Timing and Grounding in Motor Skill Coaching Interaction: Consequences for the Information State

Julian Hough[1,2], **Iwan de Kok**[1,2,3], **David Schlangen**[1,2], and **Stefan Kopp**[2,3]
[1]Dialogue Systems Group, [2]CITEC, [3]Social Cognitive Systems Group
Bielefeld University
`julian.hough@uni-bielefeld.de`

Abstract

While tutorial dialogues have been well-studied, the nature of dialogue in physical coaching scenarios is much less well understood. We present a corpus study on coaching interactions wherein a coach trains a trainee to improve a motor skill. We show how our findings put novel requirements on pedagogic dialogue act taxonomies, grounding criteria and information state update models of situated dialogue. One of these requirements is to distinguish between grounding in the traditional sense along an understanding dimension and grounding in terms of a motor program schema, the latter being due to the coach's goal to transfer knowledge of a physical movement to the trainee. Another requirement is that a fine-grained notion of time, both in absolute and relative terms, must become a first class citizen of the dialogue state to be able to model motor skill coaching. A final requirement for an information state model is characterizing what is under discussion and in the established common ground– in these kind of domains this is generally not questions and propositions, but skills and their desired and observed parameters.

1 Introduction

Dialogue in pedagogic domains presents interesting challenges for corpus studies and formal dialogue models. In contrast to more commonly studied task-completion oriented dialogues where an instructor influences their instruction follower's action towards a successful outcome (i.e. to do something), in pedagogic domains the intended outcome for the instructee is *learning gain*– that is, measurable improvement at the task at hand (i.e. to learn *how* to do something or improve upon it).

Dialogue research in tutorial domains requires a relevant dialogue act (DA) taxonomy that deals with grounding understanding of a given skill or piece of knowledge, such as those in (Boyer et al., 2007; Boyer et al., 2008; Boyer et al., 2009). The most well-developed DA taxonomies designed for task-completion dialogues (e.g. DAMSL (Core and Allen, 1997)), while sharing certain communication management DAs, require extensions with DAs that capture 'know-how'– that is, the transmission of skill, knowledge and technique from tutor to tutee. The nature of the *feedback* on tutee attempts, both successful and unsuccessful at the task at hand is also crucial to the taxonomy. Particularly, the degree of positive affect with which a tutor gives feedback can influence learning outcomes. Tutorial dialogue systems such as (Litman and Silliman, 2004; Graesser et al., 2005) use the insights from DA-based corpus studies in their systems to generate appropriate dialogue acts to maximise learning gain.

Situated dialogue, where participants are either physically co-present or have access to a commonly shared virtual space, presents other challenges for DA taxonomies. Grounding DAs such as feedback and repair need not only reference previous verbal utterances, as in (Schegloff et al., 1977), but can also reference non-verbal actions which concern a physical task at hand. In this regard Raux and Nakano (2010) study three types of non-verbal action corrections in a computer-game dialogue whereby a manager guides a player through a task in a virtual environment. Failures in communication are addressed by the manager via correction of errors of three observed types: *Commission* (failure to do the expected or appropriate action), *Omission* (failure to react to an instruction) or *Degree* (appropriate type of action carried out but falling short of the intended outcome by some real value). They showed the three correction types were uniformly distributed, but

showed differences in timing: Commission and Degree corrections were likely to be produced much closer to the error-containing action's start time (on average 2.3 seconds and 2.4s) than normal non-corrected instructions were to correct actions (3.8s).

Embodied situated dialogue becomes even more complex to analyse when gesture and speech interact. Lücking et al. (2013) provide a rich gesture taxonomy and morphology with a proposed interface to the semantics of speech. While this mark-up is comprehensive, it focuses on the transfer of spatial scene descriptions. The domain is face-to-face route description whereby a route giver will make frequent use of iconic and deictic gestures to indicate where their dialogue partner should be locationally during their route. This is a complex type of multi-modal knowledge, but not the embodied procedural knowledge required in motor skill learning which we focus on here.

