

# Uncertainty in Spoken Dialogue Management

Kristiina Jokinen  
University of Tartu  
and  
University of Helsinki


## Natural Interaction

**Goal oriented**

- speakers have intentions
- task

**Multimodal**

- speech, gesture, face, posture



**Symbolic**

- learnt conventions
- communicative signs, words

**Affective**

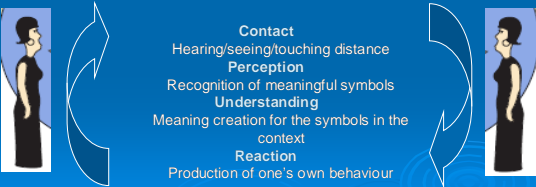
- affinity, social bonds
- emotions, attitudes

**Rational agents**

- coordination of action
- cooperation
- grounding = build shared knowledge

## Constructive Dialogue Modelling

➤ The agents monitor each other and each other's actions in the communicative situation and react to the situation according to their beliefs, intentions and interpretation of the situation, to build shared knowledge and achieve a goal (Jokinen, 2009)



**Contact**  
Hearing/seeing/touching distance

**Perception**  
Recognition of meaningful symbols

**Understanding**  
Meaning creation for the symbols in the context

**Reaction**  
Production of one's own behaviour

Oct 8 2010
Baltic HLT / K. Jokinen


## Hesitation/Uncertainty Situations

- Jokinen & Allwood (2010)
- lack of own ability to continue
  - knowledge, skills
- lack of permission to continue
  - situational issues
- lack of willingness to continue
  - attitude


Oct 8 2010
Baltic HLT / K. Jokinen

## Examples

*it's just a...*



*Make a distance.... probably .... I don't know*



Oct 8 2010 Baltic HLT / K.Jokinen

## Hesitation-related phenomena

- Fillers: filled pauses, discourse markers, editing terms, parentheticals
- Disfluencies
- Self-corrections (repairs)
- Retractions: reformulation or restart of one's utterance
- Non-verbal aspects: hand gestures, body posture

Oct 8 2010 Baltic HLT / K.Jokinen

## Perception of hesitation signals

- Hesitation markers (*uhh, umm*)
- Pauses, silence
- Slower speaking rate
- Higher pitch
- Fundamental frequency F0 rises before pauses that occur in major syntactic boundaries
- Gesturing
- Compensatory pattern Carlson & Gustafson (2006)
  - F0 contours, pausing, retardation, creaky voice, syntax
  - the *total duration increase* counts rather than the contribution by syntactic or prosodic factors as such

Oct 8 2010 Baltic HLT / K.Jokinen

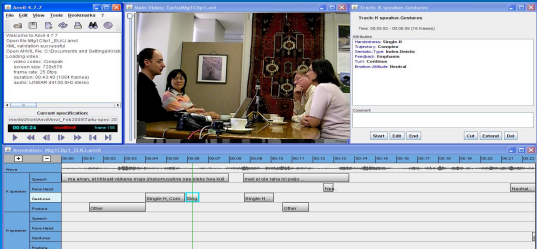
## Multi-level Hybrid Method

- Top-down analysis of the collected data
  - at different meaning levels
    - words, syntactic phrases, dialogue acts, gestures, face, posture,...
  - different tagging schemes: AMI, MUMIN, standardisation
  - manual annotation of what the human observe
- Bottom-up analysis of the collected data
  - at different signal levels:
    - speech, eye-gaze, face, gesture recognition
  - different technical constraints: accuracy
  - Automatic annotation of what "happens"
- Correlations and classifications between
- Crossing points

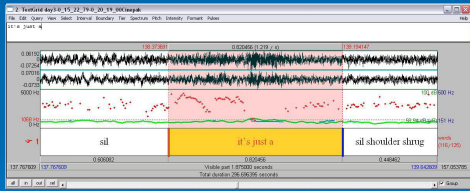
Oct 8 2010 Baltic HLT / K.Jokinen

## Annotation – human observations

- Speech, facial expressions, gestures, body posture
- Using the Anvil annotation tool



