

Engagement and Affects in WikiTalk Interactions

Kristiina Jokinen

University of Helsinki /
University of Tartu

Where do I come from?





University of Helsinki

- Oldest in Finland
 - Founded 1640 in Turku, moved to Helsinki in 1828
- Bilingual: Finnish and Swedish; (English)
- Over 38 000 students (women almost 24 000)
- Staff 7800; professors & researchers 3800
- 11 faculties, 300 different subjects and course programmes
- <http://www.helsinki.fi/>



30/09/2015 Aix-en-Provence



University of Tartu

- Founded in 1632 by King Gustavus Adolphus of Sweden
- 9 faculties and 4 colleges
- 70 bachelor, 80 master and 35 doctoral study programmes
- 17,000 students (including over 800 international students)
- 3,800 employees, including 1,800 academic employees (190 professors)
- <http://www.ut.ee>



30/09/2015 Aix-en-Provence 4

Structure of the talk

- Spoken Interaction
 - Constructive Dialogue Modelling
- Engagement in interaction
 - Feedback and nodding
 - Turn-taking and eye-gaze
- Human-robot interaction
 - Open-domain dialogues
 - Topic and NewInformation
 - Affect and sentiment analysis

30/09/2015

Aix-en-Provence

5

Spoken Interactions

- **Cooperative** social activity
which manifests itself in
 - Accurate timing and **alignment**
of gaze, gestures, body, and speech
concerning
 - Turn-taking, feedback and topic **management**

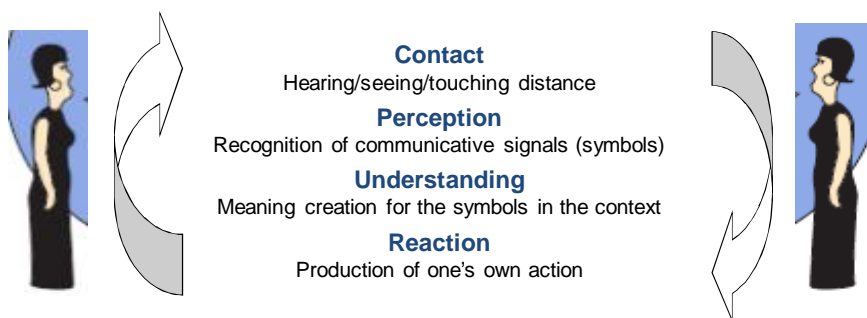
30/09/2015

Aix-en-Provence

6

Constructive Dialogue Modelling

- Rational agents coordinate their actions, observe each other and the environment
- React to situations according their beliefs, intentions and understanding of the context
- Construct shared knowledge to achieve (communicative) goals, maintain social bonds, express emotional state (Allwood, 1976; Jokinen 1996, 2009)
- Multimodal signalling of the basic enablements: gesturing, eye gaze, body posture
- Rational interaction: cooperation, synchrony



30/09/2015

Aix-en-Provence

7

Feedback

- Feedback giving is mostly vocal
 - Backchannelling: *ah, ah-ah, mm, uh-huh*
 - Lexical feedback: *yes, no, right, awful, super*
 - Repetition: last/core words uttered by the partner
 - Reformulation: express meaning of the message
 - Relevant continuation
- Multimodal feedback
 - Head movement (84%): nods, shakes, tilt...
 - Facial expressions (9%): smile, rising eyebrows,...
 - Hand gestures (4%)
 - Body posture (3%)
- Explicit and implicit feedback

30/09/2015

Aix-en-Provence

8

Nodding and feedback

- Navarretta, Ahlsén, Allwood, Jokinen & Paggio (2012). Feedback in Nordic First-Encounters: a Comparative Study. *Proceedings of the Language Resources and Evaluation Conference (LREC-2012)*. Istanbul, Turkey.
- Toivio & Jokinen (2012). Multimodal Feedback Signaling in Finnish. In *Proceedings of the Human Language Technologies – The Baltic Perspective*. Published online with Open Access by IOS Press, distributed under the terms of the Creative Commons Attribution Non-Commercial License. doi:10.3233/978-1-61499-133-5-247
- Cultural differences in nodding:
 - Repeated nods vs single nods:
 - Finns tend to use single nods as opposed to Danes who use repeated nods
 - "invisible" feedback
 - Down-nods vs. up-nods
 - Acknowledgement of the expected information in the shared context (down) vs. acknowledgement of unexpected, novel information in the shared context (up)
 - However, Swedes tend to use up-nods more than Finns who use them more than Danes
 - Cf. user interruptions in WikiTalk : the robot uses up-nods

30/09/2015

Aix-en-Provence

9

Turn-taking



30/09/2015

Aix-en-Provence

10

Turn-taking and eye-gaze

- Prosody usually signals turn-taking
 - Hirschberg, Swerts, Edlund, Gustafsson et al.
 - Lowering of voice -> give turn
 - Raise voice -> take turn
- Eye-gaze shows where the person's attention is focused:
 - which participant the speaker is talking to
 - if the partner has understood/is interested in the speaker's topic
 - if the partner is present/reliable/bold (cultural politeness codes)
- An effective means to coordinate turn-taking
 - Mutual gaze as an agreement of turn-taking (Kendon, 1967)
- In multiparty conversations, also head-movement is an important signal of attention and the speaker's intention to take turn

30/09/2015

Aix-en-Provence

11

Turn-taking and eye-gaze

Jokinen, Furukawa, Nishida & Yamamoto (2013). Gaze and Turn-taking behaviour in Casual Conversational Interactions. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, Special Section on Eye-gaze and Conversational Engagement, 3:2.

