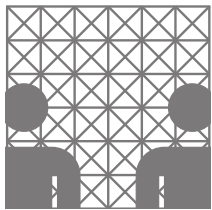
Coordinating Speech Delivery to Gesture Progress

Timo Baumann
Universität Hamburg

baumann@informatik.uni-hamburg.de
www.timobaumann.de/work

Coordinating Speech Delivery to Gesture Progress

e.g. for deictic expressions:

„*move **this** piece over **there** through **that** gate.*“

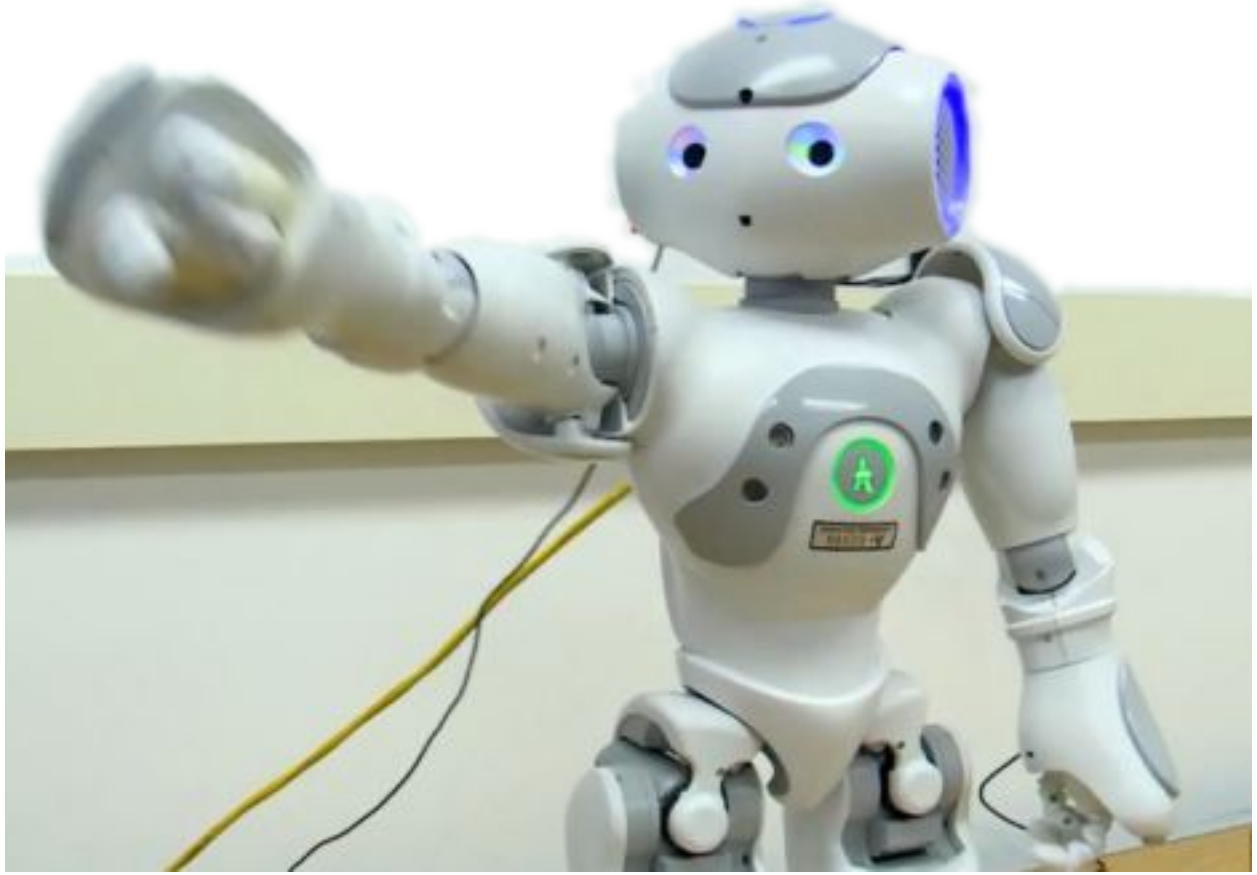


Image courtesy of Vanderbilt University.