In this paper, we address the intersection of pedagogic and embodied situated dialogue found in the domain of motor skill coaching, and describe findings which reveal part of the nature of the information state of the participants in these interactions. The rest of the paper is as follows: §2 describes the challenge and uniqueness of motor skill coaching dialogues, §3 outlines our research questions, §4 describes our findings on timing and grounding in a coaching corpus study, §5 describes the consequences for information state update approaches to dialogue modeling and §6 concludes.

2 Motor Skill Transmission in Coaching

For a technical skill such as computer programming, learning gain can be assessed by tutors and tutorial systems by generating open questions about procedural knowledge such as "What should you do now?" to which the tutee can provide an answer (e.g. "I will use an array") to show evidence of their competence (Boyer et al., 2008). The tutor evaluates the tutee's progress by such question answering, and gives appropriate feedback. The goal is fairly clearly presented in this cases to the tutee as the learning gain criterion is set out in advance– for instance the achievement of higher scores.

However, in motor skill learning, for a human coach, task success is much more difficult to evaluate and communicate, particularly if the outcome is not directly observable by the trainee. In such purely technique-oriented tasks, the feedback from a coach is vital to learning success, and for novices, this feedback defines it.

Furthermore, under McMorris (2014)'s definition of a physical skill as "the consistent production of goal-oriented movements, which are learned and specific to the task", the requirements beyond factual knowledge learning increase again: the situated, embodied nature of a motor skill means feedback both from the coachee's own perceptual self-monitoring and externally from the coach is time-critical, with online instructions being of utmost importance.

We will assume the coach's goal is to induce in the trainee a *motor program schema* (Schmidt, 1975), and evidence as to whether the coachee has induced it or not is observed through their demonstration of the desired outcomes. The feedback on successful learning is relayed to the coachee to ground the fact it was successful.

Two types of grounding: understanding and skill To model motor skill coaching interactions, we propose there are two types of grounding at work– $grounded_{understanding}$ and $grounded_{skill}$. For a skill to become $grounded_{understanding}$, it has to be subject to communicative grounding requirements in the spirit of (Clark, 1996). However, this domain centers around communicating non-propositional information of physical movement which is only observable by *consistent demonstration of success* by the coachee– only then, after positive feedback by the coach will this become $grounded_{skill}$.

The reason we make this division is that it is possible the coachee could resolve all linguistic and intentional information in a description of an exercise but still not have grasped the skill, either in kind and in degree. The physical, embodied nature of learning a motor program schema means this representation is not straightforwardly translatable into symbolic means for information transmission but needs analogue values for trajectories, speeds, distances and pressures. It is clearly challenging for a coach trying to make this information common ground, both in terms of the dialogue acts and nonverbal actions they use and the *timing* they employ to do this.

3 Research questions

To investigate timing behaviour and grounding strategies in coaching interactions we conduct a

corpus study which focuses on the following overarching research questions:

q1 **Characterizing dialogue acts for motor skill coaching:** What is an adequate taxonomy of dialogue acts and non-verbal actions for the motor skill coaching domain, and how does it differ to existing task-oriented, tutorial and situated taxonomies?

q2 **Modelling dialogue context:** What type of information is in the *dialogue context* for the coach during the current coaching interaction as it unfolds in terms of the skill elements addressed so far? When does the coach take skills introduced to be understood by the trainee (grounded$_{understanding}$) and when does the coach take a skill to be part of the coachee's motor program schema (grounded$_{skill}$)?

q3 **Modelling decisions:** Which elements of the context influence the type of dialogue act the coach will use to address it? Specifically, does the status of the skill element as *given* (grounded$_{understanding}$) or *new* affect the dialogue act type, and does the status of the skill element in the common ground as having been routinized and mastered (i.e. being grounded$_{skill}$), affect the way the coach talks and acts non-verbally concerning the skill element?

q4 **Timing:** When do dialogue act and non-verbal actions happen with respect to coachee's skill attempts on a fine-grained time-line? As timing is a critical part of motor skill acquisition this becomes more vital than other tutorial domains.

