Technology Integration in the Daily Activities of the Elderly

Stefan Soutschek, Werner Spiegl, Stefan Steidl, Joachim Hornegger, Hellmut Erzigkeit, Johannes Kornhuber

A continuous increase of the anticipated average life is expected in the near future. We have to deal with the consequences of aging, such as loneliness in an environment characterized by technology or the need to help elderly people. Technical solutions are required that solve the occurring problems associated with aging and are accepted and used. This work faces the challenge of improving the early detection of diseases like dementia, using an intelligent dialogue system that improves the chance for integrating applications of telemedicine in an intelligent home environment.

1 Introduction

1.1 Demography

Due to continuous improvement of medical care, the high standard of living in the first world and many other influences, an increase of the anticipated average life is expected (see Figure 1) [1].In Germany, the population of people older than 80 years is the fastest growing segment of population. From currently four million people this part of the population will increase to approximately 10 million people until 2050. This implies a huge challenge to society, government and industry but also for each individual person. It is necessary to address the different kinds of limitations, for example due to illnesses, needs and other things, that come up with aging. Sustainable solutions need to be created that support elderly people in their everyday activities and to provide assistance in case of emergency. Interdisciplinary research is required for the development of solutions that help to cope with the upcoming changes in our society.

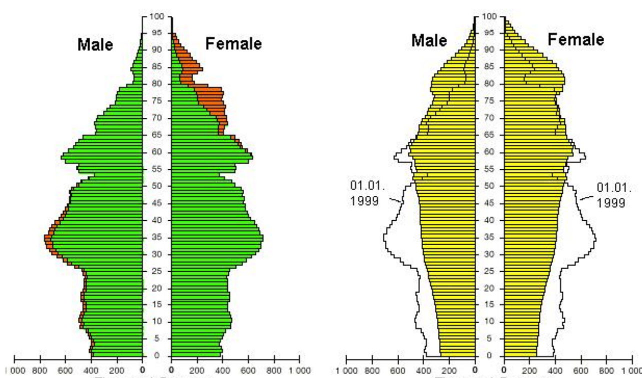

Figure 1: Population pyramid for the years 1999 (left) and 2050 (right) [1].

1.2 Dementia

One of the most prevalent and also expensive diseases of elderly people is dementia. Only considering Alzheimer's disease, approximately one percent of the 65 year old people are affected. Bickel [2, 3] points out that the prevalence of dementia correlates with age and doubles every five years in age. As a consequence ten percent of the people suffer from dementia at the age of 80. The incidence rate shows a similar tendency. Annually, about 150,000 new people are assumed to be affected with dementia.

The first symptoms of the Alzheimer's disease show up by a light decrease of peoples efficiency. Over the course of time this worsens to a total loss of mental abilities and personality. Although other forms of dementia differ in their symptoms, the irreversible and progressive behavior is similar.

An improved early detection of dementia would improve the chances of therapy and thus also reduce the effort of taking care of the affected people [4].

2 Medical and Methodical Aspects

From a clinical point of view the prevention or the early detection of diseases maximizes the chances for better therapies. The early detection of dementia symptoms for instance is based on cognitive deficiency. During the last 15 years, the main focus was on the detection of changes in the fields of social competence and emotionality, which resulted in the main research topics **ADL** (Activities of Daily Living) and **BPSD** (Behavioral and Psychological Symptoms of Dementia) [6].

These cognitive symptoms concerning behavior and health condition are discriminating features for the design and implementation of economical screening instruments, which enable the early detection of dementia [6]. With the incorporation of biological markers, nowadays the reliable diagnosis of dementia and an estimation of the expected sickness degree is possible.

These developments also play an important role in the investigation of new ways for therapeutic intervention. In case of dementia, these interventions comprise cognitive and behavior-oriented training routines as well as medicinal therapies. One question that directly comes along while talking about diagnostic measures concerns appropriate criteria for an evaluation. Traditionally, emphasis is put on economy or cost-benefit analysis, but the impact of acceptance and attractivity of screening instruments becomes more and more important. The recent achievements in telemedicine seem to provide a proper framework [7].

