

„Continuous Interaction with a Virtual Human“

**eNTERFACE'10 MIDTERM**

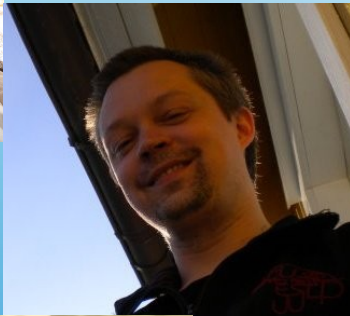
- Dennis Reidsma



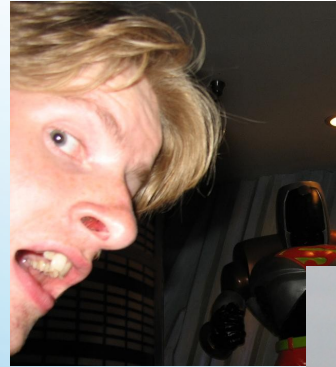
- Bart van Straalen



- Daniel Neiberg



- Iwan de Kok

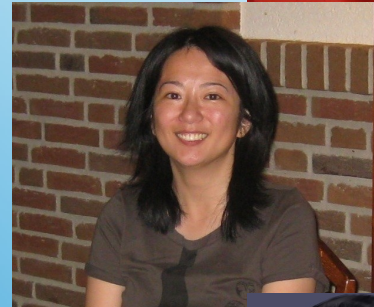


- Herwin van Welbergen

- Sathish Pammi



- Khiet Truong



- Elckerlyc



e10CI TEAM

This eNTERFACE'10 project was kindly sponsored by:



Game research  
for training and  
entertainment



Social Signal Processing Network

**e10CI SPONSORS**

The project: goals, application, and approach

**e10CI PROJECT**

So far, VH systems tend to be developed using a push-to-talk paradigm (half duplex).



**e10CI PROJECT**

## Examples of interaction between humans

- without overlap: this is what VHS can already do to a certain extent
- with overlap: this kind of behavior is the ultimate goal. Who is the Speaker? Who is the Listener here?

Long-term goal: Making a VH capable of Continuous Interaction.

- VH capable of listening while it is speaking
- VH capable of expressing itself while listening
- VH capable of the continuous mutual coordination that humans exhibit in conversation

**e10CI PROJECT**

Detailed goal:

Taking the first step towards the global goal by making a VH capable of actively dealing with Listener Responses from the user, while the VH is speaking.

- 1) Elicit Responses from Listener
- 2) Deal with them, timely and adequately

**e10CI PROJECT**

- Using the open source Virtual Human platform „Elckerlyc“



**e10CI PROJECT**

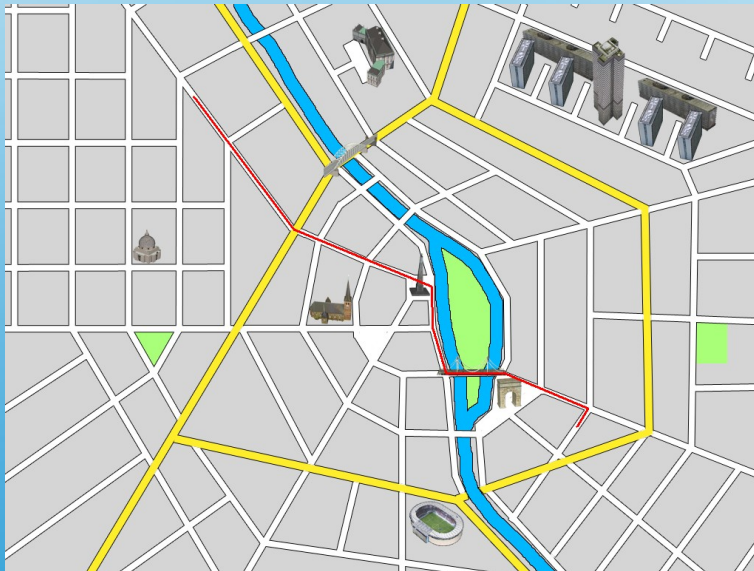
Eliciting Responses from the User

**eLICITING RESPONSES**

It only makes sense to have a Virtual Human that can deal with Listener Responses if the user actually gives such responses to the Virtual Human!

