Exercise Support System for Elderly: Multi-Sensor Physiological State Detection and Usability Testing

Jan Macek and Jan Kleindienst

IBM Czech Republic
V Parku 2294/4
Prague, Czech Republic
{jmacek2, jankle}@cz.ibm.com

Abstract. We present an interactive system for physical exercise of older people and provide results of a usability study with target user group. The system motivates an elderly person to do regular physical activity based on an easy exercise in a monitored environment without a direct supervision from care-givers. Our system employs multi-modal interface including speech synthesis and speech recognition, as well as distance measurement using an ultrasound range finder. The system coaches the user through a sequence of body movements in the exercise utilizing an underlying human activity model.
For evaluation of the performance of the user we present a statistical human activity model to estimate physical load of the user. The system tracks user load by monitoring heart rate and by scanning movement patterns using statistical estimators. At well-defined moments and when the scanning suggests there is a problem with the user, the user is asked to verify his ability to continue with the exercise.
The system was tested on a set of elderly users to gather usability data and to estimate the acceptance of the system. While simplicity of the setup proved to work well for the users, suggestions for further extensions of the system were gathered. Usefulness of the concept was verified with a physiotherapist.

1 Introduction

With the European population getting older, increasing number of supportive work with older people will need to be handled by automated systems. We introduce an interactive multi-modal system to support physical exercise of older people. The system allows a user to perform parts of daily exercises on his own and helps him to keep track of his physical activity in the longer term.

The activity described in this contribution is the chair squat exercise. The user performs the exercise with an armchair using the armrests for support. And he goes through a simple stand up/sit down movement sequence. Although the nature of the exercise is simple for a healthy young person, it is important for an older person who needs to practice movements of daily living regularly to keep his independence level.

The system is composed of Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) components, as well as of ultrasound sensor for distance measurements

and of heart rate monitor. The advantage of the ultrasound sensor is in the reduced need of costly pre-processing of input signal, typically done for video input, before it is used in the actual system.

The system is designed with high modularity in mind in terms of adding sensors and inputs to monitor the user. To make addition of new sensors to the system easy, we use a classifier fusion approach in the user state model.

In the last few years, the area of exer-gaming became highly active with devices like Wii and Xbox offering new types of controllers and games controlled by whole body movements. Although these devices attracted interest among the older users [1], they do not provide games tailored to their specific needs. There exists a system SilverFit [2] tailored directly to the needs of the elderly. This system uses 3D vision system and comes with rather high price tag. We approach the problem with an expandable set of inexpensive and processing "low cost" sensors.

Related work in the field of multimedia includes Lin et al. [3], who propose a meta-classifier approach based on classification of composed feature vectors. These vectors combine the predictions of classifiers from various modalities. Also, the approach copes with asynchronous nature of the outputs from the classifiers.

In medical care, wireless sensor networks are developed [4]. Patients are equipped with wireless vital sign sensors to allow caregivers to monitor their status.

The presented paper expands on the work [5] in following directions. It adds the heart rate sensor to the inputs of the system. It evaluates various classifiers for the expanded set of attributes. And finally, it presents more thorough usability study of the expanded system with the target user group.

In the first part of the present paper, we demonstrate the performance of a selected set of features based on observation of user's movements using the ultrasound range finder and of user's heart rate. We use these features as inputs to four types of classifiers and we compare their performance. This allows us to conclude on usefulness of the measured variables in the framework of user's physiological state estimation. Further, the comparison of single and fused classifier performances allows us to show their value for the modular user activity model.

In the second part of the paper, we present a usability study of our system on a target user group that evaluates various aspects of the design as well as its overall acceptance.

2 Technical realization

We present an extension of the system PHEASY [5], an interactive multi-modal exercise support system for elderly. We add a heart rate monitor to the setup which combines voice technology with a distance measurement sensor. This setup allows the user to interact with the system primarily using body movements, leaving voice as a complementary modality of interaction used when the system needs to verify the user state. An avatar-based text-to-speech interface is used to improve the naturalness of the interaction. The heart rate monitor improves recognition of the physiological state of the user which is then reflected by the system during the exercise.