4 Corpus Study: Timing and Grounding in Coaching Dialogue

To address the research questions we study a corpus of coaching dialogues where a coach trains a trainee in the exercise of a body-weight squat (a squat done with no weight or barbell)– see Figure 1. It is a simple and closed skill in that it is not subject to environment change (i.e. it is not an interactive sport) and can be practiced alone. However, it is an interesting skill from a dialogue perspective in that it is an exercise without a tangible outcome (such as scoring a goal in practicing taking football penalties), and relies on the expertise of a coach to provide feedback to indicate success.
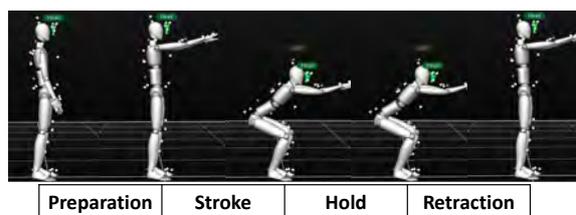

Figure 1: The four phases of a squat

We invited 8 participants to interact with 2 different professional fitness coaches (4 participants per coach). The average length of the sessions was approx. $4\frac{1}{2}$ minutes. The participants had various different levels of expertise with squats ranging from novice to doing it on a monthly basis. None were professional athletes but all partook in recreational exercise.[1] All sessions were in German and all participants were native German speakers.

4.1 Dialogue act and non-verbal action annotation

The dialogues were transcribed, translated and utterances were segmented into dialogue act units.[2] To address question q1 we did an initial analysis of 2 sessions, one from each coach, and created our annotation scheme for verbal and non-verbal acts in Table 1. The verbal dialogue acts specialized to this domain are as follows:

- *Instruction*[*directive*]: Imperative command to carry out a skill[3] (e.g. "Do three or four squats")

- *Instruction*[*attempt*]: Request to carry out a skill to the best of the participant's ability (see e.g. (1)).

  (1) **Coach:** also langsam so weit runterarbeiten .. wie du runterkommst
      so slowly go down .. as far as you can

- *Instruction*[*mentalize*]: Imperative to imagine something not present that will help with the skill, or to pay attention to the feeling of a particular part of the body during skill attempts (e.g. "Imagine there is a wall in front of you and you do not want to touch it")

- *Acknowledge*[*skill*]: Signal of recognition of a skill attempt with neutral sentiment, analogous to standard backchannels (e.g. 'Right' or 'Okay' said after a squat has been completed)

---

[1]We do not look at the effect of expertise or prior experience here, but intend to in future work.

[2]Like the slash-unit of (Meteer et al., 1995).

[3]This is analogous to the Action-Directive in (Core and Allen, 1997).

- *Adjust*: An instruction where the *degree* of an element of the skill is directed to be changed, as in (2). This is similar to (Raux and Nakano, 2010)'s Degree Error Correction, however an *Adjust* has a different notion of success– in a task-completion dialogue such as object selection there are binary notions of success and failure and reinforcement of good practice is not vital, whereas here the reaction from the corrector is especially important in terms of its motivational affect and long-term learning outcome.

  (2) **Coach:** Stell mal die Beine etwa schulterweit auseinander ...
  also noch [ei]nen Hauch weiter
  Plant your feet about shoulder-width apart .. a bit more

- *Repair[skill]*: An other-repair of a skill attempt which repairs misunderstanding of the intended outcome (rather than a linguistic other-repair repairing an utterance) or recovers from a lack of uptake via repeating or reformulating (e.g. "No, not that way, the other way").[4]

- *Explanation*: An explanation of why a certain skill is important (rationale) or consequences of mastery (e.g. "this will help your power transmission"), or an elaboration on an instruction with more descriptive detail (clarification).[5]

- *Feedback[positive]*: Evidence of approval that the skill is being performed well (e.g. "You kept your back nice and steady").

- *Feedback[negative]*: Criticism of the way the skill is being performed (e.g. "At the moment your knees are buckling a lot").

- *Commentary[self]*: Commentary description on the current action by the speaker (e.g. "I'm now tensing my stomach and back muscles").

- *Commentary[other]*: Commentary description on the current action by the addressee (e.g. "You're now getting into what we call the neutral position").

- *SetGoal*: Announcement that the session will turn its attention to a given skill element (e.g. "Let's focus on the width of your stance").

---

[4]This is similar to the Commission and Omission errors in (Raux and Nakano, 2010) described above.