Figure 2: Standard material for SKT [5]

3 Optimization of Diagnosis and Therapy with Telemedicine

Methods developed in the field of telemedicine offer advantages compared to "classical" methods at a comparable diagnostic and therapeutic valence:

- Economic handling
- Highly objective data collection
- Easier verification of the reliability of the collected data
- Easier implementation of intelligently administrated databases
- Automatic adaptation and optimization of standards for expectancy values of clinically relevant features for specific groups to support the decision process
- Possibility to measure control variables for routine duty on a large scale

With the help of the technological progress in telemedicine the amount of collected data and information has highly increased. One big advantage such a large pool of data has is supporting the diagnosis. With the help of methods provided by pattern recognition for example, current test results of a patient can be analyzed by comparing these results to a database that was built from the collected data. In addition to supporting the diagnosis also the execution of tests is optimized. With the help of computers, tests can be simplified on the patient side as well as on the operator side by automatization of the interfaces.

An example for such a test is the SKT (Syndrom-Kurztest), first published in 1977 in Germany [5]. It is a short cognitive performance test for assessing deficits of memory and attention within clinical settings. The SKT is used for staging and monitoring the severity of cognitive impairment in clinical basic research, psychological or psychiatric reports, individual patients, drug trials proving therapeutic efficacy [5].

Traditionally, this test has been performed without using any computer assistance just utilizing material as shown in Figure 2, but during the last months this test was digitized by the Dr. Hein GmbH[1].

In the new digitized version of the SKT, the materials (wooden playboards and forms, see Figure 2) are replaced with a touch-screen monitor on patient side and a mobile computer on the operator side. An obvious improvement is the automatization of the analysis of the acquired data. That means the operator has an optimal support when making a diagnose. A further aim of the implementation was to simplify the usage of the digitized version of the SKT, especially on the operator side but of course also on the patient side. Here the main advantage is that due to the simple user-interface, the operator can concentrate more on the patient and the correct execution of the test.

Another important point is that modifications are easy to realize. With the digitized version it is possible to change for example, colors. This is important for people who have problems to differentiate between certain colors.

[1] http://www.dr-hein.com

4 Integration of Telemedicine in Home Environments

With the possibilities that are provided by the technological progress in telemedicine, a completely new way of testing has been opened. One can think of having mobile stations for example in pharmacies which are connected via Internet to a clinic, where experts analyze the test results and inform the user via e-mail or phone call whether everything is fine or whether an expert should be contacted.

Another way of providing such methods of telemedicine is to integrate these tests in an intelligent home environment. An occupier for example, can repeat this test regularly which offers the possibility to perform a continuous screening that allows for a long term monitoring of the state of health and more specific the deficits of memory and attention for the SKT.

Also for the intelligent home environment, the result is sent to the responsible expert who decides whether it is necessary to visit a clinic or not. For a lot of people, especially for elderly people this represents an easement in their daily life as they do not need to go to the treating physician every time a test needs to be performed. They can sit on their couch in their familiar surroundings and get the instructions from the intelligent system integrated in the home. In our intelligent home environment a dialog system (see section 4.2) has already been integrated.

4.1 Strategies

For maximizing the acceptance of these integrated methods of telemedicine, the user needs to be motivated. As mentioned, the fact that it is not necessary anymore to visit a clinic for performing the test is already a strong motivator. Another point that must be considered is the user-friendliness of such a system, that means, it should not overexert the user. One starting point to minimize the time needed to cope with the handling of the test, is to use already existing hardware, for example television and the dialog system that is part of our intelligent home environment. A further method we will investigate to motivate the occupier is to arouse his ambition by embedding the test into a game environment. If the test is part of the daily or weekly routines, the user is more motivated, when he has the chance to improve his high score or to unlock a piece of an artwork like a puzzle every time he successfully finishes the test.