Figure 1 The PHEASY system in action

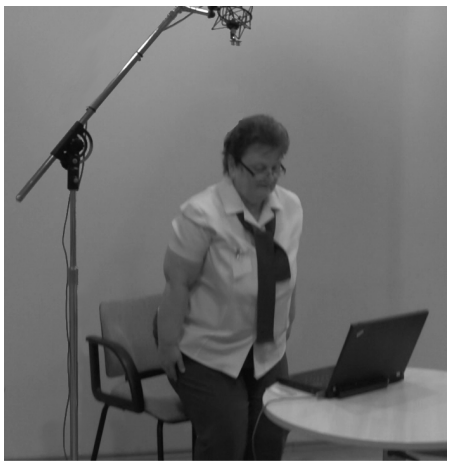
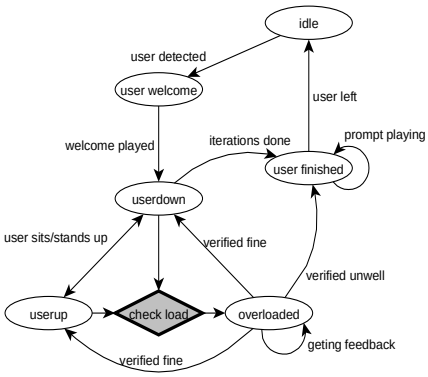


Figure 2 Simplified state model of the exercise system



We used an ultrasound sensor (SRF08 Ultra sonic range finder) to measure the posture height of the user during the exercise. Distance measurements were sampled at 5Hz, a rate sufficient to provide enough information about the subject's activity to the model. Further, we used a heart rate monitor that sampled the heart rate of the user every two seconds. The synchronization of the sonar and heart rate data was done by reading the variables at the same time (every 200ms) while the subroutine handling the reading from the heart rate sensor refreshed the underlying value every 2 seconds. We used the same microphone (AKG400) that was used in training of the acoustic model for the automatic speech recognition (ASR) in the Embedded ViaVoice system (EVV) [6]. For the speech interface, we used the EVV text-to-speech (TTS) engine embedded in a talking head avatar [7]. The interaction of the components and the logical control of the system were implemented in the Conversational Interaction Management Architecture (CIMA) dialog framework [8, 9]. The system setup with a user in action is shown in Figure 1.

The logical design of the system is presented in Figure 2. The calibration of the ultrasound sensor is crucial for well defined transitions between the states of the exercise. In the start, the system performs calibration of the chair height and goes to the "idle" state waiting for a user. When the user comes and sits (natural to do in our setup) the system goes to the "user welcome" state and records height of the seated user while playing the welcome message and instructions to the user. Based on the two measured heights, the system is able to detect the body postures as well as the event of user leaving the system. After the welcome and instructions are played to the user the system moves to the state "user down" and starts the cycle of prompting the user to perform respective posture change and waiting for the user to finish the change. This forms the main line of the exercise resulting in switching between the two states "user down" and "user up". During switching between the states, the switch durations and state dwelling durations are recorded. These values then translate to input features used by the diamond highlighted "check load" box to detect "high load" state of the user; if such an event occurs the system moves to the state "unwell". Here the user is prompted if he is feeling unwell and the exercise should terminate. If the

user confirms the exercise terminates; if the user indicates that he feels fine the estimated state "unwell" is overruled and the exercises continues. When the user feels unwell or the final iteration was reached the system goes to the state "user finished". The goodbye prompt is played and after the user leaves the state "idle" takes over.

User Activity Model. The cornerstone of the system is the user activity model. It is used to predict user's state based on the input features. The input features are extracted from the distance measurements and heart rate monitor readings during the exercise. In our experiments we use following input features: sit-down duration/average sit-down duration ratio, stand-up duration/average stand-up duration ratio, sitting/standing duration, difference of going up duration from average, difference of going down duration from average, sit-down/stand-up ratio, heart rate, and heart rate difference from the user's average. The durations in the input features are measured from the moment the avatar prompts the user with the respective instruction until the user starts moving. The load estimation task is the highlighted diamond box in Figure 2. A classifier is used to estimate user's physiological load and to classify it to the two classes "high" and "low".