[5]We encourage the attributes [rationale] and [clarification] to be added to the tag where annotators are confident which one it is, however we do not calculate agreement levels on this.

The other acts shown in Table 1 have standard definitions for dialogue act tags. The non-verbal acts specialized to this domain are:

- *SkillAttempt[preparation|stroke|hold|retract]*: An instance of an attempt at a phase of the overall target skill. For squats the 4 phases are as in Figure 1. The *preparation* phase can consist in several parts from adopting the stance to raising or crossing the arms. The *stroke* is the main phase focused on by our coaches. The *hold* at the lowest position is often short and occasionally too short to annotate at all. The quality of the squat has largely been determined before *retraction* back to the upright position.

- *Demonstration[exaggerated|positive|negative]*: Presentation of a movement either as it is meant to be done (*positive*), else an example of what not to do (*negative*). This can be *exaggerated* to emphasize an element of the skill.

- *Iconic[modelling|shaping]*: Gestures which represent objects, which are invariably parts of the body involved in the skill, either through using other body parts to represent them (*modelling*) or *shaping* their outline in the air.

- *Deictic[self|other|thirdperson|touch]*: Gestures used to refer to something in the environment. These include touching of the body in this domain, both one's own (*self*) or one's partner's (*other*) to point out physical details of movements– see Figure 2 C for an example of a $Deictic[self]$ gesture.

The other non-verbal acts are *Beat* gestures, *Head* movements (including nods) and *Discourse* gestures. The category *OtherAction* includes concurrent movement of the participants around the experimental space.

Skills Under Discussion In addition to these acts, for each act decision the annotators chose the particular element(s) of the motor skill being talked about. We will call these tags the *Skills Under Discussion*– their relationship to Questions Under Discussion models discussed in Section 5. The labels form a closed set and consist of values such as $StanceWidth$, $ArchedBack$ and other squat-specific skills. These approximate the *content* of the acts in this domain.

Annotation agreement and overall distributions Three annotators annotated the corpus and

we checked annotation agreement on verbal dialogue acts only for one transcript between two of them. We had an acceptable Cohen's κ of 0.69. The main source of disagreement was over what constituted *Commentary[other]* and what was *Feedback[positive]*– the boundary between these can be vague upon manual inspection.

As can be seen from the distributions in Table 1, as expected the verbal side of the interaction is heavily dominated by the coach. The majority of the coachee's verbal contributions were acknowledgements in the form of short backchannels indicating understanding of the coach's dialogue act.

30.0% of the coach's DAs are directive instructions, and adjustment acts such as in Figure 2 D are relatively common (81 total, 9.1% of the coach's DAs). Positive feedback is overwhelmingly more common than negative feedback (112 occurrences vs. 18) and the acknowledgement signal *Acknowledge[skill]* that a skill has been seen by the coach is also frequent (11.1% of coach DAs). We will discuss grounding strategies below.

For non-verbal acts, there is again asymmetry in the distributions, which is unsurprising given the domain– the coachee attempts a skill whilst the coach demonstrates it. What is more interesting however is the frequent use of iconic and deictic gestures, together constituting over 40% of the coach's non-verbal actions.

