswerk des Verbandes Deutscher Tonmeister (VDT), Bergisch-

Gladbach. München: Sauer-Verlag, 1999, S. 1051-1066.

[99sch2]   Schreyer, A.; Suda, P.; Maderlechner, G.: A Formal Approach To Textons and its Application to Font Style Detection. Lee, S.-W.; Nakano, Y. (Eds.): Document Analysis Systems; Theory and Practice, Third IAPR Workshop, DAS'98, Nagano, Japan, 04.-06.11.1998, Selected Papers, Springer-Verlag, 1999, Lecture Notes in Computer Science, Vol. 1655.

[99ste1]   Stemplinger, I.: Beurteilung, Messung und Prognose der Globalen Lautheit von Geräuschimmissionen. München: Herbert Utz Verlag Wissenschaft, 1999, 112 S.

[99wie1]   Wiegrebe, L.; Hirsch, H.S.; Patterson, R.D.; Fastl, H.; : Time constants of pitch processing arising from auditory filtering. JASA Vol. 105, No. 2, Pt. 2 (1999), S. 1234.

[99zwi1]   Zwicker, E.†; Fastl, H.: Psychoacoustics Facts and Models. 2nd updated edition. Berlin/Heidelberg: Springer-Verlag, 1999, 416 S., 289 Abb.

2000

[00bau1]   Baumann, U.: Identification and segregation of multiple auditory objects. In: Auditory Worlds: Sensory Analysis and Perception in Animals and Man (siehe [00man1]), S. 274-278.

[00cha1]   Chalupper, J.; Fastl, H.: Simulation of Hearing Impairment Based on the Fourier Time Transformation. Proc. ICASSP'2000, Istanbul, Türkei, 5.-9.6.2000. IEEE, Vol. 2, S. 857-860.

[00cha2]   Chalupper, J.: Modellierung der Lautstärkeschwankung für Normal- und Schwerhörige. Tagungsband DAGA 2000, 26. Jahrestagung Akustik, Oldenburg, 20.-24.3.2000, S. 254-255.

[00cha3]   Chalupper, J.: Aural Exciter and Loudness Maximizer: What's Psychoacoustic about "Psychoacoustic Processors"? 109th AES Convention in Los Angeles, USA, 22.-25.9.2000, Audio Engineering Society Preprint, Nr. 5208.

[00cha4]   Chalupper, J.; Wimmer, H.; Schmid, W.: Sprachübertragung mit Induktionsschleifen (Speech Transmission by Induction Loops: Physical and Psychoacoustic Measurements). Tagungsband der 21. Tonmeistertagung – VDT International Audio Convention, Hannover, 24.-27. November 2000, Verband Deutscher Tonmeister.

[00fal1]   Faltlhauser, R.; Pfau, T.; Ruske, G.: On-line Speaking Rate Estimation using Gaussian Mixture Models. Proc. ICASSP'2000, Istanbul, Türkei, 5.-9.6.2000. IEEE, Vol. 3, S. 1355-1358.

[00fal2]   Faltlhauser, R.; Pfau, T.; Ruske, G.: On-line Speaking Rate Estimation Using a GMM/NN Approach. Tagungsband ITG-Fachtagung "Sprachkommunikation", Ilmenau, 9.-12.10.2000, VDE Verlag, Berlin, Offenbach 2000, Hrsg.: ITG, S. 101-105.

[00fal3]   Faltlhauser, R.; Pfau, T.; Ruske, G.: On the Use of Speaking Rate as a Generalized Feature to Improve Decision Trees. Proc. of ICSLP 2000, Peking, China, 16.-20.10.2000, China Military Friendship Publish, Vol. 1, S. 317-320.

[00fas1]   Fastl, H.: The presentation of stimuli in psychoacoustics. In: Auditory Worlds: Sensory Analysis and Perception in Animals and Man (siehe [00man1]), S. 246-247.

[00fas2]   Fastl, H.: Masking effects. In: Auditory Worlds: Sensory Analysis and Perception in Animals and Man (siehe [00man1]), S.247-251.

[00fas3]   Fastl, H.: Basic hearing sensations. In: Auditory Worlds: Sensory Analysis and Perception in Animals and Man (siehe [00man1]), S. 251-258.

[00fas4]   Fastl, H.: Loudness and noise evaluation. In: Auditory Worlds: Sensory Analysis and Perception in Animals and Man (siehe [00man1]), S. 258-267.

[00fas5]   Fastl, H. for Zwicker, E.†: Psychoacoustically-based models of the inner ear. In: Auditory Worlds: Sensory Analysis and Perception in Animals and Man (siehe [00man1]), S. 76-80.