**eLICITING RESPONSES**

- We use a route-description application to elicit Listener Responses



**eLICITING RESPONSES**

At the end of each sub step in the route description, Responses are elicited using:

- vocal features (new „pitch contour“ markup in OpenMary)
- face and head (new repertoire of expressive behavior)
- both of the above
- none of the above

based on corpus and literature

**eLICITING RESPONSES**

## Results

- Analysis to follow...
- ...but at the least, we did get a lot of Responses:
- only vocal, or only non-vocal, or both
- Understanding+, Understanding-, Interruptions, „Repeat instruction“, politeness („thank you!“),

**eLICITING RESPONSES**

## Results

- Video of Virtual Human
- Video of User (10C#420, 102#514, 10A#515, 114#760-790)
- Recorded 10 sessions with eNTERFACE participants

**eLICITING RESPONSES**

- Analysis is ongoing, but....
- ...at least in one video we see Responses for the „combined“ condition, but not for the „no-elicitation“ condition.

**eLICITING RESPONSES**

- Analysis is ongoing, but....
- ...at least in one video we see Responses for the „combined“ condition, but not for the „no-elicitation“ condition.

**+500!**

**eLICITING RESPONSES**

Dealing with Responses, in an adequate and  
timely manner

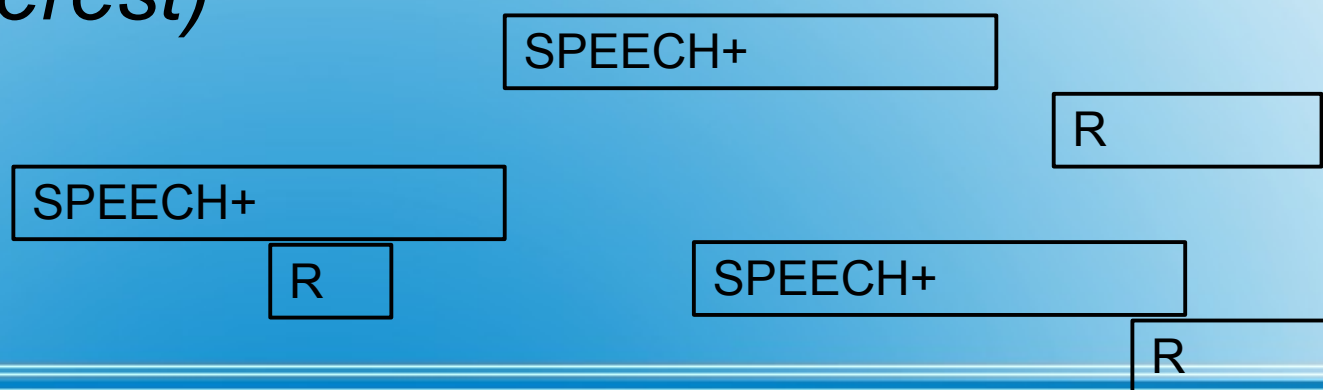
**dEALING WITH RESPONSES**

- What responses do humans give?
- How do human speakers deal with responses?
- Can we classify the responses automatically?
- Can we deal with them automatically?

## dEALING WITH RESPONSES

## Descriptions for Responses

- Cooperative / Competitive
- Understanding / Nonunderstanding
- (*Agreement / Disagreement*)
- (*Attentiveness*)
- (*Interest*)
- ...