4.2 Timing of Adjust and Instruction acts

Having gained an insight into the interactions through dialogue acts, we now focus on the timing of these acts, and in particular, the timing of a coach's dialogue act production relative to a skill attempt by a coachee. As just described, *Adjust* moves are very common and have interesting timing properties in terms of turn-taking– see Figure 3 for a time-line of an adjusting event. The coach, constantly monitoring the coachee's action keeps incrementing his contribution with the adjunctive phrase 'noch ein bisschen' ('a bit further') until the coachee has achieved the desired foot stance. Notice the timing here is incredibly fine-grained, with the coachee's reaction being close to human reaction time ($\approx$0.2s) from the *middle* of the adjust instruction. Adjustments are inherently able to be concurrent with the non-verbal channel of the coachee's action, so tight coupling of the coachee's movement and the coach's feedback, although appearing like a normal dialogue turn-
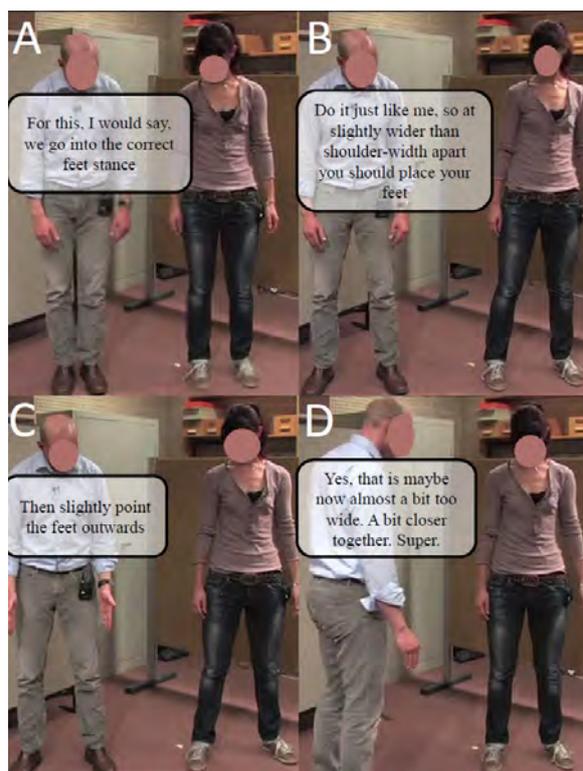

Figure 2: A typical coaching interaction: In A the coach commentates on where the foot position should be. In B the coach elaborates on the instruction. In C he uses a deictic gesture relative to his own body to show the correct width and in D he repairs the coachee's over distance and *adjusts* her stance until satisfied.

taking structure on first pass, can afford a great deal more overlap.

In non-adjustment instructions, timely reaction from both parties is also common. In fact, we observe coachees often *anticipate* the instructions even in these forward looking acts. To investigate this observation empirically we calculated the mean and standard deviations for the time between the end of an *Instruct[directive]* act and a skill attempt corresponding to the phase of the squat it is instructing. We do this both for instructions to enter the stroke phase such as 'geh nochmal runter' ('go down again') and for the retraction instructions like 'komm wieder hoch' ('come up') and find the means (vertical red lines) and probability density plots as shown in Figure 4. We find the mean interval from the end of the instruction to the start of the skill attempt was negative for the stroke phase at -0.274s (st.d.=1.204) and even more so for the retract phase at -0.410s (st.d.=0.510), meaning on average the coachees
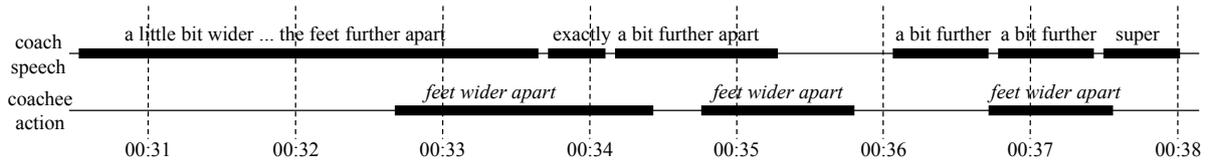
Figure 3: The fine-grained interaction between the coachee's actions and the coach's adjust moves

Dialogue Acts	Count (%)	
	Coach	Coachee
Instruct[directive]	267 (30.0)	1 (0.5)
Instruct[attempt]	41 (4.6)	0 (0.0)
Instruct[mentalize]	20 (2.2)	0 (0.0)
Acknowledge[skill]	99 (11.1)	0 (0.0)
Adjust	81 (9.1)	0 (0.0)
Repair[skill]	10 (1.1)	0 (0.0)
Explanation	80 (9.0)	0 (0.0)
Feedback[positive]	112 (12.6)	3 (1.4)
Feedback[negative]	18 (2.0)	0 (0.0)
Commentary[self]	32 (3.6)	8 (3.8)
Commentary[other]	2 (0.2)	0 (0.0)
SetGoal	17 (1.9)	0 (0.0)
Acknowledge[verbal]	32 (3.6)	150 (70.4)
Question	31 (3.5)	4 (1.9)
Answer	3 (0.3)	22 (10.3)
FloorManagement	21 (2.4)	2 (0.9)
StatementOther	11 (1.2)	17 (8.0)
Social	9 (1.0)	2 (0.9)
ClarificationRequest	5 (0.6)	4 (1.9)
Non-verbal Acts		
SkillAttempt	1 (0.2)	398 (75.2)
Demonstration	132 (21.5)	0 (0.0)
Iconic	134 (21.9)	1 (0.2)
Deictic	114 (18.6)	2 (0.4)
Beat	39 (6.4)	0 (0.0)
Head	25 (4.1)	28 (5.3)
Discourse	45 (7.3)	21 (4.0)
OtherAction	123 (20.1)	79 (14.9)