Classifier fusion. Classifier fusion techniques combine classifiers that operate on distinct inputs. They work on a higher level of generalization than data fusion and feature fusion techniques (that we label *"all attributes"* in our result tables). The method of classifier stacking combines classifiers by training a meta-classifier that uses outputs of the partial classifiers to give the final classification. The classifier ensemble methods use classifiers with identical inputs which are trained using various types of data reweighting, resampling or bootstrapping and which are then assigned weights in the final ensemble.

3 Evaluation

In this section we present experimental results which verify the possibility to recognize various levels of physical load of the user. The physiological load is estimated using statistical methods based on outputs from the distance sensor and from the heart rate monitor. All experiments were performed using a group of younger users, different from the target user group (60+ years). Data for the target user group will be collected in a longer term once the system is initially tuned.

To evaluate performance of the proposed method, we collected a dataset of feature readings on eight subjects in their thirties. Two types of data were collected for this user group. In the first fold, the subjects came afresh to our measuring device and his performance during the exercise was recorded via the sonar distance measurements and heart rate sensor. In the second fold, the users were asked to perform physical exercises to achieve significant load on cardiovascular system prior to the exercise.

As a result we obtained a dataset containing two classes in the examples, the "low load" class and the "high load" class. We collected a total of 166 examples: 76 cases of the class "low load" and 90 cases of the class "high load".

Table 1. Results of the baseline classification experiments. Precision, recall and F-measure shown per class. *All attributes used*.

Classifier	Accuracy	Precision		Recall		F-measure	
		high	low	high	low	high	low
Random forest	82.5%	0.802	0.862	0.9	0.737	0.848	0.794
C4.5 (decision trees)	78.9%	0.796	0.781	0.822	0.750	0.809	0.765
AdaBoost with decision stumps	76.5%	0.763	0.768	0.822	0.697	0.791	0.731
Bagging with REPTrees	72.9%	0.742	0.712	0.767	0.684	0.754	0.698
Naïve Bayes	64.5%	0.607	0.905	0.978	0.250	0.749	0.392

Experiments. To verify our hypotheses (1) about separability of the two classes in data and (2) about the improvements of classification accuracy by adding the distance based input features, we performed experiments by constructing following types of classifiers.

Initially, we trained two separate classifiers on both types of data; i.e. one classifier just used the distance readings and the other one used heart rate readings. Next, we performed data fusion on the attribute level by merging the heart rate and sonar distance attributes into a single vector. Using this new merged data, we trained a combined classifier. Finally, we applied the classifier fusion technique. Here, we took the two classifiers trained on the two separate datasets and combined them using stacking into a single fused classifier [10].

As an addition and alternative to the previous schemes, we also experimented with an attribute selection algorithm that, using Best First search [11], selects a subset of attributes used to the train a classifier.

In the experiments we used WEKA [12] to train and evaluate the classifiers. As the base classifier, we used the random forests that performed best in comparison to the other classifiers. In these classifier selection trials we compared Random Forests (10 trees built on 5 random attributes) to Bagging with REPTrees (Reduced Error Pruning decision trees), C4.5 (decision trees), AdaBoost with Decision Stumps, and Naïve Bayes. The results of the baseline experiments are shown in Table 1.

Experimental Results. Table 2 shows the results for classification experiments with the selected type of base classifier over the range of attribute selection and classifier combination techniques. The results were obtained using 20-fold cross validation run on all 166 instances available. Apparently, the attributes collected proved to separate the two load levels.

Taking the random forest built on all attributes as a baseline the performance of a random forest classifier drops when it uses sonar attributes and improvement when it is built on heart rate attributes. Further improvement is achieved when using the Best First search for attribute selection and the random forest classifier (labeled as “attribute selection” in Table 2).

