

Turn-Taking

Seminar „Sprachverarbeitung“
Presentation by Benedikt Adelman

Turn-Taking from the Perspective of a Dialogue System

Human



System



Realistic?

irritating!

intimidating!

USERS DON'T LIKE

being interrupted
being talked their ears off

THEY DO LIKE

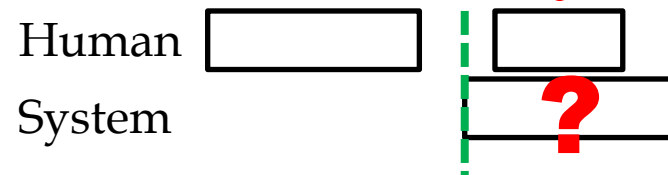
taking their time talking
interrupting the system

“the best units for dialogue systems are the very same ones that humans use”

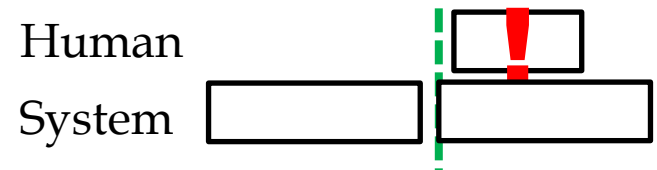
(Edlund, Heldner & Gustafson, 2005)

The System Wants to ...

- ... know whether the user has **(not yet)** finished speaking

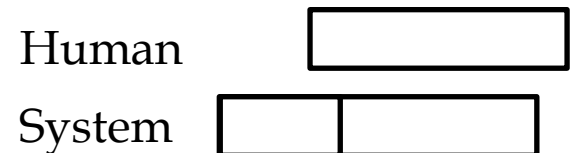


- ... imply it has **(not yet)** finished speaking itself



We might also want to ...

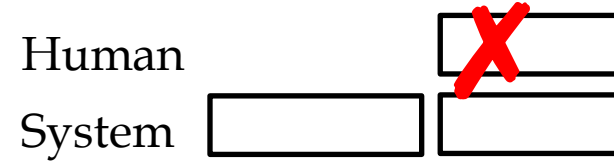
- ... allow the user to interrupt us
- ... be sure the user is listening ⇒ 'backchannels'



What shall we do?

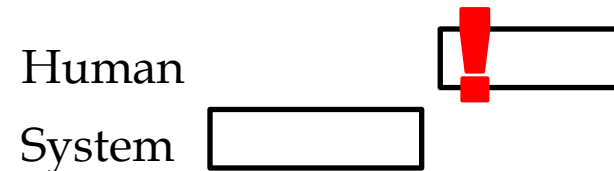
- **system needs to pause without ending the turn**

Include as few turn-yielding cues as possible.
e.g. leave utterances incomplete



- **system has finished its turn**

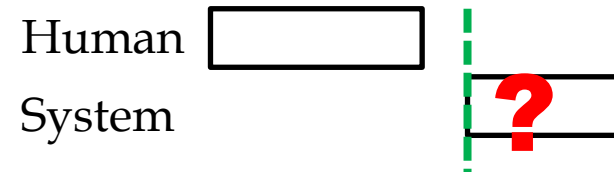
Include as many turn-yielding cues as possible.
e.g. falling intonation



- **user pauses – end of turn?**

Check for turn-yielding cues by the user.

the more present, the more likely end of turn



- **system talking – user listening?**

Include backchannel-inviting cues.

e.g. certain phrases, like 'right?', ...



Problem?

Turn-yielding cues seem to depend on language and culture.

English (e.g.)

- acoustics
- prosody
- intonation
- lengthening words

German (e.g.)

- usually no clear markers
- syntactic completion

Japanese (e.g.)

- 'understanding state' (Satō et al.)
- categories of words
- modal particles

Turn-yielding cues seem to depend on other factors.

