ROYAL INSTITUTE
OF TECHNOLOGY

User feedback in human-robot interaction: Prosody, gaze and timing

Gabriel Skantze, Catharine Oertel, Anna Hjalmarsson

Dept of Speech Music and Hearing
KTH, Stockholm, Sweden

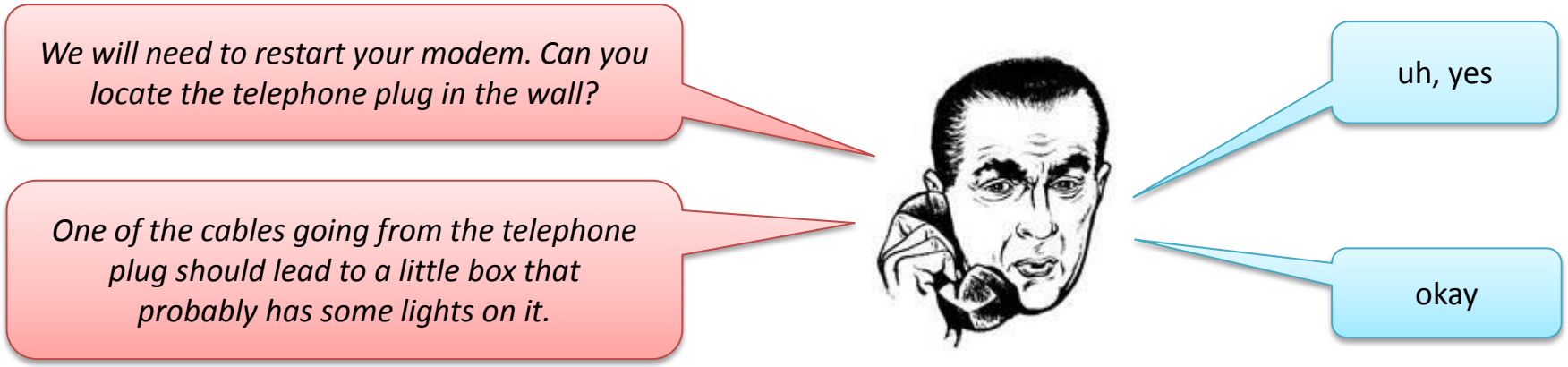
Feedback in dialogue systems

- Many studies on how the system should give feedback
  - Explicit/Implicit confirmation
  - Backchannels
- Not so many on how to interpret user feedback
- Feedback is a very broad (and vague) term
  - clarification requests, reprise fragments, etc, etc
- We will focus on **acknowledgements**

> *Acknowledgements are utterances consisting of short phrases such as "okay", "yes", and "uh-huh", that signal that the previous utterance was understood without necessarily signaling acceptance.*
>
> DAMSL (Allen & Core, 1997)

Two functions of acknowledgements

- Signal task completion



Example from Boye (2007)

- Express (un)certainty
  - Uncertainty in human-computer utterances (Forbes-Riley & Litman, 2011)
  - Uncertainty in human-human acknowledgements (Lai, 2010; Ward, 2004)

Human-Robot Map Task

Example video

Research question

- How does the form of the user's acknowledgements relate to their communicative function?
- Functions:
  - **Task completion**
  - **Uncertainty**
- Form:
  - **Lexical item** (*okay, yes, yeah, etc.)*
  - **Prosody** (pitch, duration, intensity)
  - **Response time**
  - **Gaze**