Collection of eye-gaze data (Doshisha University, Kyoto)
Manual annotation using a modified MUMIN scheme

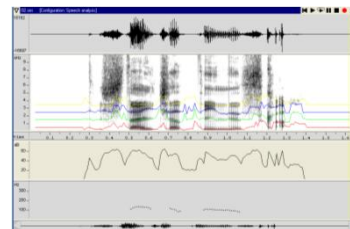


For each utterance unit extract 8 speech features:

- maximum, minimum, and mean values of F0 and intensity (power)
- difference between the respective minimum and maximum values of F0 and power
- Normalize with respect to each speaker

Utterance vector: 12-dimensional integrated vector

- manually annotated dialogue act (1)
- eye-gaze features (3)
- the speech signal features (8 features)



Dact	Gz1	Gz2	Gz3	Sp1	Sp2	Sp3	Sp4	Sp5	Sp6	Sp7	Sp8
------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

30/09/2015

Aix-en-Provence

12

How well does eye-gaze signal turn change vs. turn hold

Jokinen, Harada, Nishida & Yamamoto (2010). Turn-alignment using eye-gaze and speech in conversational interaction. *Proceedings of Interspeech 2010*, pp. 2018-2021.

Feature		Precision	Recall	F-measure
DA + Speech		0.669	0.655	0.581
DA + Speech + Eye-gaze		0.752	0.755	0.753

Total	x	y	← classified as	Total	x	y	← classified as
68	65	3	x=turn-give	68	51	17	x=turn-give
42	36	6	y=turn-hold	42	16	26	y=turn-hold
DA+Speech				DA+Speech + eye-gaze			

Eye-gaze helps especially to distinguish turn-holds:

- pauses and gaze aversion => hesitation, turn-hold
- pauses and gaze to partner => turn-give

30/09/2015

Aix-en-Provence

13

Social Robot Applications

- WikiTalk (Wilcock & Jokinen, 2013, 2015)
 - **Knowledge** from Wikipedia
 - **Multilingual navigation** (any language with speech technology, e.g. Sámi)
 - <http://vimeo.com/62148073>
 - Yle Ođđasat: <http://estreats27.rssing.com/browser.php?indx=14371100&item=1298>
- MoroTalk (Laxström, Wilcock & Jokinen, 2014)
 - News reading application for the Aamulehti newspaper
 - <https://www.dropbox.com/sh/dzsb9dy1m46kqxp/AACNwVjeNNeMp7JtVqWLSx9Ea/2014-09-24%2009.06.49.mp4?dl=0>
 - In collaboration with the *Koti* ("Home") project: well-being and health-care of elderly in future smart homes
 - **care-taking and companionship with social robots**



Euroopan unioni
Euroopan aluekehitysrahasto



Saastamalan
Koulutuskeskuntayhtymä
sasky.fi



30/09/2015

Aix-en-Provence

14

WikiTalk video (eNTERFACE, Metz 2012)

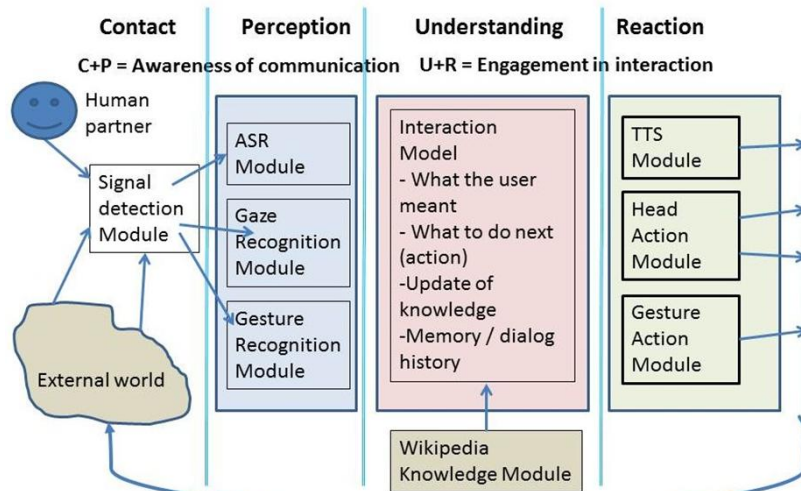


30/09/2015

Aix-en-Provence

15

CDM and Nao WikiTalk



30/09/2015

Aix-en-Provence

16

Open-Domain Dialogues

- Example: Talking about Shakespeare
 - Using information from the web
 - Wikipedia articles on any topic
- If the topic can be identified:
 - Get Wikipedia article about Shakespeare
 - Start reading out the first paragraph
- How to continue the conversation?
- How to switch topics?

