

Developing Enabling Technologies for Ambient Assisted Living: Natural Language Interfaces, Automatic Focus Detection and User State Recognition

Florian Hönig¹, Christian Hacker¹, Volker Warnke², Elmar Nöth¹, Joachim Hornegger¹, Johannes Kornhuber³

¹Institute of Pattern Recognition, University of Erlangen-Nuremberg, Germany ²Sympalog Voice Solutions GmbH

³Department of Psychiatry and Psychotherapy, University Hospital Erlangen, Germany

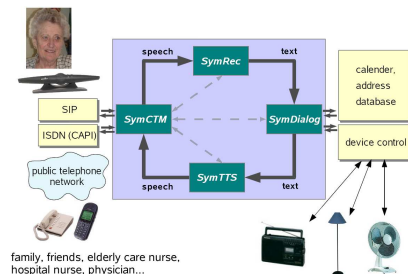
Abstract. Intuitive and practical user interfaces play a key role for successfully applying technologies for ambient assisted living (AAL). We take an approach based on automatic speech recognition that exemplifies such an interface: We present a prototypical system that provides e. g. control of household appliances, initiates and accepts telephone calls and implements reminder functions for taking medicine. It has a natural-language interface: spontaneous speech is allowed; the user does not have to learn special commands. To avoid a "push-to-talk-button", we provide algorithms for automatic recognition of the intended addressee of speech. Additionally, we classify the focus of attention from video recordings which can be useful if the system's interface uses an avatar. Furthermore, we analyze body functions such as heart rate to classify the affective state of the user. In the future, we will apply these generic solutions to health monitoring.

Introduction

- Demographic change will lead to shortage of health care services
- Ambient assisting technologies at home could
 - increase autonomy and quality of life of elderly people
 - support medical therapy
 - avoid or postpone need to move to a health-care facility

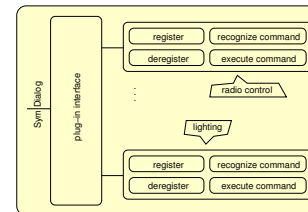
ISA-House

- Prototypical system for studying technology supporting elderly people
- Natural language interface
- Central, unified interface to various technical equipment such as radio, heating or household appliances
- Hiding device-dependent details, communication takes place on the abstraction level of a natural-language dialogue
- Currently implemented functions of the demonstrator:
 - Free speech input
 - Autonomous handling of outbound telephone calls
 - Control of the radio (volume and broadcast station)
 - Control of simple devices like lamp or fan
 - Query function for date and time
 - Plug-in architecture for easy functionality upgrade
- *System architecture* of ISA-house:



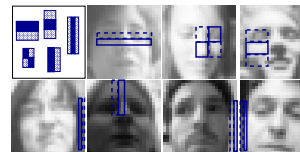
- Speech input via head-mounted microphone, microphone array or telephone/cell phone
- Core of ISA-house: ensemble of tools from Sympalog (spin-off company of the Institute of Pattern Recognition operating several commercial telephone-based dialog systems, www.sympalog.com)
 - * *SymCTM*: call and task manager handling incoming requests and outgoing messages; when voice activity is detected, SymRec is activated
 - * *SymRec*: decodes the spoken utterance
 - * *SymDialog*: easily configurable and powerful dialog manager
 - * *SymTTS*: speech synthesizer generating answers or inquiries issued by SymDialog
- For executing commands or acquiring external information, SymDialog communicates with several devices via a unified *plug-in-architecture*:

- Four interfaces: "register", "deregister", "recognize command" and "execute command"
- SymDialog uses "recognize command" to determine which device is being addressed and calls "execute command"
- New appliances or sensors can easily be integrated with this framework



Classification of User Focus

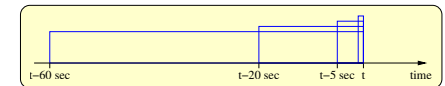
- For natural speech interface, "push-to-talk-button" should be avoided
- Currently, ISA-house combines voice-activity-detection with the "recognize command" interface
- For a more robust classification of the addressee of speech: novel methods to classify the user's focus using both audio and video input
- Audio module analyses the spoken utterance:
 - Classification of *On-Talk* or *Off-Talk* (user talking/not talking to system)
 - 100 features representing prosodic information (computed from energy, pitch, jitter, shimmer, rate-of-speech and the duration of words, syllables and pauses)
 - Further features containing scenario-independent linguistic (information computed from the sequence of nouns/verbs/particles)
- Video module (can be used if system uses e. g. an avatar displayed on television screen):
 - Classification of *On-View* or *Off-View* (user is / is not looking at screen)
 - Using Viola-Jones face detector based on Haar-like wavelet features:



- Combining prosodic, linguistic and visual analysis:
 - 85 % recognition rate for *On-Focus* vs. *Off-Focus* (user does / does not address the system)
- Scenario-dependent linguistic analysis can further increase recognition rate

Recognition of Affective User State

- Recent research in Human-Machine-Interfaces: *affective user state*
- Appropriate reaction to user state increases pleasantness, effectiveness and safety
- Physiological signals give clues about affective user state
- Wireless and integrated sensors currently being developed by number of research institutes
- New methods for automatically recognizing affective user state from physiological signals:
 - Six physiological signals: electrocardiogram, electromyogram, skin conductivity, skin temperature, blood volume pulse, respiration
 - Artefact detection; each undisturbed channel is analyzed separately:
 - * Multi-resolution analysis windows:



- * Large number of generic features per analysis window
- * Reduced by Fisher linear discriminant analysis
- * Classification with Gaussian mixture model (stressed vs. non-stressed user state)
- * Two versions of generic features:
 - 50 efficiently computed features;
 - 44 features that additionally can be computed with few memory
- Fusion of channel decisions: recognition rates between 89 % to 96 % on Drivework (Driving under Varying Workload) Database
- Results can be further improved by user adaption
- Due to generic and data-driven architecture, approach is auspicious for health monitoring

Discussion

- Future work
 - Integrate classification of user focus and user state into ISA-house
 - Collect speech and other data from target group to increase recognition rate
 - In a cooperation between physicians and engineers, use ISA-house to develop new therapy and screening paradigms
 - Research Project *Fit4Age* funded by the Bavarian research foundation
- Summary
 - ISA-house is a promising research platform for exploring the possibilities of AAL
 - Unique features like natural language interface can readily be completed with new services due to plug-in architecture
 - Novel solutions for user focus recognition in audio and video
 - New approach for affective state recognition in physiological signals; good starting basis for medical screening, another promising field in AAL



Universitätsklinikum
Erlangen



Friedrich-Alexander-Universität
Erlangen-Nürnberg