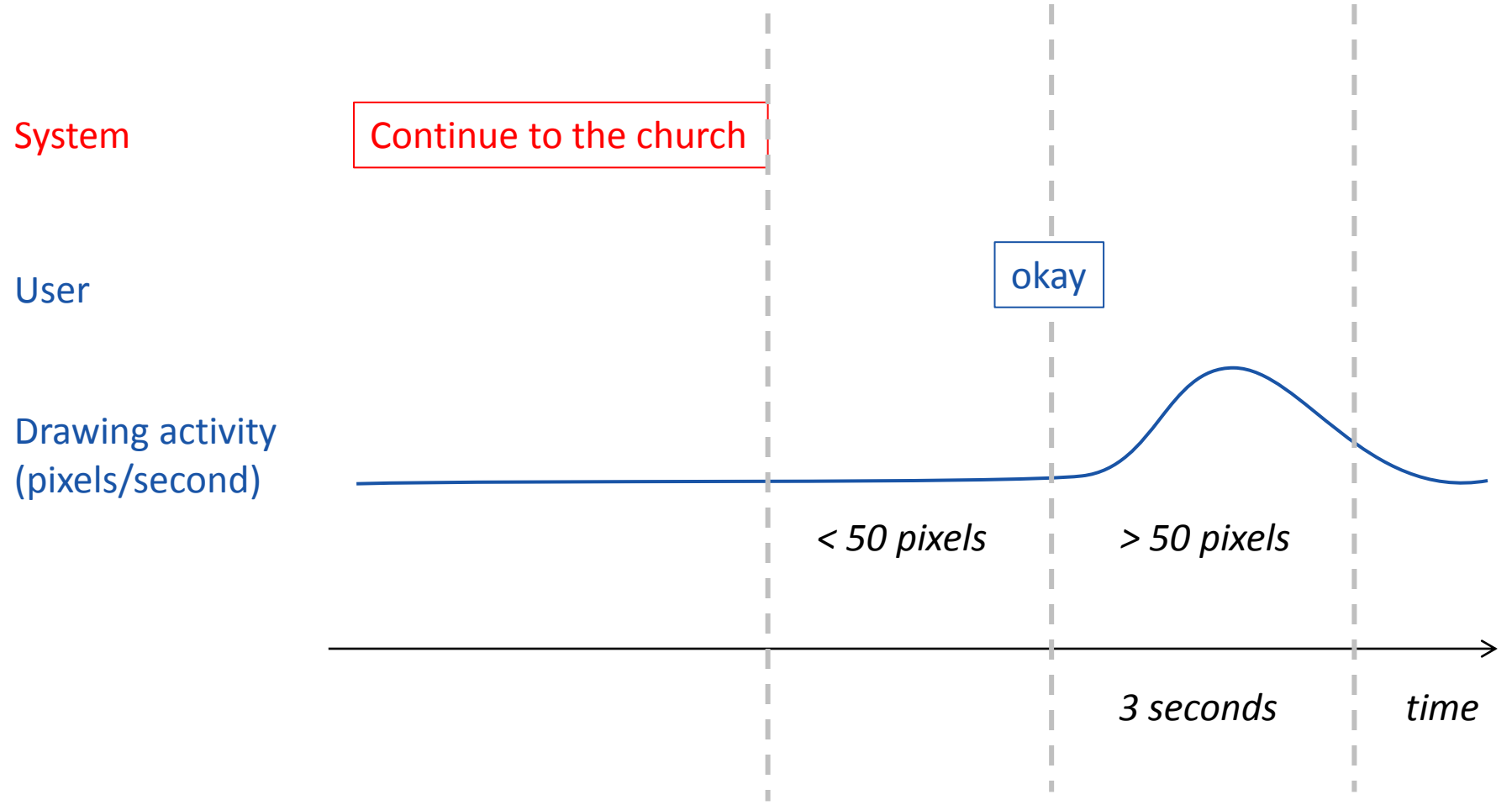
Data collection

- 24 subjects
- 6 dialogues per subject
- 144 dialogues
- ~1500 acknowledgements
  - *okay, yes, yeah, mm, mhm, ah, alright*

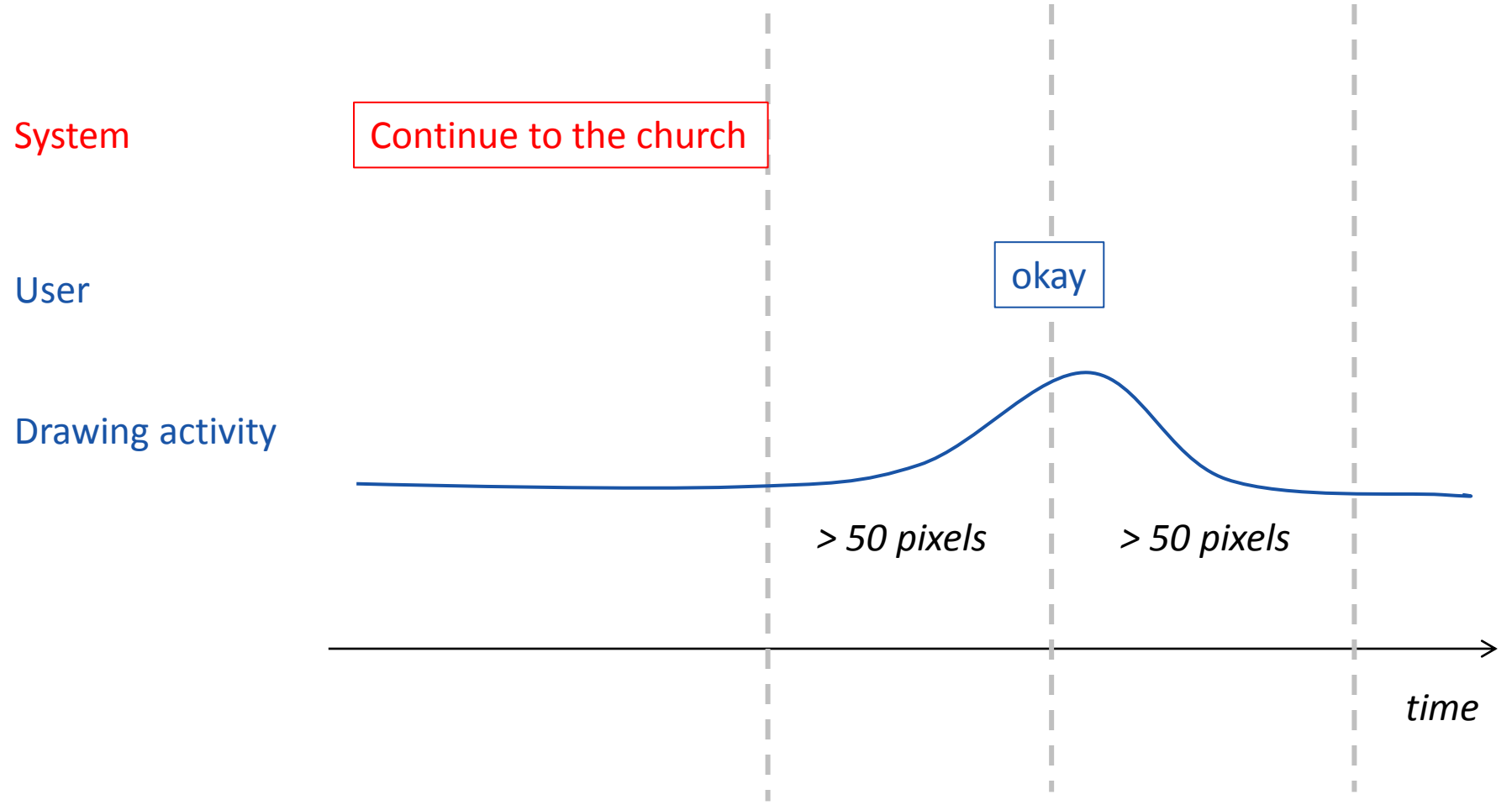
Task completion

- Based on the user's drawing activity
- 4 classes:
  - Before drawing
  - While drawing
  - After drawing
  - No drawing