4.2 Intelligent Dialog Systems

A common dialog system consists of three components: the input interface, the dialog manager and the output interface. In a speech based system a speech recognizer is the input interface and a speech synthesizer the output interface. The dialog manager is the "brain" of the system. In this component the inquiries from the input interface are processed and an appropriate answer is sent to the output interface. With such a dialog system intelligent conversations are possible. One application is an interactive control and test environment for telemedicine as it is proposed in this work. For this use the dialog system constitutes an attendant for the patient/client while he is completing a test (e. g. the SKT). With context based feedback, the patient gets the right instructions for the next step during the complete test. The dialog system must also be able to answer to queries concerning the concrete test situation.

5 Definition of Interfaces

As already pointed out, to keep the motivation as high as possible, the system should be easy to use and as intuitive as possible. This also includes the human-machine-interfaces (HMI) that are necessary to provide interaction between a user and the intelligent home environment. Besides the user-friendliness of the application itself the HMI is one of the most critical parts. If this interface does not satisfactorily address the requirements the user is expecting it will not be accepted. In our home environment two different kinds of HMIs are integrated in addition to classical devices like the standard computer mouse. Both systems are operating in the background, which means that necessary hardware is installed nearly invisible in the home environment and does not need to be adjusted by the user.

The user is able to choose whether he likes to interact with the system using a speech recognition system (see section 5.1) that extracts the necessary input out of natural speech or a gesture-based navigation (see section 5.2) that interprets predefined gestures which represent the necessary interactions. As these interfaces use completely different features (natural speech and hand-gestures respectively) as input, also possible diseases that disable the occupier from using one of the interfaces can be compensated. A user that for example suffers from impairment of motor functions could still use the speech recognition system whereas a user suffering from dysarthria still has the possibility to use the gesture based variant.

5.1 Speech Recognition Incorporating the Peculiarities of the Aging Voice

The current dialog system installed in our home environment is a product of the *Sympalog Voice Solutions GmbH*[2]. Among others it was used in the central administration of the *Sixt AG*[3] as a telephone exchange system, which is able to process spontaneous speech. The average recognition rate of the speech recognizer is 75-80 % with a word lexicon of approximately 3500 words. The system achieves a dialog success rate of more than 90 %. These results are achieved for the average voice, but the clients/patients in our proposed system are elderly people. Hence, a special version of a speech recognizer must be developed: a recognizer that incorporates the peculiarities of the aging voice. As Wilpen and Jacobsen pointed out in [16] there is a dramatical increase of the error rate for people over 70, even if one has enough training data. With any improvement of the

[2]http://www.sympalog.com
[3]http://www.sixt.de

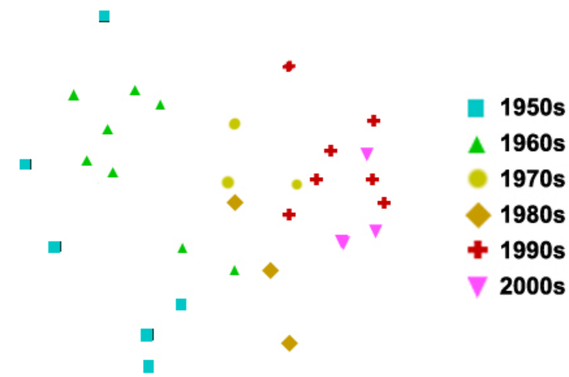

Figure 3: Sammon map of the reduced feature set from the Queen's Speech. Every item represents the recording of one year (1952 to 2002, some years missing). For better illustration the decades have the same icon.

recognition system the error rate for elderly people "remained at least 50 % higher than that for the middle age groups"[16, p. 325]. For further improvements of a speech recognizer, it is essential to understand, what are the characteristics of the aging voice and what makes the difference to the average voice.