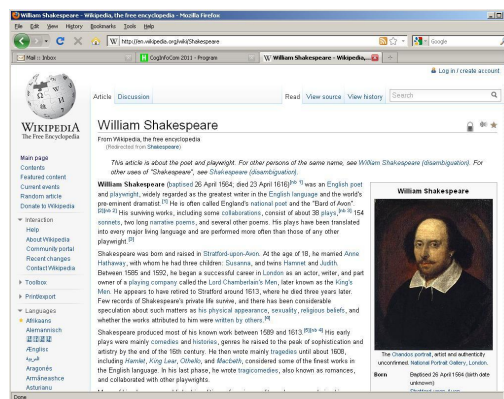
30/09/2015

Aix-en-Provence

17

How to Continue?

- Split text into chunks
 - Speak a chunk, pause for feedback
 - Is the human interested or not?
- Feedback may be non-verbal
 - Gaze, facial expression, body language
- If human seems interested:
 - Continue same topic, next chunk



30/09/2015

Aix-en-Provence

18

How to Switch Topics?

- Dialogue, not monologue
 - Don't read out entire article ("talk less!")
- Hypertext links
 - Links identify possible topic shifts
 - Topic navigation by clicking on links
 - Clicking shows human's positive interest
- Need to know the current topic
 - What is talked about vs. what is said about it
 - Need to keep track of topic shifts
- NewInfo
 - Newsy, informative vs. old, known or expected
 - Gives some new information about Topic
 - Dialogue response is based on NewInfo

30/09/2015

Aix-en-Provence

19

Topic and NewInfo

- What about Shakespeare? What can you tell about him?

<i>Shakespeare</i>	<i>was born and raised in Stratford-upon-Avon</i>
Topic	Comment
Old (background)	NewInfo

- What about Shakespeare? Where was he born and raised?

<i>Shakespeare</i>	<i>was born and raised</i>	<i>in Stratford-upon-Avon</i>
Topic	Comment	
Old (background)	NewInfo	

30/09/2015

Aix-en-Provence

20

Sentiment Analysis

- Texts are classified into positive, negative and neutral categories based on the words associated with negative, neutral or positive sentiments
- Opinion mining from text corpora
 - Various businesses and companies
 - feedback on products, customer preferences, business trends, market development
 - Governments and policy makers
 - favorable/unfavorable attitudes of the audience towards initiatives and view-points
- Early work:
 - polarity classification of documents concerning product reviews, movie rankings, and restaurant reviews (e.g. Pang & Lee 2004, 2005; Snyder & Barzilay 2007)
- Recent work:
 - opinion mining from social media texts with unconventional text characteristics
 - shortenings, abbreviations, spelling mistakes, special character strings, etc.
 - big data processing techniques due to the high volume of data

30/09/2015

Aix-en-Provence

21

Sentiment analysis and interaction

- Can the standard sentiment analysis techniques be applied to conversational data?
 - Transcribed dialogue texts resemble tweets or short SMS messages rather than long documents
 - However, conversations are unique in that face-to-face conversations are highly situational
 - Previous work used speech signal properties and dialogue context to recognize emotions (Lee et al. 2002; Devillers et al. 2003; Litman and Forbes, 2003)
- Can we infer the speaker's emotional state from the utterance content, or the speaker's attitude towards the utterance topic?
 - We don't aim at the participants' emotion analysis!
 - Rather: can the topics be mapped onto the speaker's affective state, and onto evaluation of the interaction



30/09/2015

Aix-en-Provence

22

Preliminary work

- Collaboration with Birgitta Ojamaa and Kadri Muischenk (University of Tartu)
- Analysis of spoken utterance transcripts and estimating the speaker's attitude towards the whole dialogue on the basis of their opinions expressed by utterances
- Data: MINT (Multimodal INTERaction) project (Jokinen and Tenjes 2012).
 - First encounter conversations: the participants meet for the first time
 - 23 encounters, each about 8 min long, 2902 utterances
- Simple decision:
 - positive/negative sentence contains mostly positive/negative words and no sentence negation, otherwise neutral sentence

30/09/2015

Aix-en-Provence

23

Results and participant experience

- Positive utterances: 20%
- Negative utterances: 1,3%
- Ambiguous utterances : 0,001%
- Neutral utterances: 79%
- Manual evaluation of 200 utterances:
 - Overall correctness 85%
- Questionnaire after each encounter:
 - Did you find the interaction enjoyable, friendly, impressive, nice, interesting, relaxed, anxious, natural, happy, tense, awkward, angry?
 - Participants found the interaction enjoyable, interesting and happy (Jokinen, 2012)
 - This accords with sentiment analysis

30/09/2015

Aix-en-Provence

24

Further work

- Statistical methods
- Speech signal analysis
- Context and additional conversational cues
- Multimodal features
- Methodology and terminology

30/09/2015

Aix-en-Provence

25

Thank you! Kiitos! Tänan!

Questions?
Comments!



30/09/2015

Aix-en-Provence

26