English: other study

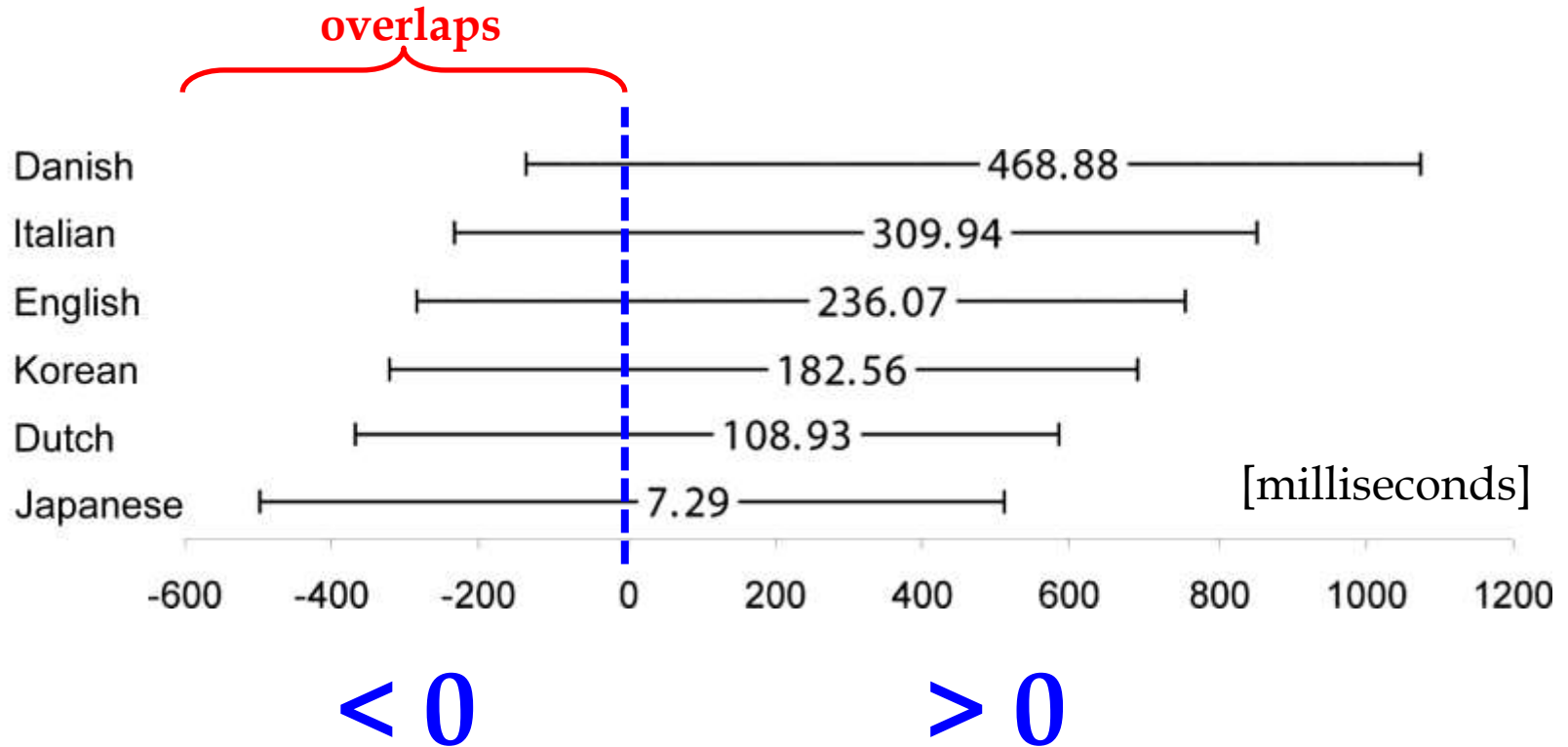
- information flow
- 'minimum necessary information'