A good starting point for this is to visualize features from recordings of the same person at different ages (longitudinal data) and figure out, if there is any tractable voice change over the course of time. For this experiment the annual Royal Christmas Message of Queen Elizabeth II. (Queen's speech) is used. Altogether 30 recordings are available from 1952 to 2002 (some years are missing) with 5-10 min each and 2.5 h in total.

In a first step for every single recording, i. e. for every year, common features are extracted as they are used in speech recognition: Mel Frequency Cepstrum Coefficients (MFCCs). After this, a so called universal backround model (UBM) is created on the whole set of feature vectors, i. e. a Gaussian Mixture Model (GMM) with 128 densities is build on the MFCCs. Based on this UBM, for every single year a GMM is adapted following [18]. This results in a high-dimensional feature vector for every year. It is assumed, that feature vectors of succeeding years are adjacent to each other in the feature space. Due to the high-dimensional feature space, the dimension must be reduced for the purpose of visualization. To get a representation in a lower dimensional space with the maintained topological structure of the data, *Sammon Mapping* is used in the experiment [17]. The results of this operation on the Queen recordings are presented in Figure 3, which describes a snap-shot of the reduction to three dimensions. Each data point in the picture represents the features of one recording, i. e. one year. For better visualization the data of the same decade have the same icon. As one can see the items of the same decade can be clustered and there is a trend from the 1950s to the 2000s: the data points are ordered from left to right by increasing age. Concerning the experimental setup these results lead us to the conclusion that there are tractable features in the recordings en bloc, but not in the voice per se, i. e. the visualized changes also arise from the recording setup, for instance from different microphones. This leads to the channel problem.

To reduce such channel effects, more or less channel inde-

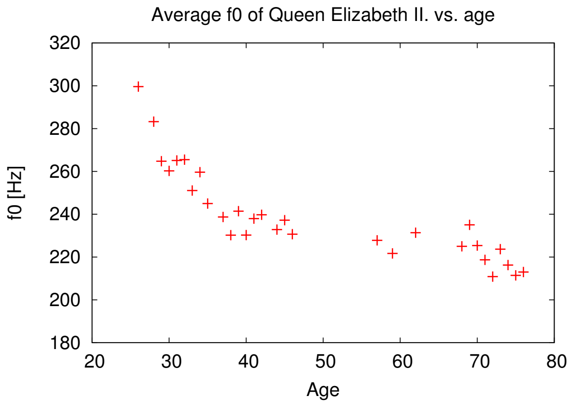

Figure 4: The averaged f0 per year of the Queen's speech. With increasing age the f0 decreases with an age correlation of -0.83/-0.93 (Pearson/Spearman correlation).

pendent features are extracted from the voice. For example the fundamental frequency f0 of the voice. The results from [19] and our experiments on the Queens speech show also a trend in the averaged f0 per year. With increasing age the f0 decreases with an age correlation of -0.83/-0.93 (Pearson/Spearman correlation). In Figure 4 the averaged f0 per year of the Queen is plotted against her age.

To generalize our results, in upcoming experiments cross-sectional (different speakers, different ages) data will be used and additional features will be analyzed: e. g. voice quality, prososdic and linguistic features. Then, with the knowledge how the voice changes over the years and what its characteristics are, special adapted speech recognizers will be built.

A further application for the telemedicine is the diagnosis of voice disorders or emotions reflected in the voice: in a concrete test situation (e. g. the SKT) not only the answers can be evaluated, but also how the answers are presented.

5.2 Gesture-Based Navigation

Besides speech, communication through gestures is one of the most important interfaces. This becomes obvious, only thinking about deafness. For that case, the user interacts with the environment using a defined set of gestures.