Table 2. Results of the classifier and attribute combination experiments. Precision, recall and F-measure shown per class.

Classifier	Accuracy	Precision		Recall		F-measure	
		high	low	high	low	high	low
Random forest, all attributes	82.5%	0.802	0.862	0.9	0.737	0.848	0.794
Random forest, HR only attributes (1)	88.5%	0.890	0.880	0.9	0.868	0.895	0.874
Random forest, sonar only attributes (2)	78.9%	0.772	0.815	0.867	0.697	0.817	0.752
Random forest, attribute selection	85.5%	0.837	0.882	0.911	0.789	0.872	0.833
Stacking of (1) and (2) with meta-classifiers:							
- C4.5	90.9%	0.941	0.877	0.889	0.934	0.914	0.904
- logistic regression	95.7%	0.966	0.948	0.956	0.961	0.961	0.954

Finally, the best performance was achieved with the stacking technique. Initially, two classifiers were built on the *exclusive* sets of sonar and of heart rate attributes. Then they were combined into a single classifier. The top performance of 4.3% error rate was achieved when logistic regression was used as the meta-classifier.

4 Usability Evaluation

This section summarizes responses in a usability testing performed on 13 users from the target user group (distinct from the group of user used in previous section). The users chose values from semantic differential scales with even number of options. The answers were clustered in the summary plots (Figures 4 to 12) for better legibility. Age distribution is shown in Figure 3. Most of the users were in their sixties. The users had no significant impairments. Few had minor hearing and vision problems.

We asked the users if the exercise introduction and explanation was intelligible. At his point, the user was welcomed to the system and detailed description of the exercise was presented. The explanation intelligibility is shown in Figure 4.

The combination of spoken and animated UI elements proved to be confusing to some users, when they focused solely on the video animation. This caused decoupling of their movement from the spoken guidance. Sometimes, the user started with the exercise without waiting for the audio explanation. Although the system reacted properly to the movements of users, sometimes the commands were cut short due to the switch to the following state of the exercise.

Further, we looked at the acceptance of the two modalities of navigation through the exercise. The subjective usefulness of the two modalities is presented in Figures 5 and 6, respectively. The comparison of the two modalities is in favor of the animation of the exercise as indicated by 85% of the users.

An interesting point was the speed of the exercise. The command to start the following movement is always issued immediately after the user reaches a particular

Figure 3

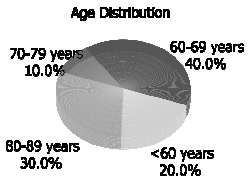


Figure 4

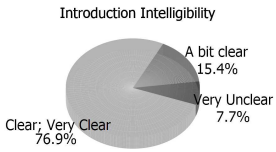


Figure 5

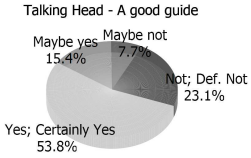


Figure 6

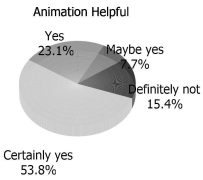


Figure 7

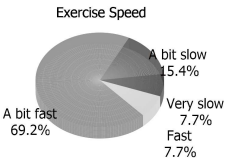


Figure 8

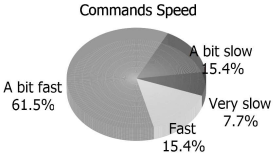


Figure 9

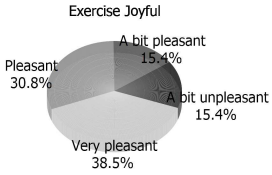


Figure 10

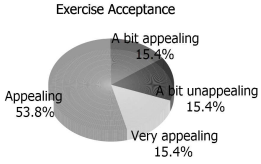
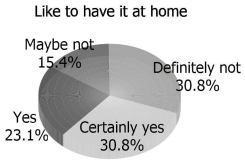


Figure 11



state. This immediacy had strong influence on the perception of the exercise speed. The subjective speed of the exercise is shown in Figure 7. Related to this issue is the perceived speed of synthesized speech as shown in Figure 8. For both qualities, we observed that most of the users felt the speeds were faster than they would prefer.