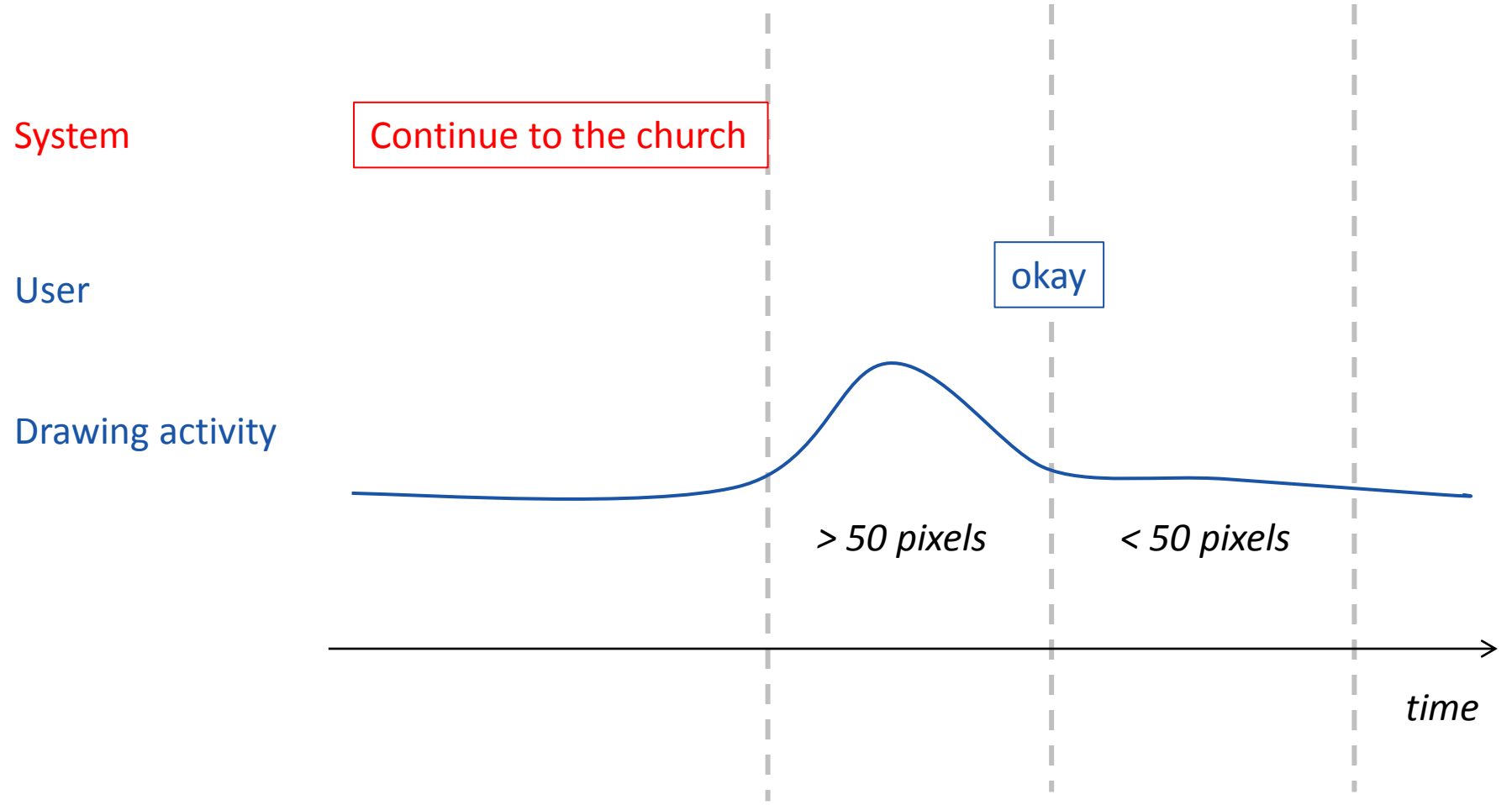
Before drawing



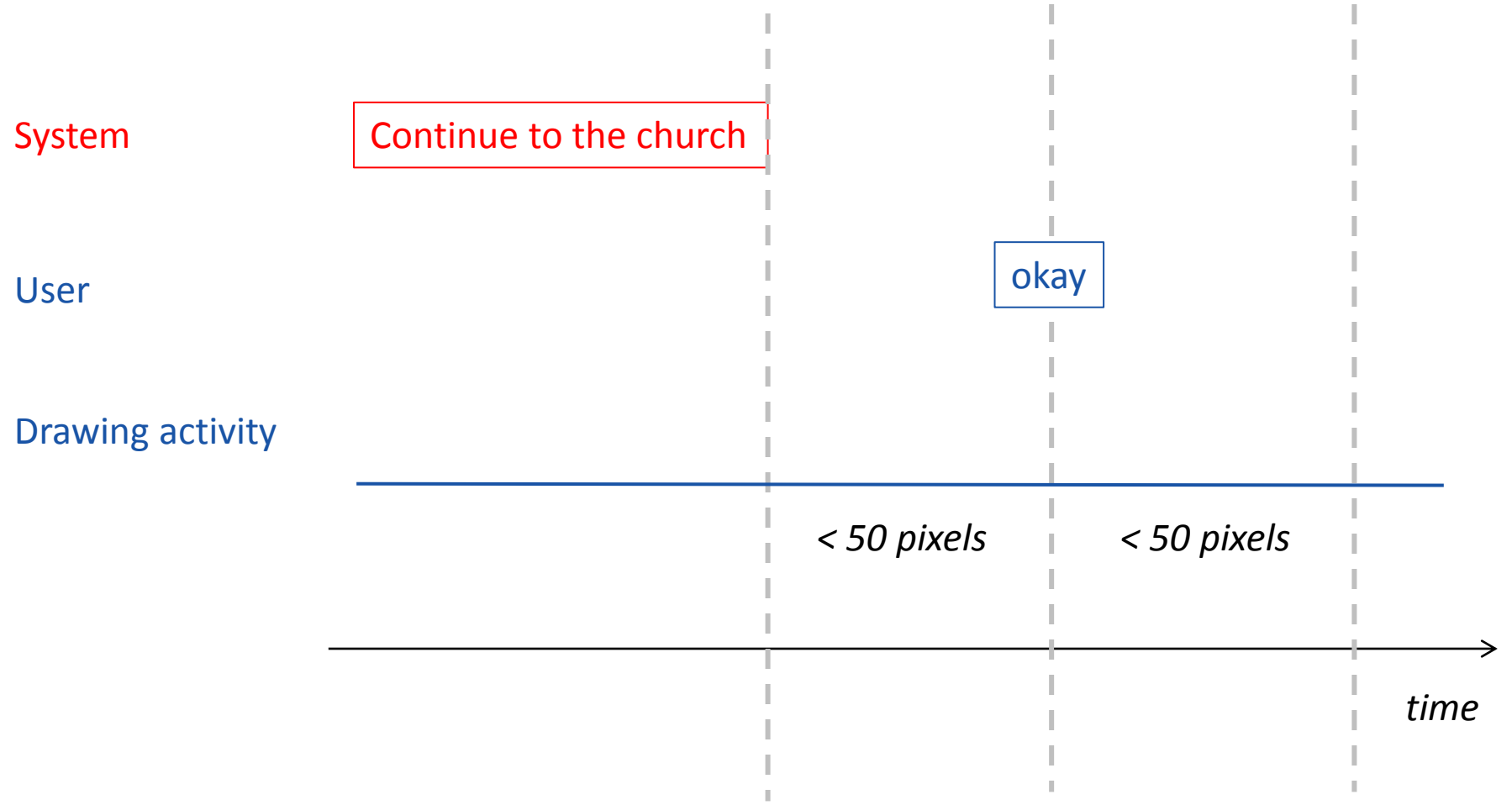
