

Dialogue Structure of Coaching Sessions

Iwan de Kok^{1,2}, Julian Hough^{1,3}, Cornelia Frank^{1,4},
David Schlangen³, and Stefan Kopp^{1,2}

¹CITEC, ²Social Cognitive Systems,

³Dialogue Systems Group, ⁴Neurocognition and Action (Biomechanics)
Bielefeld University

idekok@techfak.uni-bielefeld.de

Abstract

We report initial findings of the ICSPACE (‘Intelligent Coaching Space’) project on virtual coaching. We describe the gathering of a corpus of dyadic squat coaching interactions and initial high-level models of the structure of these sessions.

1 Introduction

While interactive tutoring systems which perform factual teaching have been established for some time (Litman and Silliman, 2004; Graesser et al., 2005), dialogue systems capable of skill coaching are much rarer. We introduce preliminary work on the ICSPACE (‘Intelligent Coaching Space’) project, which aims to create a virtual intelligent coaching agent in an interactive environment to train users to perform complex motor actions.

Coaching physical movement skills requires combining communication with real-time tracking, assessing and correcting the motor action of the coachee. In particular, giving online feedback while the coachee is carrying out an exercise (Sigrist et al., 2013) is an interesting challenge imposing specific requirements on the system.

To identify these requirements more precisely we analyse two recordings of a professional coach training individuals to perform a squat. We focus on the overall dialogue structure and observe which dialogue situations arise.

2 Recordings

We invited a professional coach to our lab to record two coaching sessions. The coach was asked to instruct coachees how to do a squat as he would teach it in the gym. The coachees (one female (A), one male (B)) were familiar with doing squats, although they had not received instruction from a professional coach before. Each interaction lasted between 4 and 5 minutes.

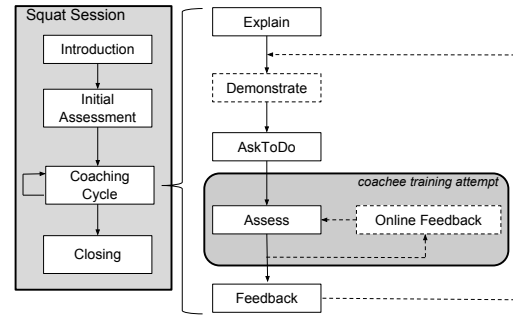


Figure 1: Overall structure of a squat coaching session (left) and structure of a coaching cycle within the session (right).

3 Dialogue Structure

In analysing the two dialogues, several commonalities to the structure of the interaction were observed, as well as some interesting differences. From these dialogues a common overall coaching structure can be inferred, represented in the left-hand shaded box in Figure 1 – temporal dependencies are represented by arrows.

In both dialogues, there was an *Introduction* phase where rapport with the coachee was established by the coach, consisting of questions about personal details and also establishing their previous experience of the squat exercise.

In the second phase of *Initial Assessment*, while in A’s session this was completed after lengthy explanation, for B the coach begins by asking the coachee to do a squat before any explanation. In both cases the coach assesses the coachee’s ability at performing the squat, identifying the sub-movements and aspects of their technique which fall short of the coachee’s potential, and subsequently planning the following coaching behaviour.

What follows are a series of *coaching cycles*, each explaining a particular area in which the movement is being executed below the potential

of the coachee, or which the coach wants to bring to attention to ensure the coachee will continue executing that aspect correctly. As soon as the coach evaluates the action as being performed appropriately after several iterations, the session is *closed*.

Coaching cycles Each coaching cycle follows a similar structure as depicted in the right-hand diagram of Figure 1 – the optional components and transitions have dashed lines. In both dialogues, the coach starts by *explaining* the particular (aspect of the) movement this coaching cycle will focus on (see visually, top-left in Figure 2). Often the coach *demonstrates* the movement and highlights the area of interest with gestures. The coach then *asks* the coachee to perform the movement and they comply, else this is done without prompting as the coachee takes initiative (as was the case in the top-right image of Figure 2). The latter case shows the possibility of mixed initiative in coaching dialogues and is analogous to question accommodation in issue-based dialogue management (Larsson, 2002), however the “answer” to the accommodated question here is non-linguistic.

As the coachee attempts the squat, either in a single effort or in repetition, the coach *assesses* the movement being performed by the coachee (bottom-left in Figure 2) and may give *online feedback* during execution to adjust the movement, either in the form of short utterances as verbal feedback, gesture or even by performing the movement synchronously with the coachee (bottom-right in Figure 2). Once the coach is satisfied with the result or can not adjust the movement during execution, he will stop the coachee and give more lengthy final *feedback* on the movement. During this stage he will often explain why this particular (aspect of the) movement is important. If the coach is satisfied he moves on to the next aspect (if there are any remaining to be corrected), otherwise the cycle will be partially repeated by either another demonstration or request to try again.

Spatial positioning of the coach The coaching cycles were not only noticeable in the dialogue structure, but also in the coach’s movement. The explanation and feedback phases were always performed in front of the coachee – the *instruction space* (see top two images in Figure 2). During the assessment and online feedback phases the coach usually moved to the side of the coachee to get a good profile view – the *observation space*



Figure 2: Images from the recordings showing *explaining* (top-left), *demonstrating* with user initiated following of (top-right), *assessing* (bottom-left) and finally giving *online feedback* on (bottom-right) the squat movement. The images also highlight the *instruction space* (top) and *observation space* (bottom).

(see bottom two screenshots in Figure 2). Initial demonstrations are usually performed in the instruction space. If the coach demonstrates the movement during the online feedback phase, this is performed in the observation space. This multi-locational instruction behaviour is similar to that exhibited by music teachers in instrumental lessons, who tend to move between the *work zone* and the *listening zone* (Duffy and Healey, 2012).

4 Conclusion

We have analyzed the dialogue structure of two exemplary squat coaching sessions to identify the requirements of a virtual intelligent coaching system. The initial recordings show the need for multi-modal turn-taking, use of different spaces for different coaching phases and the ability to generate fast incremental feedback as described by (Kopp et al., 2013) during phases of *online feedback*. As a next step we plan to collect more recordings with different coaches with coachees of different skill levels. This should give evidence for whether the dialogue structure we hypothesize generalizes to the domain, and it will inform the design of an artificial coach capable of online feedback and dialogue in coaching.

Acknowledgments

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

References

- Sam Duffy and Patrick GT Healey. 2012. Spatial coordination in music tuition. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*.
- Arthur C Graesser, Patrick Chipman, Brian C Haynes, and Andrew Olney. 2005. Autotutor: An intelligent tutoring system with mixed-initiative dialogue. *Education, IEEE Transactions on*, 48(4):612–618.
- Stefan Kopp, Herwin van Welbergen, Ramin Yaghoubzadeh, and Hendrik Buschmeier. 2013. An architecture for fluid real-time conversational agents: integrating incremental output generation and input processing. *Journal on Multimodal User Interfaces*, pages 1–12.
- Staffan Larsson. 2002. *Issue-based dialogue management*. Department of Linguistics, Göteborg University.
- Diane J Litman and Scott Silliman. 2004. Itspoke: An intelligent tutoring spoken dialogue system. In *Demonstration Papers at HLT-NAACL 2004*, pages 5–8. Association for Computational Linguistics.
- Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. 2013. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic bulletin & review*, 20(1):21–53.