Current gesture-based approaches need additional devices like data gloves, colored gloves or other objects which help segment and identify the gesture in the acquired camera image. In most cases these devices are impractical or even inapplicable and therefore lower the willingness to work with such an interface. It is not feasible for an user to wear a data glove every time he wants to interact with the intelligent home environment. But with the help of the latest developments in the field of 3-D cameras which directly provide 3-D depth information in addition to 2-D gray value images, gesture recognition systems are proposed to classify hand gestures [14] or to interpret the movement of the hand [8, 13] without the need of any additional device. Such a 3-D gesture recognition system has big advantages compared

Figure 5: ToF camera: SwissRanger 3100 [15]

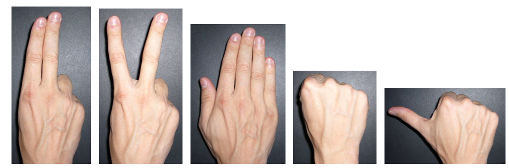

Figure 6: Set of gestures for communicating necessary information

to available 2-D systems [9, 10, 11], which are also capable of recognizing gestures without any additional device. Due to the 3-D depth features of the scene the same functionality can be provided in a faster and more robust way as these features can be utilized for several preprocessing steps for example segmentation or feature extraction. Furthermore the distance information can be used to easily enhance the functionality of such a system by providing an intuitive access to the third dimension.

In our home environment, in a first step, a user-interface which interprets gestures and enables a user to explore and navigate through 3-D data sets, based on a real-time dynamic gesture recognition system using the SR-3100 (see Figure 5) has been developed. This system can already be used to provide virtual tours through famous buildings to give an example. Based on the functional requirements that are necessary for navigating through 3-D data sets like rotation, translation, cursor movement, point selection or a possibility for a reset, we selected five gestures representing this interactions (see Figure 6). The main selection criterion for the set of gestures were to get a small, intuitive gesture set and the complexity of physically performing a certain gesture was chosen to be as small as possible.

This first application already allowed an evaluation concerning the aspects of classification rate, real-time applicability, usability, intuitiveness and training time which provides a good indication of the acceptance of such a system.

For each classifier, a ten fold cross-validation was performed using 40 data sets per gesture, acquired from 15 test persons. The test users were not advised to take off rings or pull up their sleeves which would also not be the case when using the system in the home environment. By additionally applying a principal component analysis on the extracted feature vectors a classification result of 94.3% was achieved.

Also a performance evaluation of the implemented algorithms was performed as the user will not accept the interface if it has a perceptible delay. This evaluation was performed on a single core Pentium M 1.87 GHz. The results show that a computation time of 25.77 ms meets the requirements concerning real-time applicability.

Besides this technical evaluation for the intelligent home environment, it is even more important to gain information about the usability, intuitiveness and training time of the implemented HMI. Thus, in a first evaluation, seven users were asked to use our HMI and perform the different kinds of gestures for about three to five minutes so as to familiarize oneself with the HMI after they got a two minute introduction to the system.

In the practical part, the test users were requested to answer a list of four questions on a scale from one up to five, where one corresponds to "worse" respectively "strong adaptation" and five corresponds to "very good" respectively "no adaptation".

The first question concerned the response time of the system. This question allows to combine the technical evaluation to the user impression. With a mean of 4.14, a satisfying result has been achieved considering possible optimizations. For adapting this system to the SKT, a reduction of the set of gesture could be conceivable and thus also the classification time is reduced.

The users were asked how strong they had to adapt to the system. Here the opinions concerning the adaption to the system differ. It needs to be taken into account that the test persons used the system for the first time. With a mean of 3.57, the individual results of this evaluation can still be judged as good except for one user.

Another important aspect for the evaluation of the usability which was asked was the comfort of the gestures while using the interface. With an average of 4.0, the system convinced the test users.

Also the question concerning the intuitiveness of the cursor movement and click event of the gesture-based interface that was posed to the test users achieved a satisfying result. Especially these two gestures will play an important role when using the gesture based navigation for interacting with the intelligent home environment.

6 Outlook

In this work we present one possible way to take advantage of the technological progress in telemedicine for the integration of economical screening instruments in an intelligent home environment. By that it is possible to gain information about the health situation and also to increase the comfort of elderly people. The integration of the SKT [5] in our intelligent home environment for example, focuses on the improvement of the early detection of dementia but also obviates unnecessary hospital stays.