Coordinating Speech Delivery to Gesture Progress

e.g. for deictic expressions:

„*move **this** piece over **there** through **that** gate.*“

Image courtesy of Vanderbilt University.

Coordinating Speech Delivery to Gesture Progress

e.g. for deictic expressions:

*„move **this** piece over **there** through **that** gate.“*

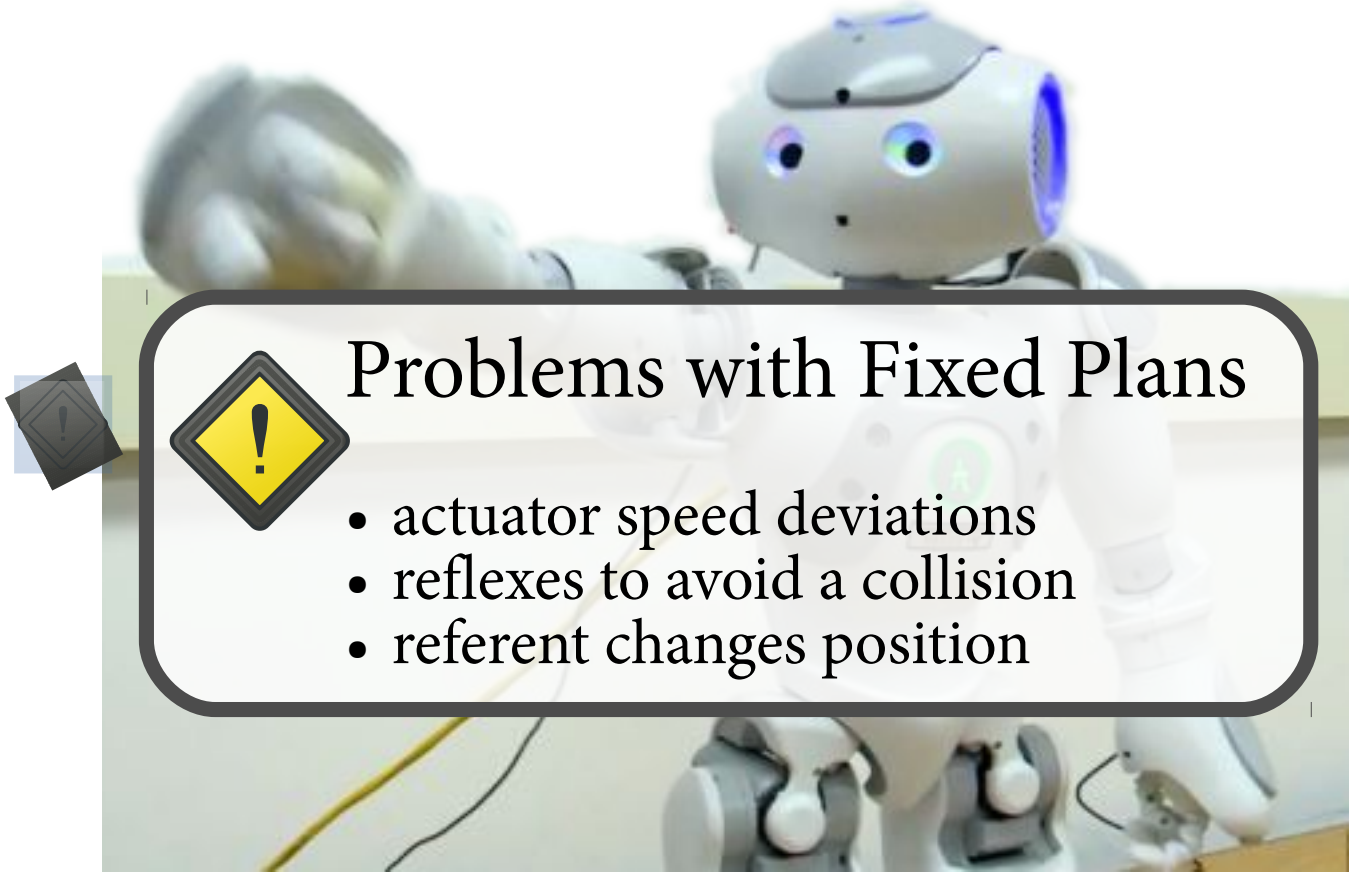


Image courtesy of Vanderbilt University.