While drawing



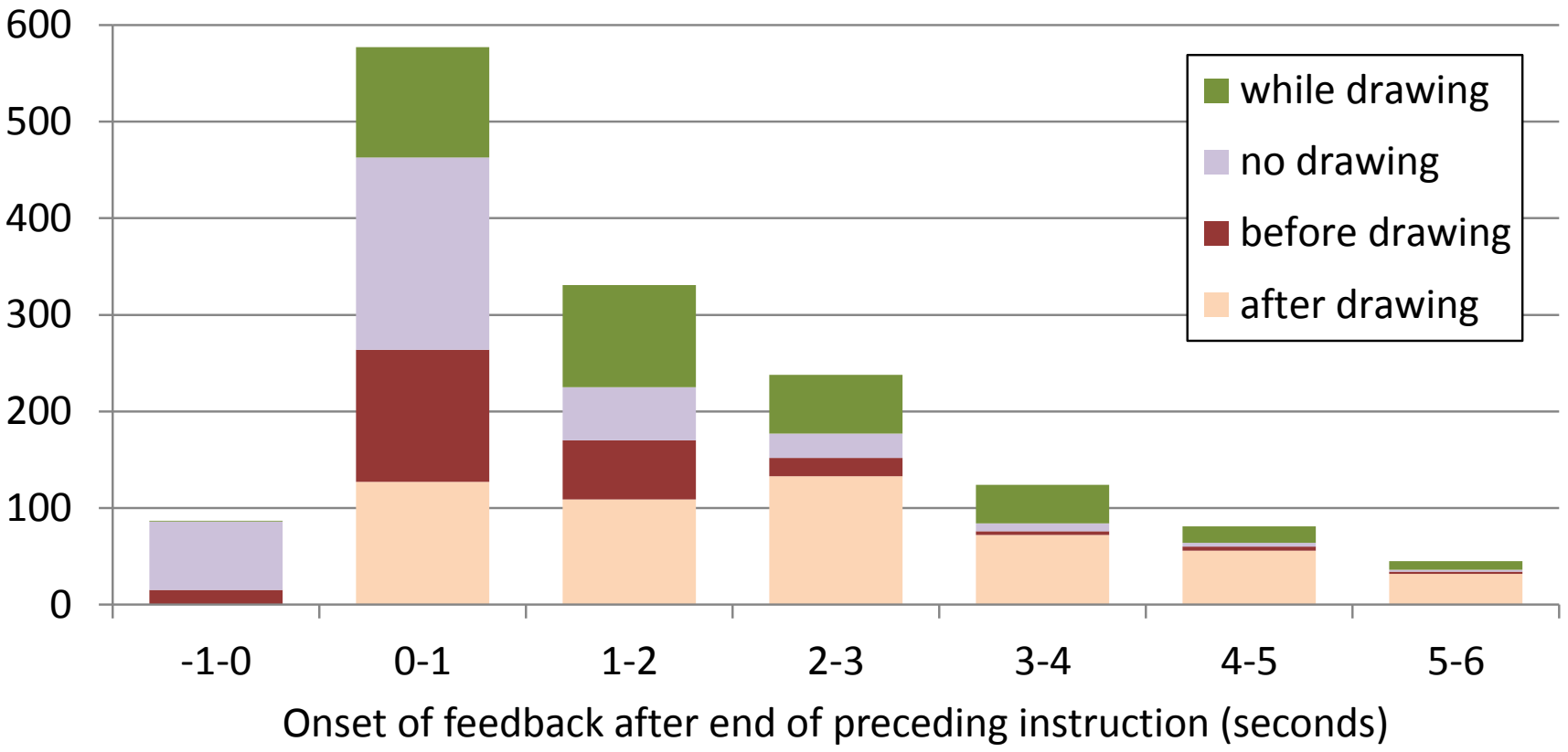
After drawing



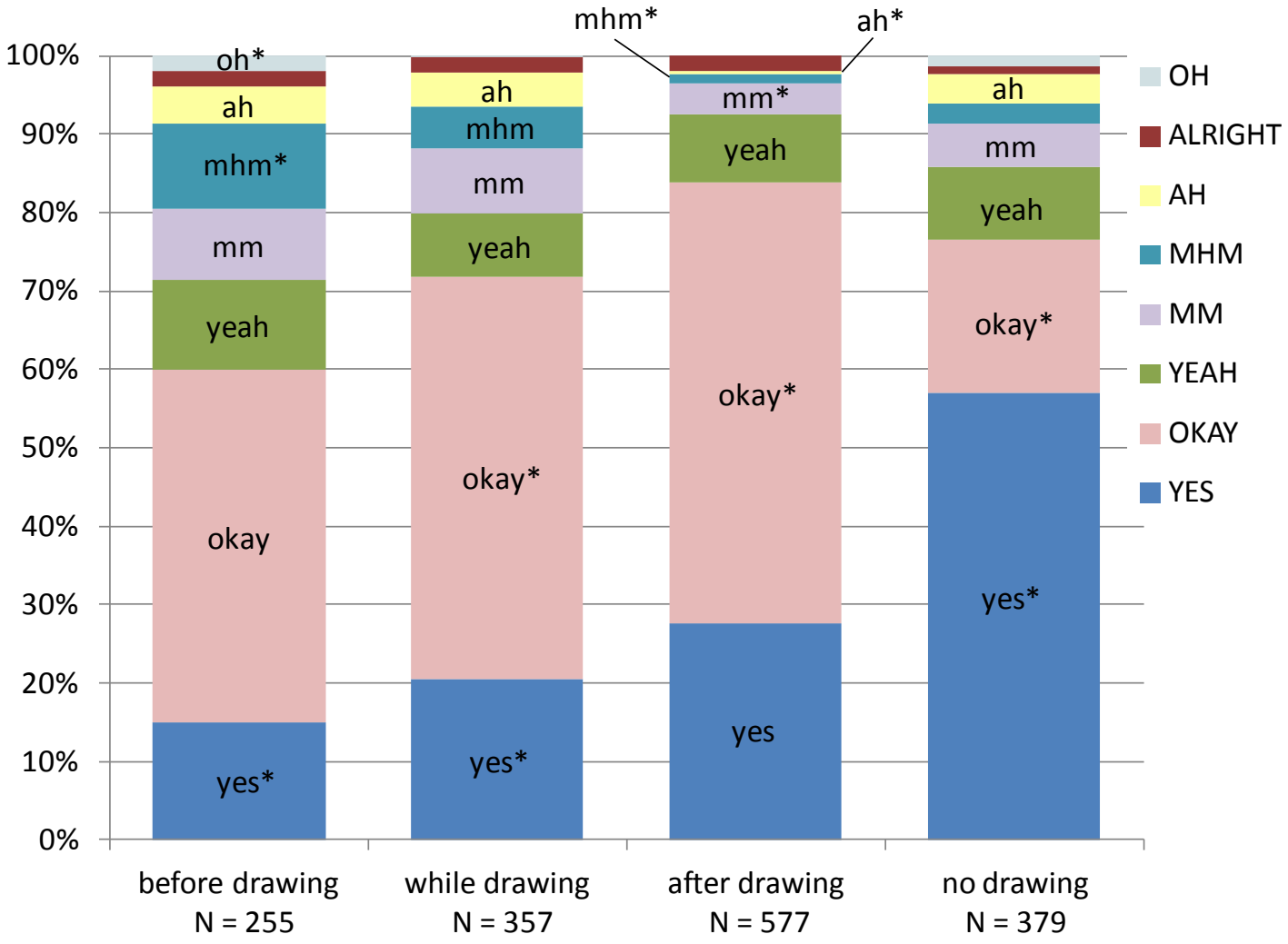
No drawing


System
Continue to the church
User
okay
Drawing activity
< 50 pixels
< 50 pixels
time