Such screening instruments will not only help to early detect diseases, which improves the chances of therapy, they will also help building up a rapid growing database, which can than be used for supporting diagnosis.

The next steps of this work will concentrate on evaluating the digitized SKT first of all in a clinical environment to collect data for setting up the database and in a next step, after integrating the test in the intelligent home environment, to evaluate the acceptance of that test. Also the human machine interfaces, that means the speech recognition system and the gesture-based navigation, will be further evaluated as possible input devices for the SKT. For the future also other tests will find their way into the home environment which will further improve the comfort of elderly people.

7 Acknowledgement

This research project is supported by the BFS[4] research association "Zukunftsorientierte Produkte und Dienstleistungen für die demographischen Herausforderungen - FitForAge"[5]. We thank Jonathan Harrington and the BBC for providing the recordings of the annual Royal Christmas Message of Queen Elisabeth II.

References

[1] M. Eisenmenger, et al., Bevölkerung Deutschlands bis 2050. 11. koordinierte Bevölkerungsvorausberechnung. Wiesbaden: Statistisches Bundesamt, 2006.

[2] H. Bickel, Demenzsyndrom und Alzheimer Krankheit: Eine Schätzung des Krankenbestandes und der jährlichen Neuerkrankungen in Deutschland. Gesundheitswesen, 62(4), 211-218, 2000

[3] H. Bickel, Epidemiologie der Demenz, In: K. Beyreuther et al. (Hrsg.): Demenzen. Grundlagen und Klinik, Thieme Verlag, Stuttgart, New York, 15-41, 2002.

[4] E. Gräßel, Häusliche Pflege dementiell und nicht dementiell Erkrankter. Teil II: Gesundheit und Belastung der Pflegenden, Zeitschrift für Gerontologie und Geriatrie, 31(1), 57-62, 1998

[5] H. Erzigkeit, SKT: Kurztest zur Erfassung von Gedächtnis- und Aufmerksamkeitsstörungen, Manual, Geromed, Erlangen, 2001

[6] H. Erzigkeit, H. Lehfeld, J. Pena-Casanova, F. Bieber, et al., The Bayer Activities of Daily Living Scale (B-ADL Scale), Results from a validation study in three European countries, Dementia and Geriatric Cognitive Disorders, 12, 348-358, 2001

[7] H. Lehfeld, H. Erzigkeit, Beeinträchtigungen der Alltagsaktivitäten (ADL) und der kognitiven Leistungsfähigkeit in unterschiedlichen Demenzstadien, Fortschritte der Neurologie - Psychiatrie, 68(6), 262-269, 2000

[8] P. Breuer, Entwicklung einer prototypischen Gestenerkennung in Echtzeit unter Verwendung einer IR-Tiefenkamera, Diplomarbeit, Universität Koblenz-Landau, Koblenz, Germany, 2005

[9] B. Deimel, Entwicklung eines stabilen videobasierten Verfahrens zur Segmentierung der Hand am Unterarm, Diplomarbeit, Lehrstuhl für Graphische Systeme, Technische Universität Dortmund, Germany, 1998

[10] S. Funk, S. Fuchs, Entwicklung eines gestisch-intuitiven Mensch-Maschine-Interfaces auf Basis der videogestützden Erkennung von Handzeichen, Technischer Bericht, Technische Universität Dresden, Dresden, Germany, 2002

[11] A. Elgammal, V. Shet, Y. Yacoob, L.S. Davis, Learning dynamics for exemplar-based gesture recognition, CVPR '03: Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.1, 571-578, 2003

[12] Y. Sato, M. Saito, H. Koik, Real-Time Input of 3D Pose and Gestures of a User's Hand and Its Applications for HCI, In VR 2001: Proceedings of the Virtual Reality 2001 Conference, p. 79, IEEE Computer Society, Washington DC, USA, 2001