Conventional Approach: Stop/Resume Speech

„move ..... this piece ..... over ... there through th..at gate.“

- just pausing the audio stream is psycholinguistically implausible
  - effects around pauses would still have to be modelled
  - it's not what humans do
- already requires delivery progress information
- no way to speed up synthesis if gesture is ahead of time

→ simply stopping/resuming doesn't cut it

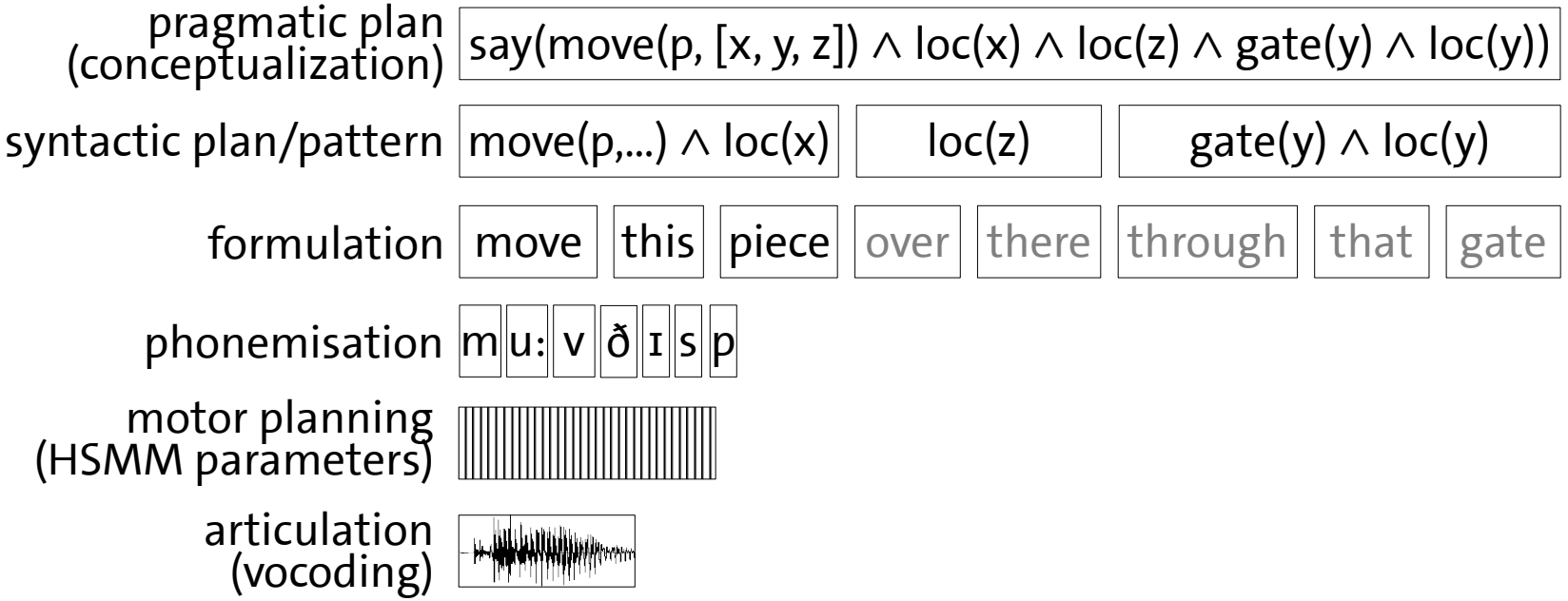
Coordinating Speech Delivery to Gesture Progress

... with **incremental speech synthesis**

- more flexible than stopping/resuming speech:
  - online speech tempo adaptations (stretch/compress)
  - change content that is to be spoken (e.g. change a referent)
  - *reflexive* behaviour, such as hesitations
  - provide detailed feedback on delivery progress