Table 1: Dialogue acts and non-verbal communicative actions marked up in our corpus with the percentage of total acts for each tag.

were moving on the Skill Under Discussion well before the end of the utterance. The coachees can take initiative and predict instruction completions easily, just as initiative is possible in other domains as described in Traum and Larsson (2003).

4.3 Grounding the motor program schema

In the outset in §2 we suggested there are two principal grounding mechanisms at work in motor skill coaching. From the coach's perspective, they must ensure not only that the coachee understands the meaning of the current Skill Under Discussion (i.e make it $grounded_{understanding}$), but that they use it to induce the appropriate motor program schema knowledge, whereupon they should provide feedback that this has been done successfully (make it $grounded_{skill}$).

We observe the coaches use various techniques to achieve the second grounding criterion. While purely instructing the coachee through directives can be effective initially, they must use other techniques if difficulties persist. In all of our sessions, one or two Skills Under Discussion were addressed at much greater length compared to the others because they were problematic for that particular coachee.

We investigate the effect of $grounded_{understanding}$ status of a Skill Under Discussion, a status we assume by virtue of the fact it has been addressed before by the coach and that the coachee has performed it to the best of their ability, for which they received acknowledgement or even positive feedback. We find that when a skill is re-referenced verbally there is a difference in dialogue act type used. We calculate the distribution of dialogue act types used based on whether the skill has been openly raised before or is new– see Table 2. While $Instruction$ acts are the most probable in both first mentions and subsequent mentions, their dominance is attenuated in the subsequent condition. $Adjust$ moves are one such way to attune the parameters of a skill as discussed, but also $Explanation$ becomes more frequent, as does $Feedback[positive]$ and $Instruction[mentalize]$ instructions. The $Acknowledge[skill]$ act, while having similar lexical and phonetic qualities to normal backchannel acknowledgements (e.g. 'okay'), is a grounding mechanism where the coach
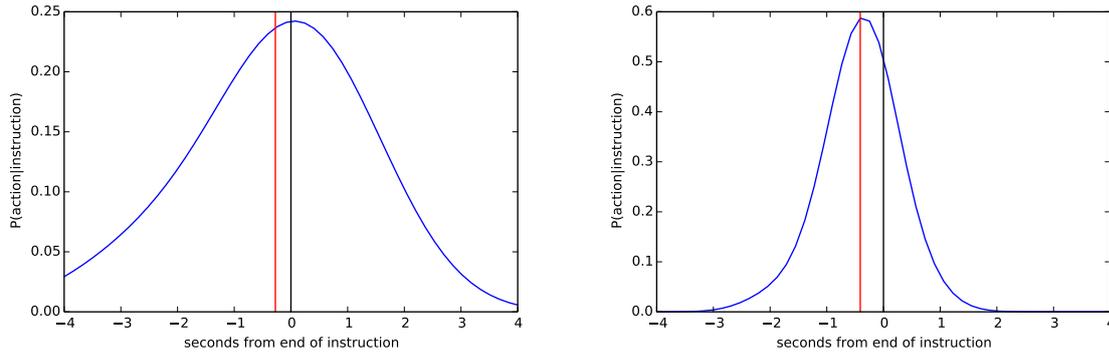
Figure 4: Anticipation in the uptake of an instruction both in the stroke (left) and retraction phase (right)

communicates a message to the effect 'I've seen you attempt this', however it is not strong enough evidence for the message 'I've seen you master this', and positive feedback is the way to convey this and make it grounded$_{skill}$.

