Demonstrating the Dialogue System of the Intelligent Coaching Space

Iwan de Kok[1,2], **Julian Hough**[1,3], **Felix Hülsmann**[1,4], **Thomas Waltemate**[1,4], **Mario Botsch**[1,4], **David Schlangen**[1,3], and **Stefan Kopp**[1,2]
[1]CITEC, [2]Social Cognitive Systems, [3]Dialogue Systems Group,
[4]Computer Graphics & Geometry Processing Group
Bielefeld University
idekok@techfak.uni-bielefeld.de

Abstract

This demonstration presents the current state of the dialogue processing of the Intelligent Coaching Space. This is a multimodal virtual environment in which users are coached in the acquisition of a physical skill. The demonstration highlights the closed interaction loop between the physical action of the user and the responses of the virtual coach.

1 The Intelligent Coaching Space

This demonstration presents the current state of the dialogue processing in the ICSPACE (Intelligent Coaching Space) project at the Cluster of Excellence 'Cognitive Interaction Technology' at Bielefeld University. In this project we are building an immersive, multimodal virtual environment in which users are coached in the process of motor skill acquisition.

A virtual coach observes the user attempting to acquire a motor skill (in our first scenario we focus on squats) and gives incremental instructions and feedback as a human coach would. The domain of physical skill acquisition creates challenges for our dialogue system not present in more traditional pedagogic domains such as tutorial systems. These challenges include fully multimodal input and generation of actions, additional grounding. See (Hough et al., 2015) for a more detailed discussion.

2 Physical Setup

The lab setup of our Intelligent Coaching Space is realized in a Cave Automatic Virtual Environment (CAVE), an immersive 3D Virtual Reality environment with front and floor projection. Users enter our environment wearing 3D glasses and a motion capture suit. They are then tracked by 10 OptiTrack motion capture cameras.

The 3D glasses are tracked to adjust the perspective in our highly responsive custom-built renderer visualizing the scene. In the scene the users see a virtual reflection of themselves in a Virtual Mirror, which is rendered using data from the motion capture suit. Next to the mirror a virtual coach is present that observes the user and instructs the user on how to do a squat.

The Motion Analyzer identifies squats in the stream of motion tracking data and classifies errors made. Based on this information the dialogue system determines the next coaching action.

3 Dialogue Processing

The dialogue system consists of several components which will be discussed in detail below.

The Coaching Strategy Manager is responsible for selecting the next coaching action. It is implemented as a finite state machine making decisions based on an information state. This information state is updated by processing the incoming user input, in this case output of the Motion Analyzer, and also feedback from the Realizer, which informs the Coaching Strategy Manager on the status of its own behaviour.

The information state keeps track of the how many squats have been performed by the user in the current interaction, the errors made during each squat and which phase of the squat the user is currently in.[1] Central to this information state model is the variable *Skills-Under-Discussion*, which unlike traditional Questions-Under-Discussion (QUD) components is not a stack of proposition-based questions, but one of action representations. Based on the current state, the Coaching Strategy Manager selects the most appropriate coaching act, which could be an instruction, demonstration, explanation or feedback.

[1]The squat is separated into a preparation phase (assuming the starting position), stroke (going down), strokehold (in the lowest position) and retraction phase (coming up).

Based on the number and severity of errors in the last squat the Coaching Strategy Manager decides whether to address existing errors by pushing the sub-action(s) performed erroneously onto the Skills-Under-Discussion (SkUD). This will modify the coaching acts that can be selected so they specifically target this aspect of the action. Like Ginzburg (2012)'s QUD, SkUD pops its top element in a stack-like fashion once the error has been corrected, and the overall interaction follows a coaching cycle as described in (de Kok et al., 2014).

Based on the decision made by the Coaching Strategy Manager, the Action Pattern Manager activates the Action Patterns required to realize the action of the coach. These Action Patterns are designed to be dynamically created, activated and/or stopped. Currently each action the Coaching Strategy Manager can choose is implemented as its own Action Pattern. All Action Patterns are their own decision makers that are free to produce behaviour fitting the constraints from earlier decision makers, typically the Coaching Strategy Manager.

Each Action Pattern can create its own information flow links to all other parts of our system. For instance, the `Incremental Instruction` pattern directly listens to the output of the Motion Analyzer, bypassing the Coaching Strategy Manager. Note that it can still be deactivated by the Coaching Strategy Manager if it decides on another action. The pattern will produce instructions to improve the on-going squat. Instruction selection is based on the errors detected in the squat and can be restricted by the Coaching Strategy Manager. E.g., if the Coaching Strategy Manager has decided that the maximal Skill-Under-Discussion is X and both an error in X and Y are observed, only an instruction to correct X will be vocalized. If no restrictions are placed, the action pattern is free to make this decision itself. It will continue giving these instructions until either no squat is currently being performed or the Action Pattern is deactivated by the Action Pattern Manager in response to a decision by the Coaching Strategy Manager. When de-activated it will immediately interrupt all its current and planned behaviours.

Other Action Patterns may be more straightforward, simply converting the action selected by the Coaching Strategy Manager into behaviour directly, without listening to any input other than that from the Action Pattern Manager.

Actions produced by the Action Patterns are realized by the AsapRealizer (van Welbergen et al., 2014). It transforms the actions into joint rotations, blend shapes and sound (using CereVoice TTS) which are passed on to the renderer.

4 Demonstration Overview

The demonstration will feature a portable version of the system presented in Section 2, which will highlight some of the challenges in dialogue management presented in Section 1.

To scale down the demonstration, the 3D CAVE environment is reduced to a single monitor. The screen will show the Virtual Mirror on which a virtual reflection of a coachee is displayed. Our virtual coach will stand next to mirror, interacting with the virtual coachee's reflection.

Instead of motion capturing people performing a squat, the demonstration will play prerecorded squats from file. These will be processed by the Motion Analyzer and played back on the screen in the Virtual Mirror. Our virtual coach will incrementally instruct the coachee during playback highlighting the tight interaction between action of the user and the coach. We will demonstrate different parameters of our coach's coaching strategy during these virtual training sessions.

Acknowledgments

This research is supported by the Deutsche Forschungsgemeinschaft (DFG) in the Center of Excellence EXC 277 in Cognitive Interaction Technology (CITEC).

References

Iwan de Kok, Julian Hough, Cornelia Frank, David Schlangen, and Stefan Kopp. 2014. Dialogue structure of coaching sessions. In *Proceedings of the 18th SemDial Workshop on the Semantics and Pragmatics of Dialogue*, pages 167–169, Edinburgh.

Jonathan Ginzburg. 2012. *The interactive stance*. Oxford University Press.

Julian Hough, Iwan de Kok, David Schlangen, and Stefan Kopp. 2015. Timing and grounding in motor skill coaching interaction: Consequences for the information state. In *Proceedings of the 19th SemDial Workshop on the Semantics and Pragmatics of Dialogue*.

Herwin van Welbergen, Ramin Yaghoubzadeh, and Stefan Kopp. 2014. AsapRealizer 2.0 : The Next Steps in Fluent Behavior Realization for ECAs. In *Intelligent Virtual Agents*, pages 449–462.