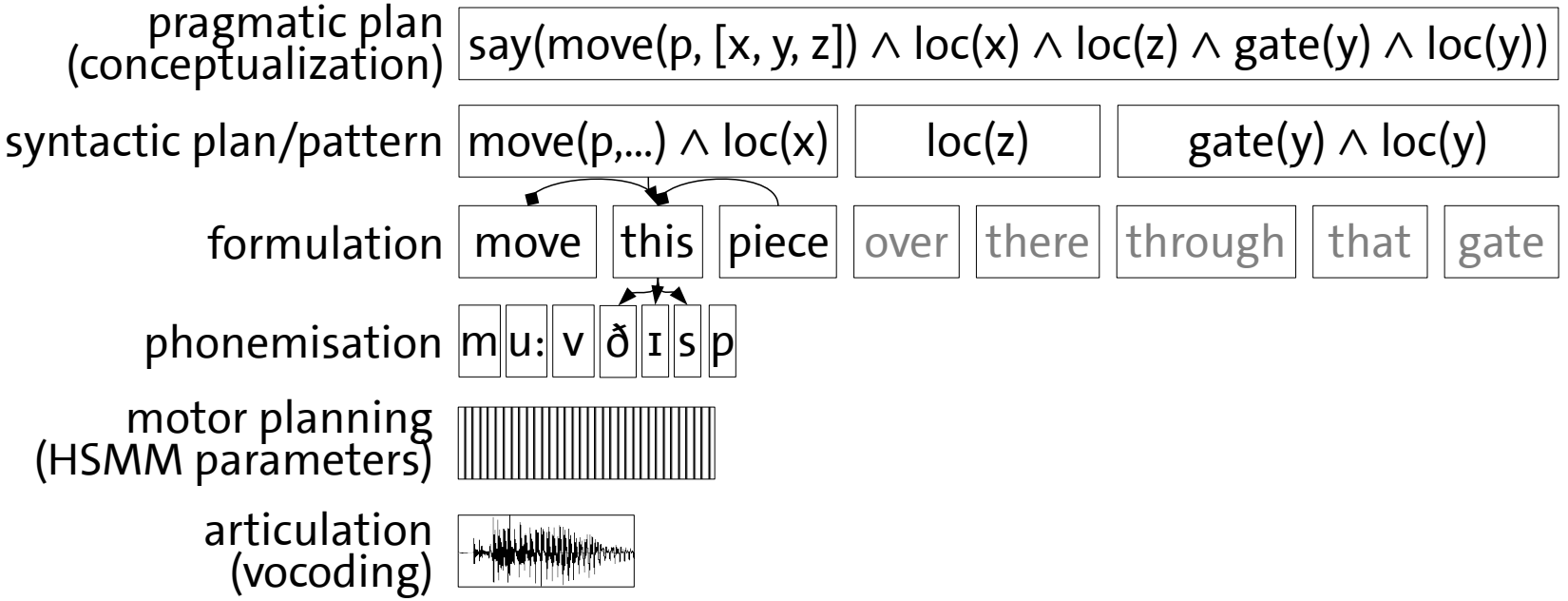
→ relatively easy to perform in the IU Framework