[00fas6]   Fastl, H. for Zwicker, E.†: Otoacoustic Emissions in human test subjects. In: Auditory Worlds: Sensory Analysis and Perception in Animals and Man (siehe [00man1]), S. 120-127.

[00fas7]   Fastl, H.: Railway Bonus and Aircraft Malus: Subjective and Physical Evaluation. Proc. of the 5th Int. Symposium Transport Noise and Vibration, TRANSPORT NOISE 2000, 6.-8.6.2000, St. Petersburg, Russland. CD-ROM publ. by EEAA.

[00fas8]   Fastl, H.: Sound Measurements Based on Features of the Human Hearing System. Tagungsband DAGA 2000, 26. Jahrestagung Akustik, Oldenburg, 20.-24.3.2000, S. 90-91.

[00fas9]   Fastl, H.; Patsouras, D.: Pure tone plus bandlimited noise as Zwicker-tone-exciter. Tagungsband DAGA 2000, 26. Jahrestagung Akustik, Oldenburg, 20.-24.3.2000, S. 306-307.

[00fas10]   Fastl, H.: Noise Evaluation Based on Hearing Sensations. Proc. of the 7th Western Pacific Regional Acoustics Conference, WESTPRAC VII, 3.-5.10.2000, Kumamoto, Japan. The Acoustical Society of Japan, Tokyo, Vol. 1, S. 33-41.

[00fas11]   Fastl, H.: Sound Quality of Electric Razors – Effects of Loudness. Proc. INTER-NOISE'2000, 28.-30.8.2000, Nizza, Frankreich. CD-ROM.

[00hai1]   Haiber, U.; Mangold, H.; Pfau, T.; Regel-Brietzmann, P.; Ruske, G.; Schleß, V.: Robust Recognition of Spontaneous Speech. W. Wahlster (Ed.): Verbmobil: Foundations of Speech-to-Speech Translation. Springer-Verlag Berlin, Heidelberg, 2000, S. 46-62.

[00hof1]   Hofmann, M.; Lang, M.: Belief Networks for a Syntactic and Semantic Analysis of Spoken Utterances for Speech Understanding. Proc. ICSLP 2000, Peking, China, 16.-20.10.2000, China Military Friendship Publish, Vol. II, S. 875-878.

[00hun1]   Hunsinger J.; Lang, M.: A Speech Understanding Module for a Multimodal Mathematical Formula Editor. Proceedings ICASSP 2000, Istanbul, Türkei, 5.-9.6.2000. IEEE, Vol. 4, S. 2413-2416.

[00hun2]   Hunsinger, J.; Lang, M.: A Single-Stage Top-Down Probabilistic Approach towards Understanding Spoken and Handwritten Mathematical Formulas. Proc. ICSLP 2000, Peking, China, 16.-20.10.2000, China Military Friendship Publish, Vol. 4, S. 386-389.

[00jun1]   Junkawitsch, J.: Detektion von Schlüsselwörtern in fließender Sprache. Aachen: Shaker Verlag, 2000, 149 S. (Berichte aus der Informatik)

[00koe1]   Köhler, J.: Erstellung einer statisch modellierten multilingualen Lautbibliothek für die Spracherkennung. Aachen: Shaker Verlag, 2000, 158 S. (Berichte aus der Informatik)

[00kru1]   Krump, G.: Der akustische Nachton: Beschreibung und Funktionsschema. Beiträge zur Vibro- und Psychoakustik 3/00. Neubiberg: Universität der Bundeswehr München, 2000, Hrsg.: H. Fleischer, H. Fastl, 101 S.

[00kuw1]   Kuwano, S.; Namba, S.; Schick, A.; Höge, H.; Fastl, H.; Filippou, Th.; Florentine, M.; Muesch, H.: The Timbre and Annoyance of Auditory Warning Signals in Different Countries. Proc. INTER-NOISE'2000, 28.-30.8.2000, Nizza, Frankreich. CD-ROM.

[00lan1]   Lang, M.; Reichwald, R. (Hrsg.): Anwenderfreundliche Kommunikationssysteme. Tagungsband vom Kongress vom 17.-19.6.1999 in München. Heidelberg: Hüthig Verlag, 2000, 376 S. (Forum Telekommunikation des Münchner Kreises, Bd. 17)

[00mad1] Maderlechner, G.; Schreyer, A.; Suda, P.: Extraction of Relevant Information from Document Images Using Measures of Visual Attention. Proc. of the 15th Int. Conf. on Pattern Recognition, ICPR'2000, Barcelona, Spanien, 03.-08.09.2000, Int. Assoc. of Pattern Recognition IAPR, Surrey, UK, 2000, Vol. 4, S. 385-388.

