Collaborating on Utterances with a Spoken Dialogue System Using an ISU-based Approach to Incremental Dialogue Management Okko Buß, Timo Baumann, David Schlangen Department of Linguistics, University of Potsdam

I. Abstract

- Silence-end-pointed SDS: user must finish speaking before system planning and generation can begin.
- Human-human dialogue: actions occur during or immediately at the end of speech input, e.g. backchannels, interruptions, delivery in instalments. This is lost in end-pointed systems.
- We explore using trial intonation in collaborative utterance building. Present and evaluate incremental dialogue manager capable of modelling such behaviour.
- Highlights of system:
 - reacts to reference resolution problems, and (un-) certainty of user, as signalled by prosody • can execute requests concurrently; raises implicit
 - "correct?" question
 - judged more human-like and more reactive than non-incremental counterpart

II. Collaboration on Utterances

- Shape of an utterance determined not only by its speaker, but also by her addressee.
- Speakers monitor addressee responses online, to determine whether successfully grounded.
- Utterance itself may contain cues for grounding responses. E.g. (from Clark 1996)
 - A: A man called Annegra?-
 - B: yeah, Allegra
 - A: Allegra, uh, replied and, uh,...

A uses try marker (Schegloff 1979): "questioning upward intonational contour, followed by a brief pause", cueing response from B, which ultimately determines shape of the overall utterance.

III. System, Domain

puzzle tiles (Pentomino):



IV. System, Overview



VI. Evaluation

- Overhearer evaluation against 800ms end-pointed system w/o concurrent actions. • 30 minutes of interactions from both setups
- recorded (discarded 10% of "outlier" interactions.) • Interactions from the incremental setting
- significantly shorter (t-test, p<0.005).

V. System, Dialogue Management

- grounding state.

Intera		
Modu	lle	
User:	Τŀ	
	th	
Floor:	EO	
Svc ·	orn	
<i>зу</i> з. лм.		
AM.	aone tho	
NLU.	{TI	
<i>S</i> ys.	'''''	
Al¶I.	aone	
User:	yes	
NLU:	۔ {د	
User:	tak	
NLU:	{ta	
Sys:	cha	
	sele	
AM:	done	
	done	

VII. References

Please see paper in proceedings.



• IS: question stack (iQUD) and a todo list. Update rules triggered by events from NLU, AM, FloorTracker, and changes to the IS.

• iQUD (variant of Ginzburg's (1996) QUD): collects relevant sub-questions, consisting of a semantic slot; a relevant non-linguistic action (RNLA); and

• NLU events downdate sub-questions, triggering rule sending RNLA to AM via todo list. Upon completion, AM rules update grounding state. FloorTracker rules cue system utterances.

ction/	ISU Rules	S
Output		Information States
he cross in	AddNegBackChannel	1
he corner		QUD:
OT_RISING		<pre>{< a (1 action=A=take; 2 prepare(A) ; 3 U), (4 tile=T ; 5 highlight(T) ; 6 U), (7 ; 8 execute(A,T) ; 9 U) > < b (10 action=A=del ;11 prepare(A) ;12 U), (13 tile=T ;14 highlight(T) ;15 U), (16 ;17 execute(A,T) ;18 U) >} TODO: <prompt(bcneg)< pre=""></prompt(bcneg)<></pre>
m 🛛	ActionPerformed	
e(BC)	UpdateGrounding	
e red one	RelevantNewInfo	<pre>{< a (1 action=A=take; 2 prepare(A) ; 3 U),</pre>
13	ActionPerformed	(4 tile=t1 ; 5 highlight(t1) ; 6 RD),
hlights t1	UpdateGrounding	< b (10 action=A=del ;11 prepare(A) ; 12 U),
e(highlight)	AddImplicitYN	<pre>(13 tile=t1 ;14 highlight(t1) ; 15 RD), (16 ;17 execute(A,t1) ; 18 U) >} {< c (19 correct=Y=yes ; 20 ; 21 I)> d (22 correct=N=no ; 23 undo(highlight) ; 24 I) > TODO: <></pre>
S	RelevantNewInfo	
correct}	ResolveAlternativ UpdateGrounding	<pre>Ves QUD: {< a (1 action=A=take; 2 prepare(A) ; 3 U), (4 tile=t1 ; 5 highlight(t1) ; 6 RDA), (7 ; 8 execute(A,t1) ; 9 U) > (10 action=A=dol :11 prepare(A) : 12 U)</pre>
ke that		(13 tile=t1 ;14 highlight(t1) ; 15 RDA),
ake}	RelevantNewInfo AddTodo	(16 ;17 execute(A,t1) ; 18 U) >} TODO: <>
anges hand	ResolveAlternativ	ves
lects tile t1	ResolveQUD	QUD: $\int c = (25 \text{ connect} - V - v \text{ or } 26 + 27 \text{ T})$
e(prepare)	ActionPerformed	f (28 correct=N=no; 29 undo(prepare); 30 I) >
e(execute)	AddImplicitYN	<pre>{< g (30 correct=Y=yes ; 31 ; 32 I)> h (33 correct=N=no ; 34 undo(execute) ; 35 I) > TODO: <></pre>

• Judgments on helpfulness (1), human-likeness (2) and reactivity (3), Likert-scale (8 subj./34 randomly selected recordings). Incr. sys rated higher on 2 and 3 (Wilcoxon rank-sum test; p=0.04, p>0.005).