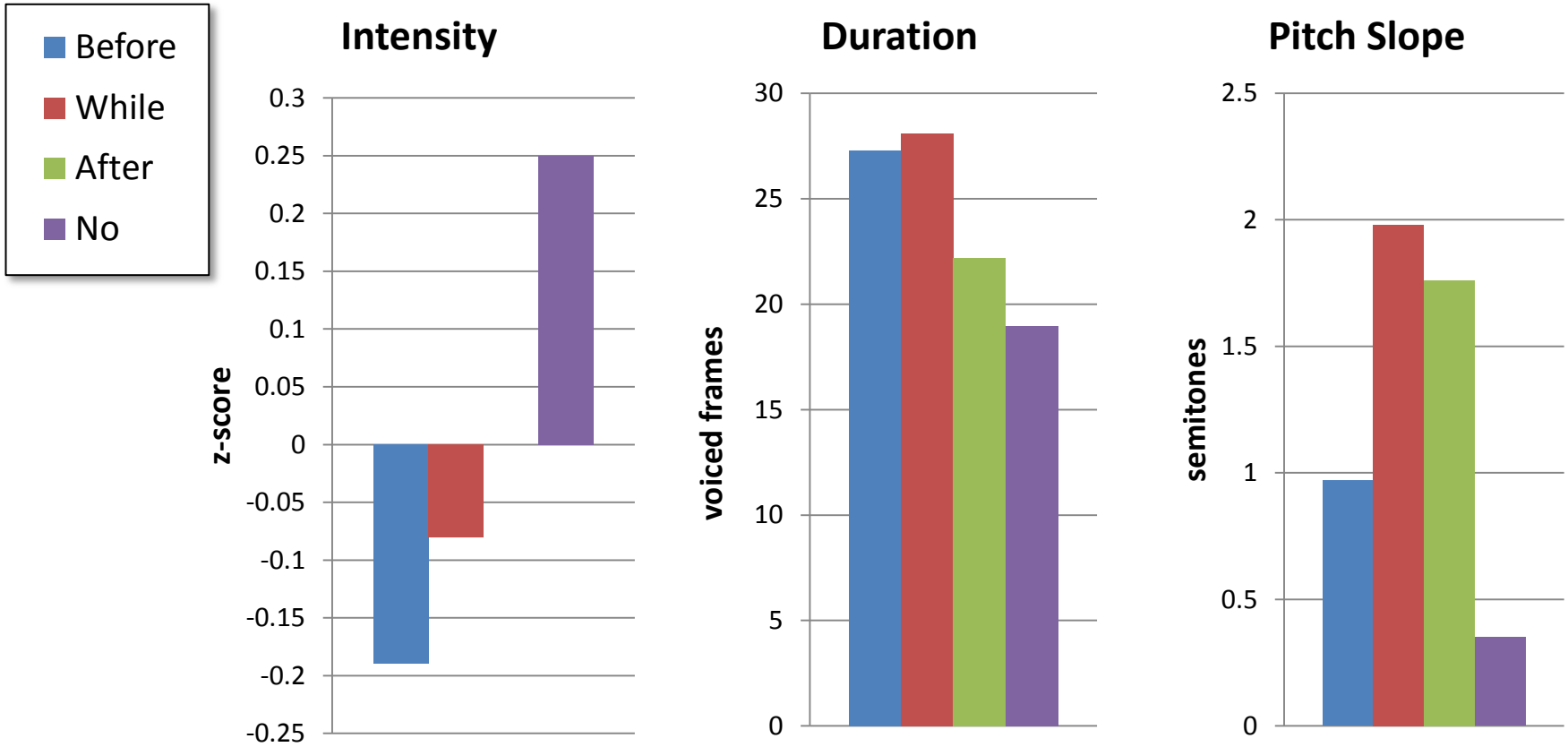
Response time vs. Task completion



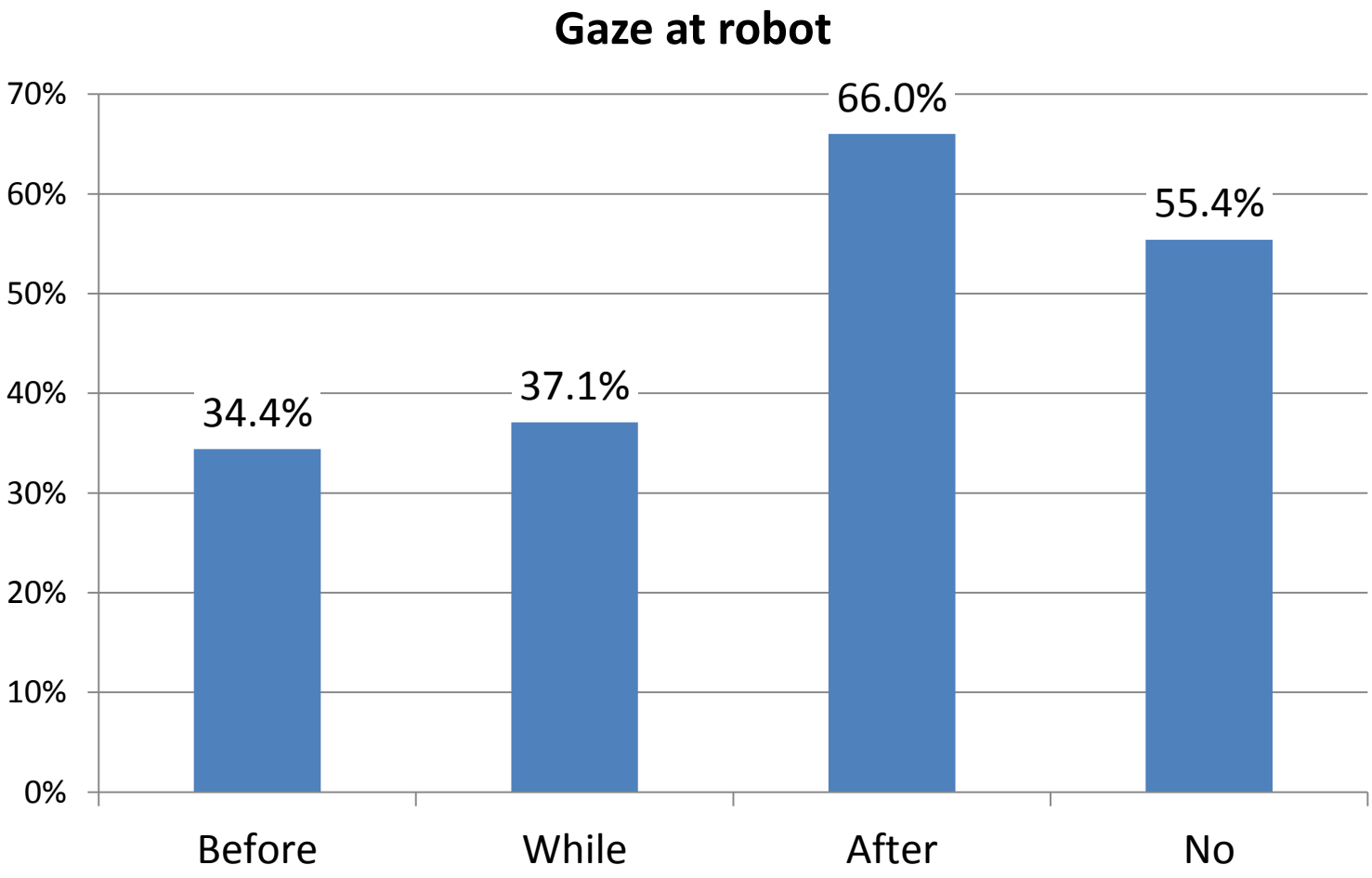
Lexical tokens vs. Task completion



Prosody vs. Task completion



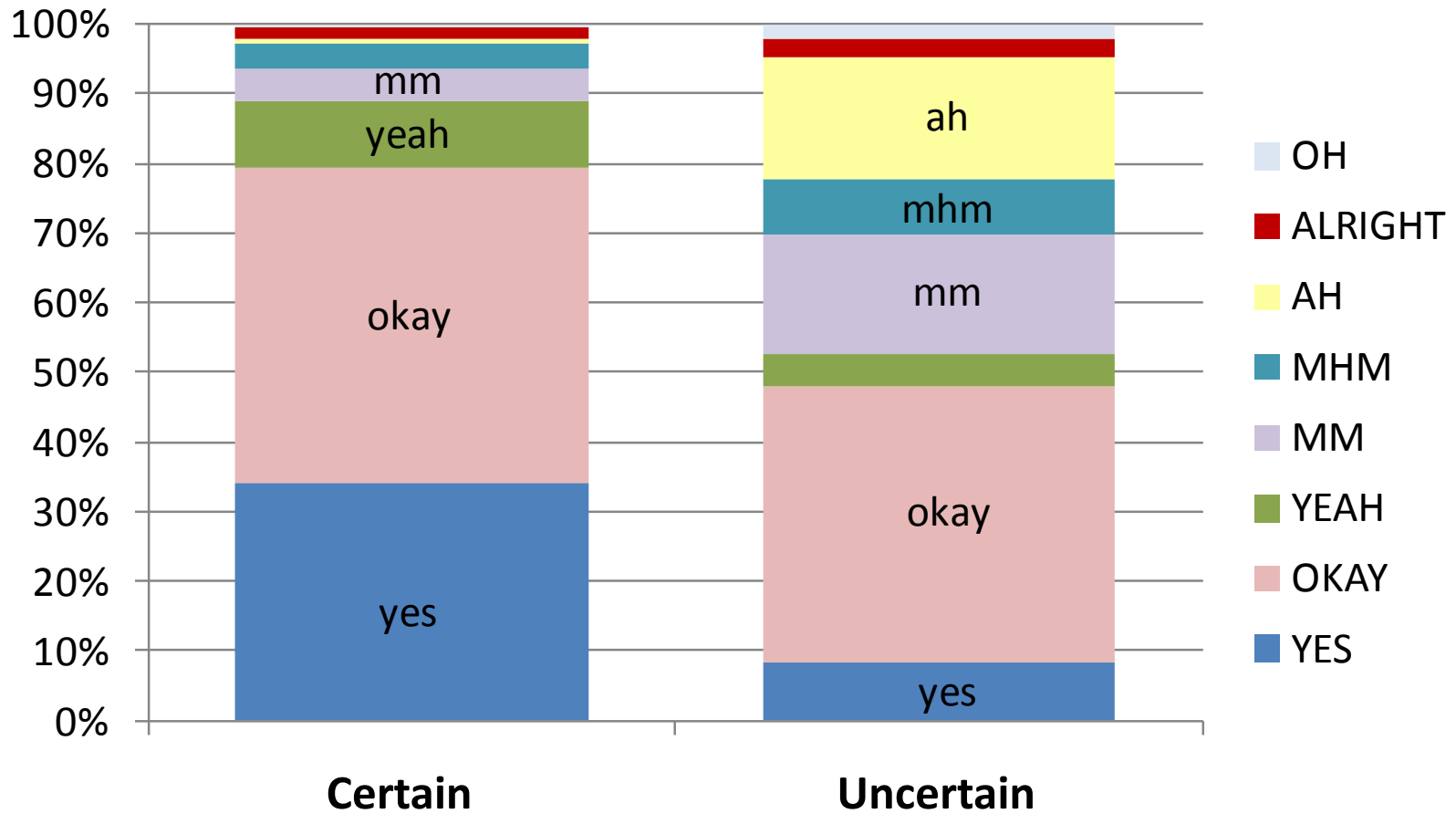
Gaze vs. Task completion



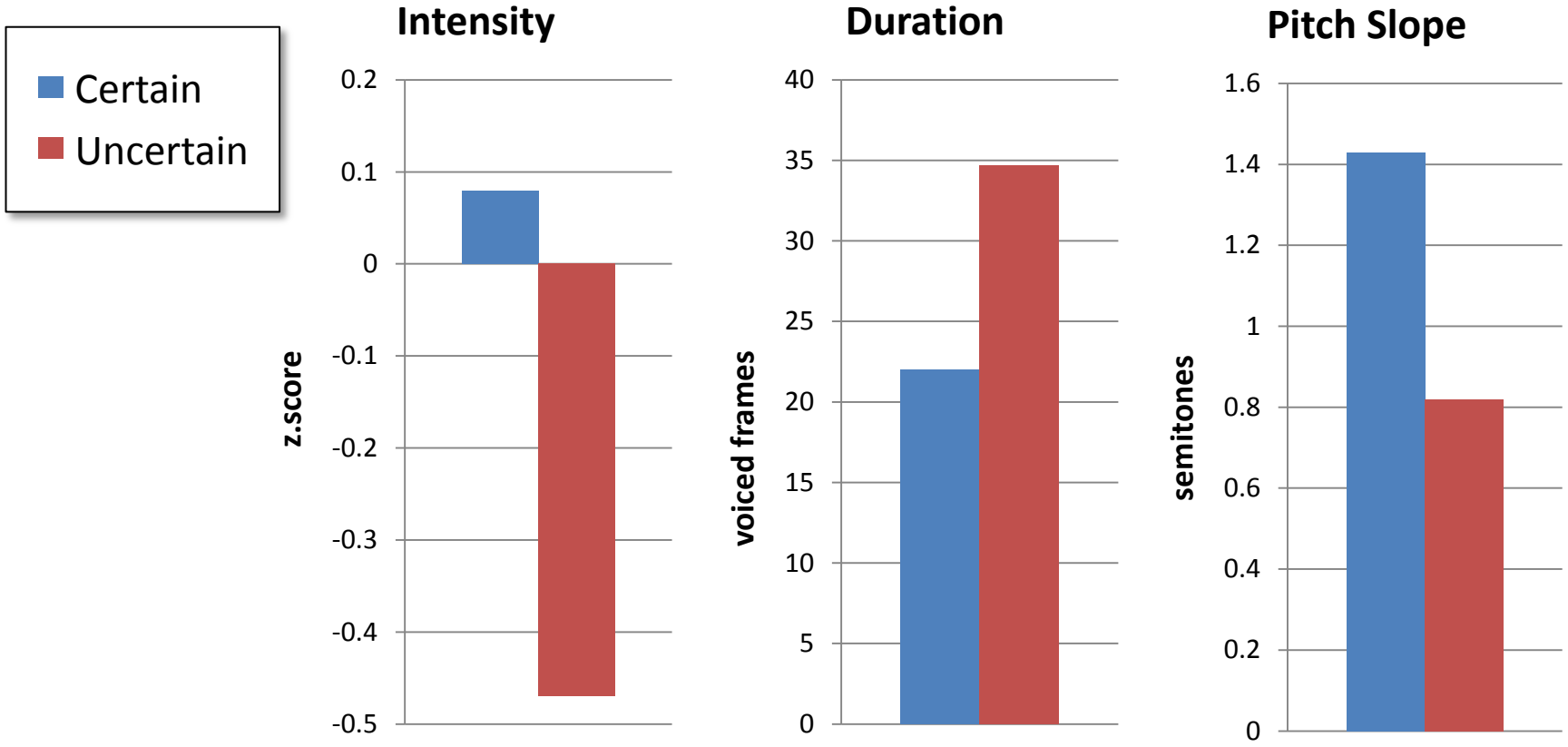
Uncertainty

- Manual labeling:
  - Certain
  - Uncertain (~12%)