Incremental Processing in the IU Framework



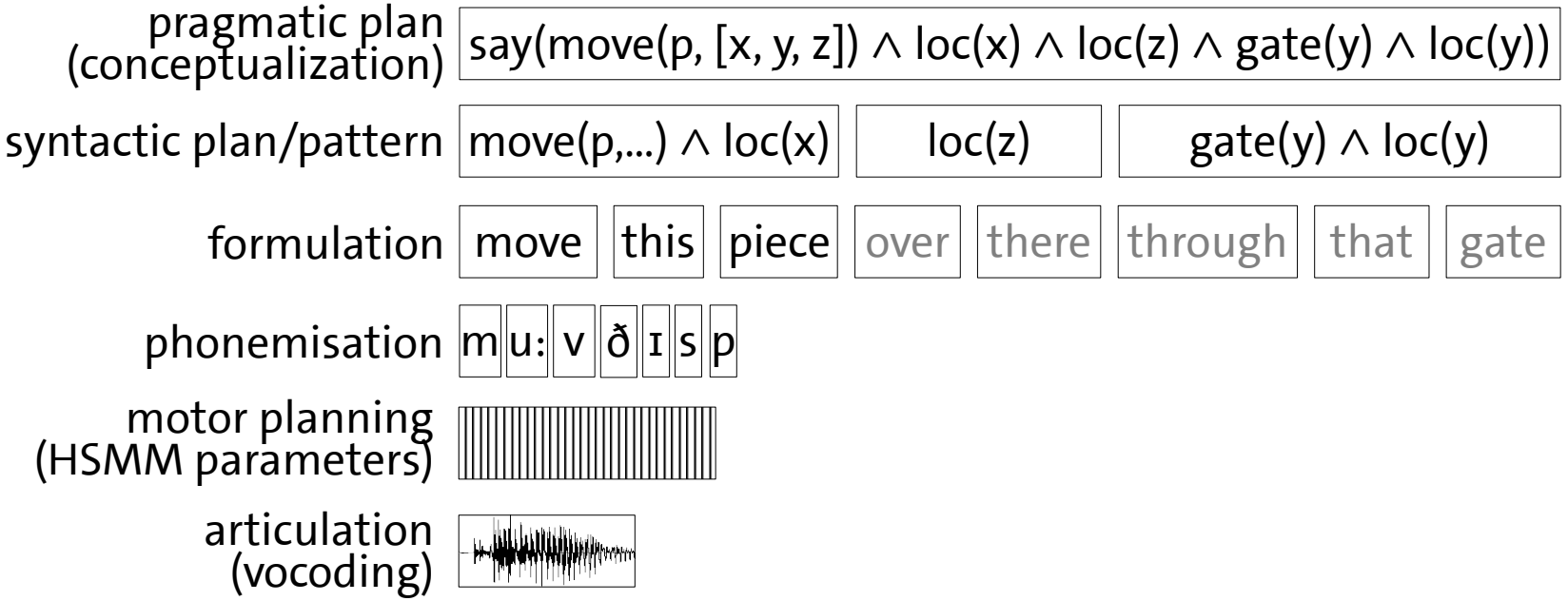
- data come as *increments* (IUs), smallest units of information on a given level of abstraction that are individually processable

Incremental Processing in the IU Framework



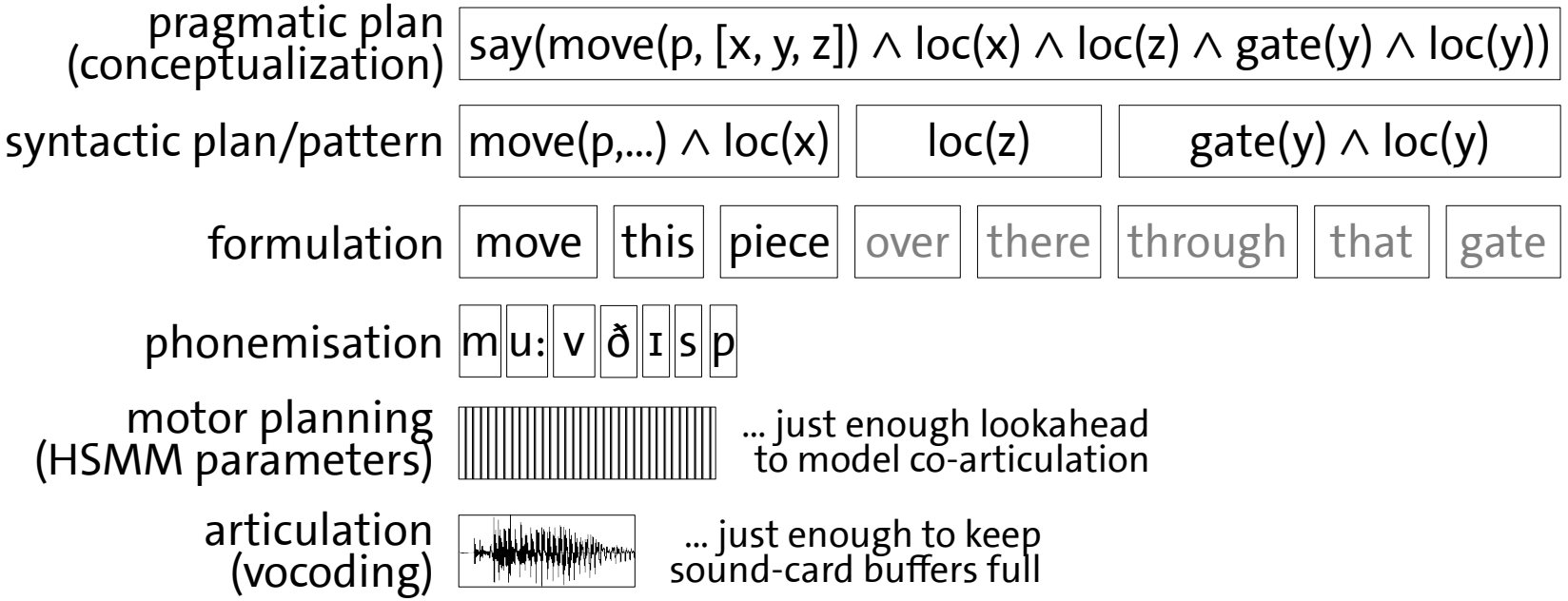
- IUs are interconnected with related IUs (those that are above/below, or on the same level)
  - IUs form a network that reflects the system state