There were a handful of mentalizing examples where the coach uses imagery and metaphor, such as in (3). These only occur in the longer sessions where a particular problem has been addressed numerous times.

(3) **Coach:** Versuch mal gedacht so ein bisschen Froschbeine zu machen das heisst wenn du runtergehst die Knie eher auseinanderzudrücke
Try to think about frog legs when your knees start getting closer together

In non-verbal behaviour, there are also differences with the gesture accompanying the dialogue acts which reference the skills. In explanations, not only are given skills likely to be accompanied by an overlapping gesture (87% new versus 97% given), also qualitatively there is a shift from *Deictic* gesture to *Iconic* gesture and *Demonstration*– see the bottom of Table 2. Analogously to the verbal case with direct instructions, directness through deixis is initially preferred to ensure grounding$_{understanding}$, but to achieve grounding$_{skill}$ several techniques of both personal demonstration and analogy with other objects and images is required.

5 Consequences for an Information State Model of Dialogue

We have argued it is useful to distinguish between grounding in the traditional sense along an understanding dimension and grounding a motor pro-

	% of acts about sub-skill	
Dialogue Acts	**1st ref.**	**Subsequent**
Instruct[directive]	46.2	25.9
Instruct[attempt]	17.3	4.6
Explanation	8.7	15.8
Feedback[positive]	6.7	11.8
Adjust	5.8	18.2
SetGoal	2.9	0.7
Question	2.9	2.4
Feedback[negative]	2.9	3.6
Commentary[self]	1.9	1.9
Instruct[mentalize]	1.9	4.1
Acknowledge[skill]	1.0	7.4
Commentary[other]	1.0	1.0
Repair[skill]	1.0	1.9
Non-verbal Acts		
Iconic	30.4	39.4
Deictic	26.1	15.2
Demonstration	8.7	15.2
Discourse	4.3	6.1
Beat	2.2	6.1
OtherAction	15.2	15.2
None	13.0	3.0

Table 2: Different dialogue acts and non-verbal acts used when a skill element is referred to the first time and subsequently. Note the non-verbal acts are only those overlapping Explanations here

gram schema, the latter being due to the coach's goal to transfer knowledge of a physical movement to the trainee. We show skill elements behave similarly to discourse referents in that their given versus new status affects the dialogue act type with which they are re-referenced. This puts the requirement on an information state model of dialogue that what is under discussion is not al-

ways propositional material, but internal representations of action sequences. Instead of issues being *resolved* like a QUD-based model (Traum and Larsson, 2003; Ginzburg, 2012), the coachee must *demonstrate* their acquisition of a motor program schema. While one could posit that demonstrations evidence propositions such as $CanDo(x)$ for a skill x, which 'answer' whether $CanDo(x)?$, bi-valued propositions may not be a useful analogy given the real values and degrees that need to be parameterized in skill representations.

Another requirement arising from our findings, and a general short-coming of the traditional Information State update approaches is the lack of timing information in the information state, which in real-time situated dialogue such as coaching dialogues is crucial. In situated multimodal dialogue interaction, the state needs to represent time to account for the plethora of overlap both inter-modally (speech and gesture occurring with various degrees of synchronization within the same agent's behaviour) and interactively (speech and gesture of different agents overlap or synchronise with one another to various degrees). In both cases the nature of the synchronization is important for meaning construction, a fact currently exploited more by the Virtual Agents community (Kopp et al., 2014) than by dialogue theorists and semanticists– however see Lücking et al. (2013).

One theoretical and practical step we are exploring is using an established temporal reasoning system, Allen's interval algebra (Allen, 1983), which describes the possible relations two temporal events can have to each other, with the primitives, or *base relations* as in Figure 5. According to the assumption of the classical information state approach, for two contiguous dialogue acts by two different agents which are related (i.e. a 'minimal pair' of dialogue acts) A and B, their relative timing would be represented $A < B$ or A m B (A ends completely before B begins, either with no gap or contiguously). However, we argue that if A was a 'forward-looking' move, such as an instruction, and B was a 'backward-looking' move related to A such as a skill attempt, all 13 Allen relations between A and B are possible, even $A > B$ and A mi B. To model coupling between two or more multimodal dialogue acts as shown here, an approach using the constraints of this temporal algebra permitting overlap and anticipation between acts and intra-act level increments is required.