[00man1] Manley, G.A.; Fastl, H.; Kössl, M.; Oeckinghaus, H.; Klump, G. (Eds.): DFG: Auditory Worlds: Sensory Analysis and Perception in Animals and Man. Final report of the collaborative research centre 204, „Nachrichtenaufnahme und -verarbeitung im Hörsystem von Vertebraten (Munich)", 1983-1997. Weinheim: Wiley-VCH Verlag, 2000, 359 S.

[00nie1] Niedermaier, B.: "Eyes free - Hands free" oder "Zeit der Stille". Ein Demonstrator zur multimodalen Bedienung im Automobil. DGLR-Bericht 2000-02 der 42. Fachausschusssitzung Anthropotechnik, München, 24.-25.10.2000, "Multimodale Interaktion im Bereich der Fahrzeug- und Prozessführung", S. 299-307.

[00pat1] Patsouras, Ch.: Zur Unterscheidbarkeit und Bevorzugung der Werckmeisterschen Temperatur gegenüber der Gleichschwebung. Tagungsband DAGA 2000, 26. Jahrestagung Akustik, Oldenburg, 20.-24.3.2000, S. 224-225.

[00pat2] Patsouras, Ch.; Fastl, H.; Widmann, U.; Hölzl, G.: Privacy versus Sound Quality in High Speed Trains. Proc. INTER-NOISE'2000, 28.-30.8.2000, Nizza, Frankreich. CD-ROM.

[00pfa1] Pfau, T.; Faltlhauser, R.; Ruske, G.: A Combination of Speaker Normalization and Speech Rate Normalization for Automatic Speech Recognition. Proc. of ICSLP 2000, Peking, China, 16.-20.10.2000, China Military Friendship Publish, Vol. 4, S. 362-365.

[00rue1] von Rücker, C.: Ein Verfahren zur Tonhöhenanalyse unter Berücksichtigung zeitlich-spektraler Kontrasteffekte. München: Herbert Utz Verlag Wissenschaft, 2000, 131 S.

[00rue2] von Rücker, C: The role of accentuation of spectral pitch in auditory information processing. In: Auditory Worlds: Sensory Analysis and Perception in Animals and Man (siehe [00man1]), S 278-285.

[00sai1] Said, A.; Fleischer, D.; Fastl, H.; Grütz, H.-P.; Hölzl, G.: Laborversuche zur Ermittlung von Unterschiedsschwellen bei der Wahrnehmung von Erschütterungen im Schienenverkehr. Tagungsband DAGA 2000, 26. Jahrestagung Akustik, Oldenburg, 20.-24.3.2000, S. 496-497.

[00sch1] Schorn, K.; Fastl, H.: Hearing impairment: Evaluation and rehabilitation. In: Auditory Worlds: Sensory Analysis and Perception in Animals and Man (siehe [00man1]), S. 286-310.

[00see1] Seeber, B.: Zum Zusammenhang zwischen Mithörschwellen und Tuningkurven. Tagungsband DAGA 2000, 26. Jahrestagung Akustik, Oldenburg, 20.-24.3.2000, S. 290-291.

[00spa1] Spannheimer, H.; Freymann, R.; Fastl, H.: An Active Absorber to Improve the Sound Quality in the Passenger Compartment of Vehicles. Proc. INTER-NOISE'2000, 28.-30.8.2000, Nizza, Frankreich. CD-ROM.

[00ter1] Terhardt, E.: Linear model of peripheral-ear transduction (PET). In: Auditory Worlds: Sensory Analysis and Perception in Animals and Man (siehe [00man1]), S. 81-89.

[00tho1] Thomae, M.; Ruske, G.; Pfau, T.: A New Approach to Discriminative Feature Extraction using Model Transformation. Proc. ICASSP'2000, Istanbul, Türkei, 5.-9.6.2000. IEEE, Vol. 3, S. 1615-1618.

[00val1] Valenzuela, M.N.: Perceived differences and quality judgement of piano sounds. In: Auditory Worlds: Sensory Analysis and Perception in Animals and Man (siehe [00man1]), S. 268-278.

[00wie1] Wiegrebe, L; Hirsch, H.S.; Patterson, R.D.; Fastl, H.: Temporal dynamics of pitch strength in regular-interval noises: Effect of listening region and an auditory model. JASA Vol. 107, No. 6, S. 3343-3350.

Imprint

U
Theresienstraße
Luisenstraße
TUM
TECHNISCHE
UNIVERSITÄT
MÜNCHEN
main campus
Arcisstraße
Barerstraße
Gabelsbergerstraße
Institute for
Human-Machine
Communication
2nd floor
S6
Königsplatz
Brennerstraße
Karolinen-
platz
U

www.mmk.ei.tum.de