- 3 person cross-annotator kappa: 0.63

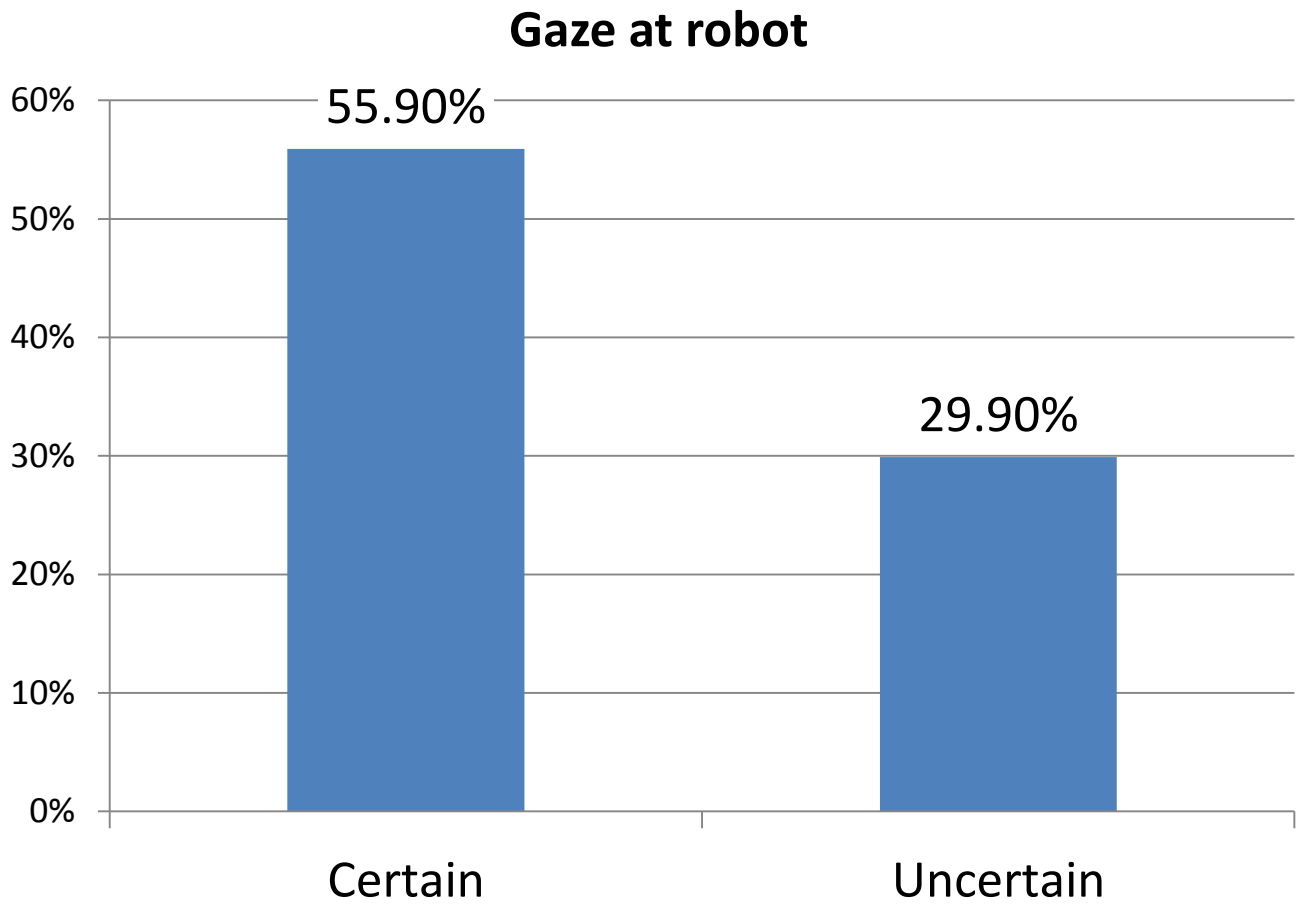
Lexical token vs. Uncertainty



Prosody vs. Uncertainty



Gaze vs. Uncertainty



Conclusions

- When the user give an acknowledgement, task completion and uncertainty can to some extent be inferred from:
  - Lexical choice, Prosody and Gaze
- The first study to investigate
  - How task completion is signalled in acknowledgements
  - How uncertainty is signalled in acknowledgements in a human-machine dialogue setting
- Future work:
  - Build classifiers for these functions
  - Employ the classifiers in the system
  - Let the system incrementally adapt its speech generation according to the user's feedback

ROYAL INSTITUTE
OF TECHNOLOGY

The End

Questions?