We asked the users if they enjoyed doing the exercise with the system and if they liked the general idea of performing exercise guided by a computer (Figure 9 and 10), and also if they would consider using the system in their homes (Figure 11).

5 Discussion and Conclusion

Our first goal was to validate the usefulness of the sonar range finder measurements of user movement and the output of the heart rate monitor for classification of user's physiological load.

To prove the usefulness of our approach, we demonstrated performance of five types of classifiers trained on this type of data. The presented results support the addition of these combined modalities to the inputs of the user activity model. The best performance of 95.7% accuracy was achieved with the stacking technique with logistic regression as meta-classifier and random forests, trained on separate datasets for movement and for heart rate, as base classifiers. This allows us to combine the two modalities, user's body movements and heart rate, as inputs to the user activity model. The usability study showed good acceptance of the system by the target user group.

Although the nature of the exercise is straightforward, it was found joyful way to exercise by most of the users. During the usability testing we had a chance to consult the usefulness of our system with a professional physiotherapist to her daily work. The concept of the system was welcomed mainly to its focus on the movements of the daily life. Suggestion on extension of the system with further exercises of similar focus on daily activities was given and use in company of users was proposed.

In the future work, we will focus on the dialog development as the corner stone of fluent interaction. Further, improvements of the user activity model with larger datasets from real use will add to the system performance.

6 Acknowledgments

We would like to acknowledge support of this work by the European Commission under IST FP6 integrated project NetCarity, contract number IST-2006-045508.

7 References

1. Theng, Y.-L., Dahlan, A. B., Akmal M. L., Myint T. Z. An exploratory study on senior citizens' perceptions of the Nintendo Wii: the case of Singapore, *Proc. of i-CREATe '09*. 2009.
2. SilverFit. http://www.silverfit.nl/. 2011.
3. Lin W.-H., Jin R., Hauptmann A. Meta-classification of Multimedia Classifiers. *Proc. of Int'l Workshop on Knowledge Discovery in Multimedia and Complex Data*, Taipei, Taiwan, May 6, 2002.
4. Shnayder V., Chen B., Lorincz K., Fulford-Jones T. R. F., Welsh M. Sensor Networks for Medical Care. TR-08-05, Div. of Engineering and Applied Sciences, Harvard University, 2005.
5. Macek J., Kleindienst J. PHEASY – Physical exercise assistance system - Evaluation and Usability Study. *IADIS Interfaces and Human Computer Interaction 2010*. 2010.
6. IBM. Embedded ViaVoice. http://www-01.ibm.com/software/pervasive/embedded_viavoice/, 2010.
7. Kunc, L. and Kleindienst, J, 2007. ECAF: Authoring language for embodied conversational agents. *Proceedings of TSD 2007*, LNCS 4629, Springer, pp. 206-213.
8. Cuřín J., Kleindienst J., Kunc L., Labský M., 2009. Voice-driven Jukebox with ECA interface. *Proc. of 13th International Conference "Speech and Computer" SPECOM'2009*.
9. Potamianos G., Huang J., Marcheret E., Libal V., Balchandran R., Epstein M., Seredi L., Labsky M., Ures L., Black M., Lucey P., 2008. Far-field Multimodal Speech Perception and Conversational Interaction in Smart Spaces. *Proc. of the HSCMA Joint Workshop on Hands-free Speech Communication and Microphone Arrays.*
10. Ghahramani Z., Kim H.-C.. Bayesian Classifier Combination. Gatsby Technical Report. University College London, UK. 2003.
11. Hall, M. A., Smith, L. A. (1998). Practical feature subset selection for machine learning. In C. McDonald(Ed.), *Proceedings of ACSC'98*, Perth, 4-6 February, 1998, pp. 181-191.
12. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009: The WEKA Data Mining Software: An Update. SIGKDD Explorations 11(1).