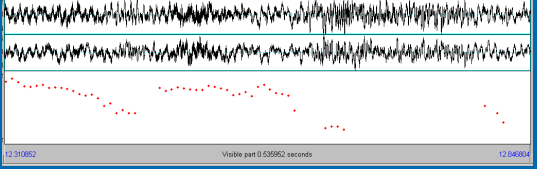
## Praat Analysis



F0 variation      pitch      intensity

Oct 8 2010      Baltic HLT / K.Jokinen

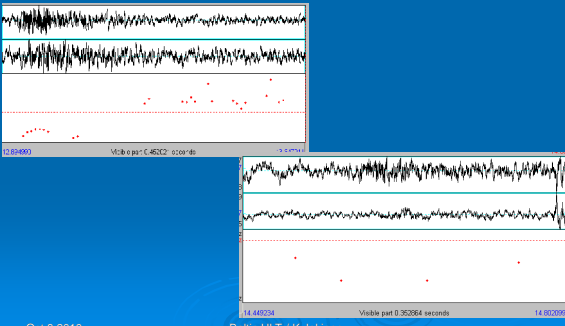
## F0 for *it's just a...*



Fundamental frequency raises before pauses at major syntactic constructs, but lowers if a pause occurs in the middle

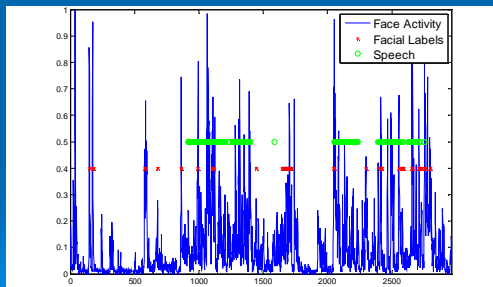
Oct 8 2010      Baltic HLT / K.Jokinen

## F0 of *probably & I don't know*



Oct 8 2010      Baltic HLT / K.Jokinen

### Face activity (blue peaks) correlates with speech activity (green circles) and manual facial gesture



Oct 8 2010

Baltic HLT / K.Jokinen

Thanks to Stefan Scherer

## Conclusions

- Verbal expressions of hesitation and uncertainty are accompanied by facial expressions and gesturing that help the partners to understand the underlying reasons for hesitation
- Situational and attitudinal hesitation seems to be accompanied by large, specific gesturing which carries social conventions of symbolic gestures
- Hesitation and uncertainty also used in the coordination of interaction and the activity that the speakers are involved in
- Semantic theme of hesitation and uncertainty: non-continuation of the current conversational topic

Oct 8 2010

Baltic HLT / K.Jokinen

## Holistic View of Interaction

- Communicative signals are used as means to manage social situation in which the agents find themselves, as a reaction to conversational understanding and they form communicative patterns rather than function as individual signs of communication
- What is the context (levels, activities, culture...)
- Cf. Compensatory pattern
  - Deviations of the expected temporal patterns

Oct 8 2010

Baltic HLT / K.Jokinen

## Future Work

- Relation between speech and gesture parameters
  - Temporal correlations
- Segmentation: what is the smallest unit
- How the speakers learn to observe hesitation signals
- Intercultural comparison: culturally accepted hesitation markers vs. interpretation of these signals (e.g. shoulder shrug)

Oct 8 2010

Baltic HLT / K.Jokinen

### Future Views in the context of Baltic HLT

- Natural language interaction extended to the whole communication situation
  - Integrates language and speech technology
- Corpus collection
  - Speech data: individual words to read speech to conversational speech
  - What type of data, activity, what kind of equipments
- Analysis
  - Annotation levels (tags for dialogue acts, gestures, etc. on pragmatic level)
  - Comparison of interaction strategies
  - Possible use of speech recognisers, morphological and syntactic parsers; also face and gesture recognition

Oct 8 2010      Baltic HLT / K. Jokinen

### Thank you!

Oct 8 2010      Baltic HLT / K. Jokinen