# dEALING WITH RESPONSES

## Corpus and annotation

- HCRC Maptask
- Official annotations (Acknowledgement vs the rest)
- Additional eNTERFACE'10 annotations (Cooperative / Competitive for the *overlapping listener utterances*)

## Classification of Listener Response with a series of classifiers

- Each solving a subset of the task
- First focusing on Competitive / Cooperative, as that is very important in the scenarios
- Enough result that we should be able to use it in the second (online) experiment

**dEALING WITH RESPONSES**

Feature(s)	300 ms	500 ms
F0	0.55	0.59
Intensity	0.60	0.62
mfcc with 0th	0.72	0.75
mfcc without 0th	0.74	0.75
duration	0.55	0.71
spectral flux	0.66	0.67
Intensity flux mfcc with 0th	0.73	0.76
Intensity flux mfcc with 0th dur	0.75	0.76
Intensity flux mfcc without 0th	0.74	0.76
Intensity flux mfcc without 0th dur	0.73	0.76

TABLE II

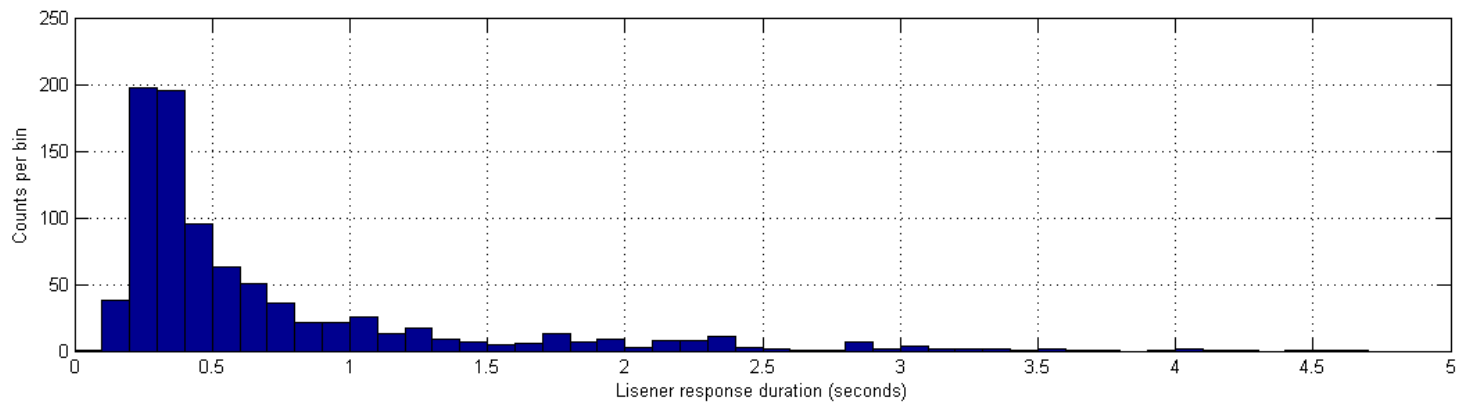
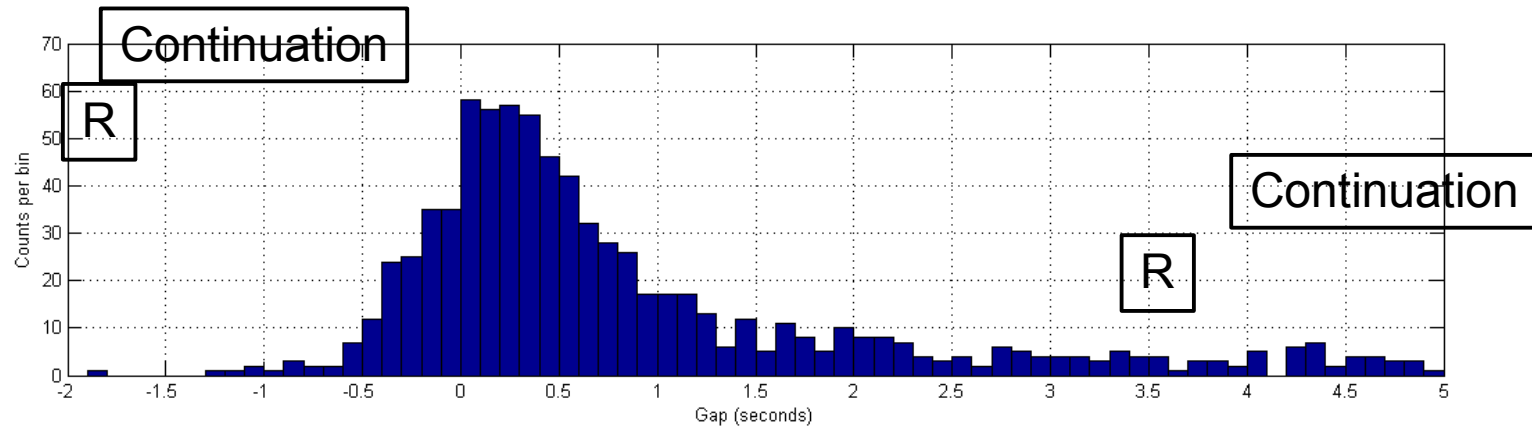
RESULTS FOR LISTENER RESPONSE VS ELSE CLASSIFICATION FOR DEVELOPMENT SET. TALKSPURTS ARE CREATED FROM CORPUS SEGMENTATION