[4] Bayerische Forschungsstiftung

[5] www.fit4age.org

[13] T. Sünkel, Erkennung isolierter komplexer Handgesten in 2 1/2D Videosequenzen mit Hidden Markov Modellen, Diplomarbeit, Universität Erlangen-Nürnberg, Germany, 2006

[14] E. Kollorz, J. Hornegger, Gesture recognition with a time-of-flight camera, In DAGM 2007: Proceedings of the workshop Dynamic 3D Imaging in conjunction with DAGM'07, 86-93, 2007

[15] http://www.mesa-imaging.ch

[16] J.G. Wilpon and C.N. Jacobsen, A Study of Speech Recognition for Children and the Elderly, ICASSP, 349-352, 1996

[17] J.W. Sammon Jr., A Nonlinear Mapping for Data Structure Analysis, In IEEE Trans. on Computers 1969, C-18(5), 401-409, 1969

[18] T. Bocklet, Speaker Recognition and Recognition of Speaker Groups Using Gaussian Mixture Models, Studienarbeit, Lehrstuhl für Mustererkennung, Universität Erlangen-Nürnberg, Germany, 2007.

[19] J. Harrington, S. Palethorpe and C.I. Watson, Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers, Interspeech, 2753-2756, 2007.

Contact

Stefan Soutschek M.Sc.
Department of Psychiatry and Psychotherapy
Schwabachanlage 6., 91054 Erlangen
Tel.: +49 (0)9131-8528977
Fax: +49 (0)9131-30 3811
Email: stefan.soutschek@informatik.uni-erlangen.de

Dipl. Inf. Werner Spiegl
Chair of Pattern Recognition
Martensstr. 3, 91085 Erlangen
Tel.: +49 (0)9131-85 28980
Fax: +49 (0)9131-30 3811
Email: werner.spiegl@informatik.uni-erlangen.de

Bild

Stefan Soutschek received his masters degree in computer science from the University of Erlangen-Nuremberg in 2007. He is a PhD student at the Department of Psychiatry at the University of Erlangen, where he is working in close corporation with the Chair of Pattern Recognition on solutions, which counter the consequences of demographic change. The main focus lies on increasing acceptance and attracting interest of such technology with instruments of art.

Bild

Werner Spiegl received his diploma degree in computer science from the University of Erlangen-Nuremberg in 2007. He is currently a PhD student at the Chair of Pattern Recognition, where he is working on solutions, which counter the consequences of demographic change with application of technology. The main focus lies on developing a modern speech recognition system incorporating the aging voice problem.

Bild

Stefan Steidl is a Ph.D. candidate in Computer Science at the Chair of Pattern Recognition at University of Erlangen-Nuremberg, where he also received his diploma degree in 2002. His primary interests lie in the area of automatic classification of naturalistic emotional user states from speech. Previous research has also included work in speech recognition and speaker adaptation.

Bild

Joachim Hornegger graduated in Computer Science (1992) and received his Ph.D. degree in Applied Computer Science (1996) at the University of Erlangen-Nuremberg (Germany). In 2003 Joachim became Professor of Medical Imaging Processing at the University of Erlangen-Nuremberg and since 2005 he is a chaired professor heading the Chair of Pattern Recognition. His main research topics are currently pattern recognition methods in medicine and sports.

Bild

Hellmut Erzigkeit graduated in Psychology in (1972) and received his Ph.D. degree in Psychology (1978) at the University Hospital of Erlangen. In 1985 Hellmut became head of the Clinical Psychology department at the Psychiatric University Hospital of Erlangen. His main research topics are in the area of psychometrics of cognitive impairment of performance, changes of health and behaviour associated with aging and their disorders.

Bild

Johannes Kornhuber obtained an appointment to a full professorship in the Department of Psychiatry at the University of Göttingen in 1978, where he was Chairman of the Gerontopsychiatric Section. Since 2000, he has been a full professor and chairman in the Department of Psychiatry at the University of Erlangen. His research interests include: the pathophysiology of Alzheimer's disease; the early diagnosis and treatment of dementia syndromes.