Relation	Illustration	Interpretation
$X < Y$ $Y > X$		X takes place before Y
X m Y Y mi X		X meets Y (*i* stands for *inverse*)
X o Y Y oi X		X overlaps with Y
X s Y Y si X		X starts Y
X d Y Y di X		X during Y
X f Y Y fi X		X finishes Y
$X = Y$		X is equal to Y
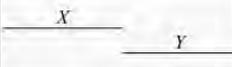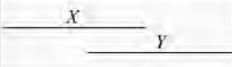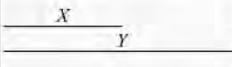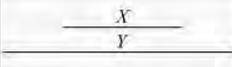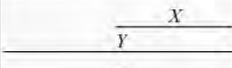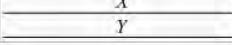
Figure 5: (Allen, 1983)'s interval algebra for describing the thirteen possible temporal relationships between two observed intervals

6 Conclusion

We have presented a corpus study with a novel dialogue taxonomy for motor skill coaching dialogues. We argue this puts requirements on formal models of situated dialogue, including fine-grained shared time representations, and characterizing what is under discussion and in the common ground– in these kind of domains this is generally not questions and propositions, but skills and their desired and observed parameters. In future work we wish to analyze skill referencing completely multimodally, rather than in the verbal sense with accompanying non-verbal acts as we do here[6] and also investigate how the grounding status, both grounded$_{understanding}$ and grounded$_{skill}$, of skills under discussion generalizes to other learning domains.

[6]Thanks to a reviewer for this suggestion.

References

James F Allen. 1983. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843.

Kristy Elizabeth Boyer, Mladen A Vouk, and James C Lester. 2007. The influence of learner characteristics on task-oriented tutorial dialogue. *FRONTIERS IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS*, 158:365.

Kristy Elizabeth Boyer, Robert Phillips, Michael Wallis, Mladen Vouk, and James Lester. 2008. Balancing cognitive and motivational scaffolding in tutorial dialogue. In *Intelligent Tutoring Systems*, pages 239–249. Springer.

Kristy Elizabeth Boyer, William J Lahti, Robert Phillips, MD Wallis, Mladen A Vouk, and James C Lester. 2009. An empirically-derived question taxonomy for task-oriented tutorial dialogue. In *Proceedings of the Second Workshop on Question Generation*, pages 9–16.

Herbert H Clark. 1996. *Using language*, volume 1996. Cambridge university press Cambridge.

Mark Core and James Allen. 1997. Coding dialogs with the damsl annotation scheme. In *AAAI fall symposium on communicative action in humans and machines*, pages 28–35. Boston, MA.

Jonathan Ginzburg. 2012. *The Interactive Stance: Meaning for Conversation*. Oxford University Press.

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. 2014. An architecture for fluid real-time conversational agents: integrating incremental output generation and input processing. *Journal on Multimodal User Interfaces*, 8(1):97–108.

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.

Andy Lücking, Kirsten Bergman, Florian Hahn, Stefan Kopp, and Hannes Rieser. 2013. Data-based analysis of speech and gesture: The bielefeld speech and gesture alignment corpus (saga) and its applications. *Journal on Multimodal User Interfaces*, 7(1-2):5–18.

Terry McMorris. 2014. *Acquisition and performance of sports skills*. John Wiley & Sons.

M. Meteer, A. Taylor, R. MacIntyre, and R. Iyer. 1995. Disfluency annotation stylebook for the switchboard corpus. ms. Technical report, Department of Computer and Information Science, University of Pennsylvania.

Antoine Raux and Mikio Nakano. 2010. The dynamics of action corrections in situated interaction. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 165–174. Association for Computational Linguistics.

Emanuel A. Schegloff, Gail Jefferson, and Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language*, 53(2):361–382.

Richard A Schmidt. 1975. A schema theory of discrete motor skill learning. *Psychological review*, 82(4):225.

David R Traum and Staffan Larsson. 2003. The information state approach to dialogue management. In *Current and new directions in discourse and dialogue*, pages 325–353. Springer.