Incremental Processing in the IU Framework



- the system state is changed by adding/removing IUs
- IUs can be managed by processors that react to network updates (add/revoke/update)

Incremental Processing in the IU Framework



- the system state is changed by adding/removing IUs
- IUs can be managed by processors that react to network updates (add/revoke/update)

Coordinating Speech Delivery to Gesture Progress

→ relatively easy to perform in the IU Framework

→ **requirements for coordination/coupling:**

- provide synchronization points between motion & speech
- synchronization should be available on *various linguistic levels* (phrases, words, syllables) and
- synchronization should *integrate with production capabilities* (NLG, prosody, articulation, synthesis)

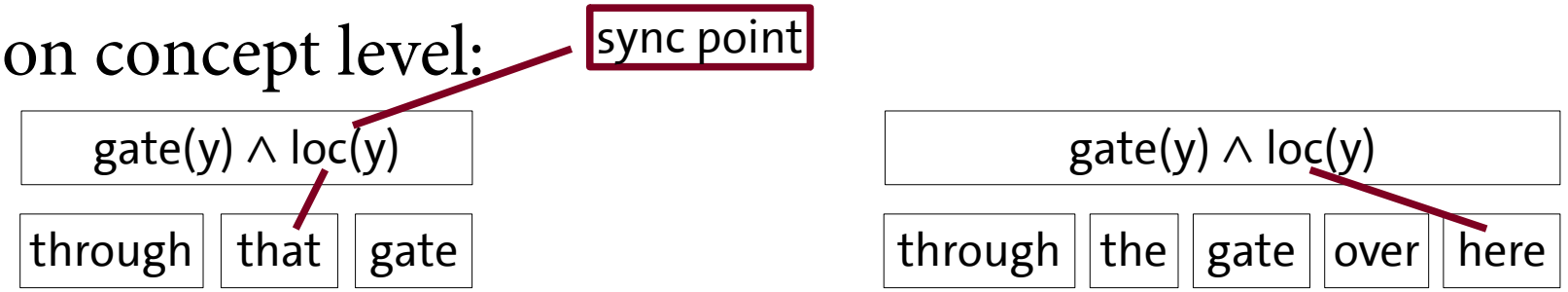
Interfacing with Execution Control

- anchor points between gesture&speech that are co-planned in advance
  - e.g. start/center/end of some IU
- robot control monitors gesture delivery and notifies of deviations (e.g. including $t_{expected}$ and expectation error)
  - notify the IU, it will automatically determine the relevant processing steps
  - expectation error could be used e.g. to determine whether stretching or hesitating should be performed
- speech delivery returns how well it is able to meet the new goal

An example

- synchronization on more abstract level leaves more freedom (and responsibility) to speech delivery:
  - on word level:

  - on concept level:



- let speech delivery decide on the best option given the timing constraints

Summary

Coordinated Speech Delivery is work-in-progress

- tempo changes not yet articulatorilly plausible
  - need to determine stretchability in given contexts
- hesitations are available (but do not sound great)
- simple re-generation would be easy, but inflexible
  - thorough re-generation is still an open question
- generic interface with gesture is yet to be determined
  - that's why I came to the workshop
- I don't have a robot

Thank you.

baumann@informatik.uni-hamburg.de,
get the code at inprotk.sf.net.

Funded by a Daimler and Benz Foundation PostDoc grant.

Raum für Notizen