max latency (ms)	Features	Avg. F-score
300	Intensity flux mfcc without 0th	73
500	Intensity flux mfcc without 0th dur	76

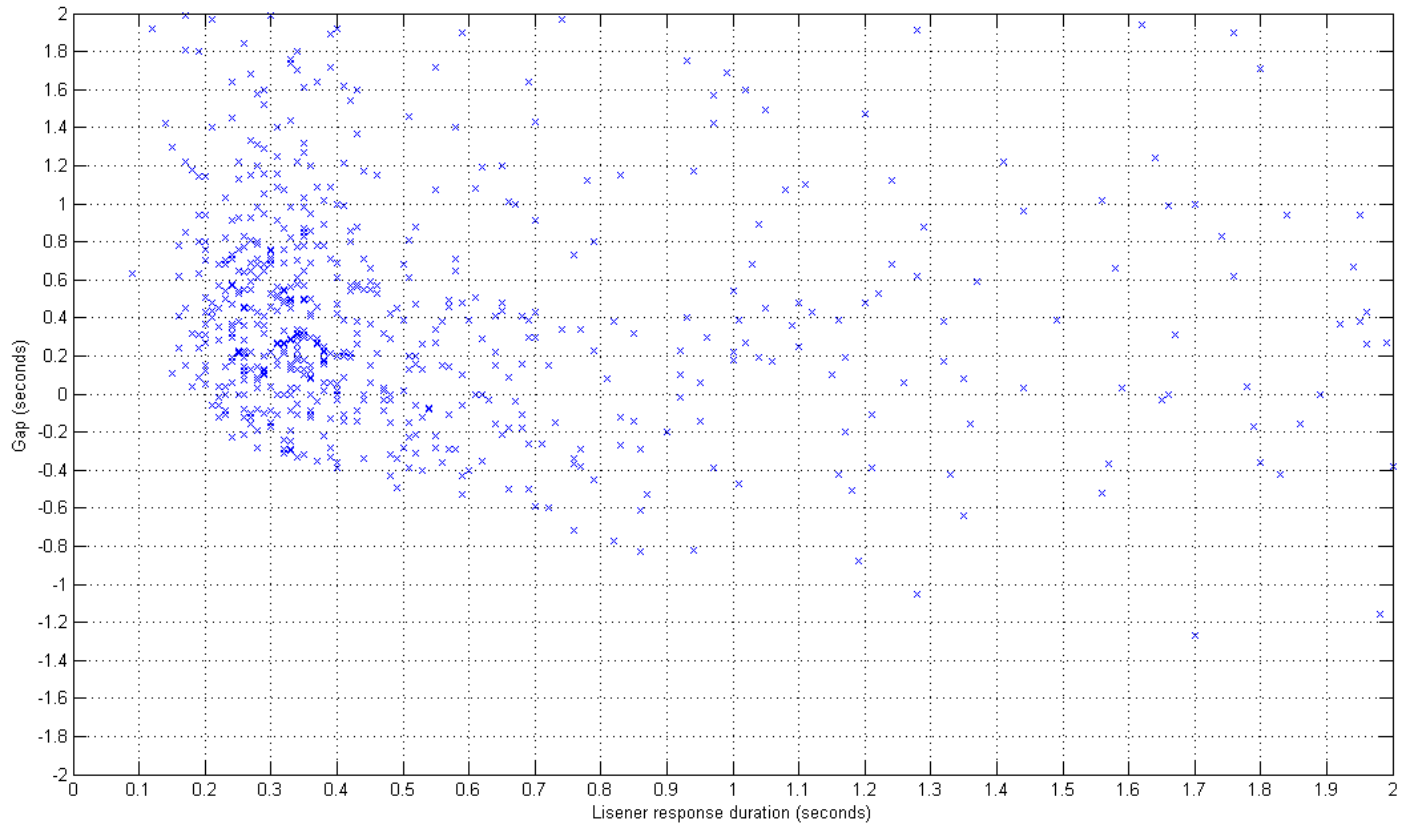
TABLE III

RESULTS FOR LISTENER RESPONSE VS ELSE CLASSIFICATION FOR EVAL-SET. TALKSPURTS ARE CREATED FROM CORPUS SEGMENTATION

- Given that we can detect and classify Responses, how shall we deal with them, in an adequate and timely manner?
- What do humans do?



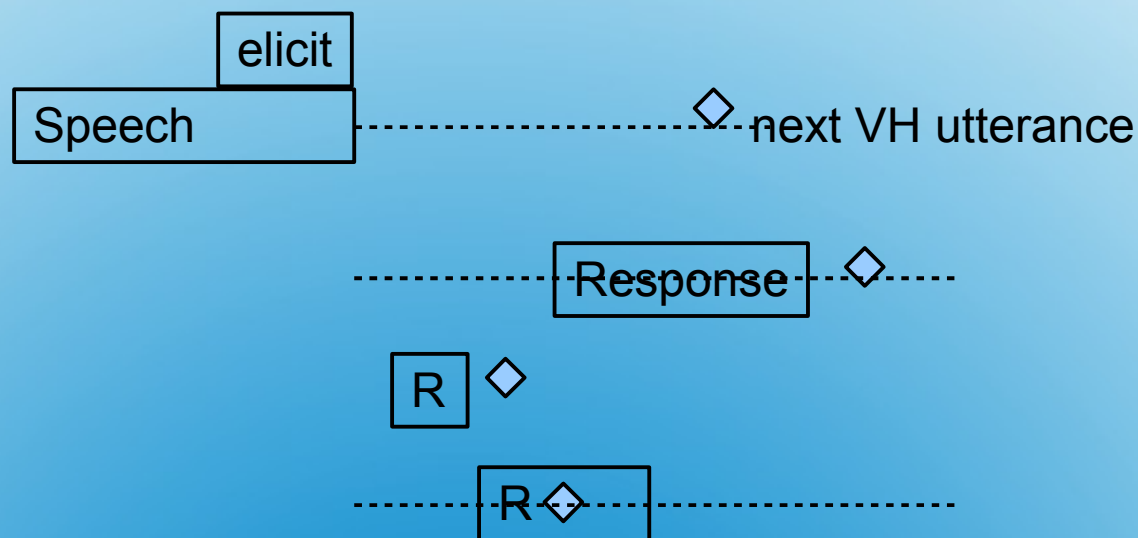
# dEALING WITH RESPONSES



# dEALING WITH RESPONSES

## The second experiment

- Involves several ways of dealing with responses, with and without overlap being generated in the Virtual Human's speech.



# dEALING WITH RESPONSES

- The system capabilities are there, but the dialog management rules still need to be quantified.

**dEALING WITH RESPONSES**

- <setup on paper>

**dEALING WITH RESPONSES**

SSPNet, GATE, Project team, Experiment  
Participants, other eNTERFACERs, Senior Project  
Advisors, Mark&Ronald, Albert

Thank you!

**tHE END**