Japanese: other study

- prosody
- duration of content words

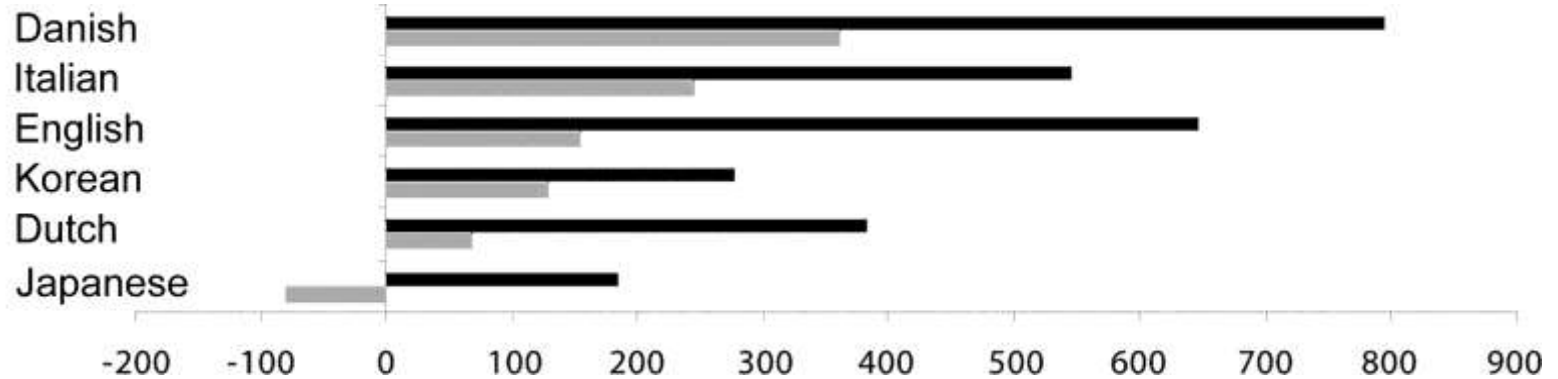
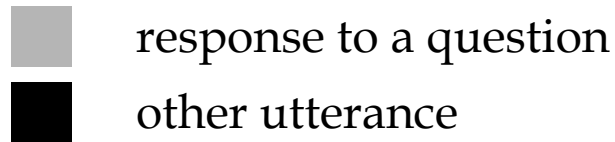
So there might be commonalities. But ...

Example: Pause Durations

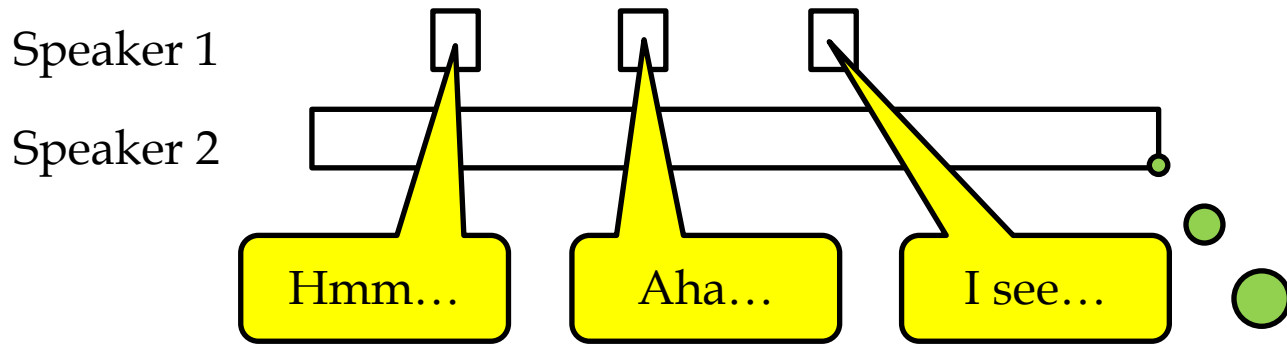


Example: Pause Durations

pause before



Backchannels



'Backchannel-inviting cues'

- at least 110ms duration
- pitch is low
- syntax & prosody important

results for English



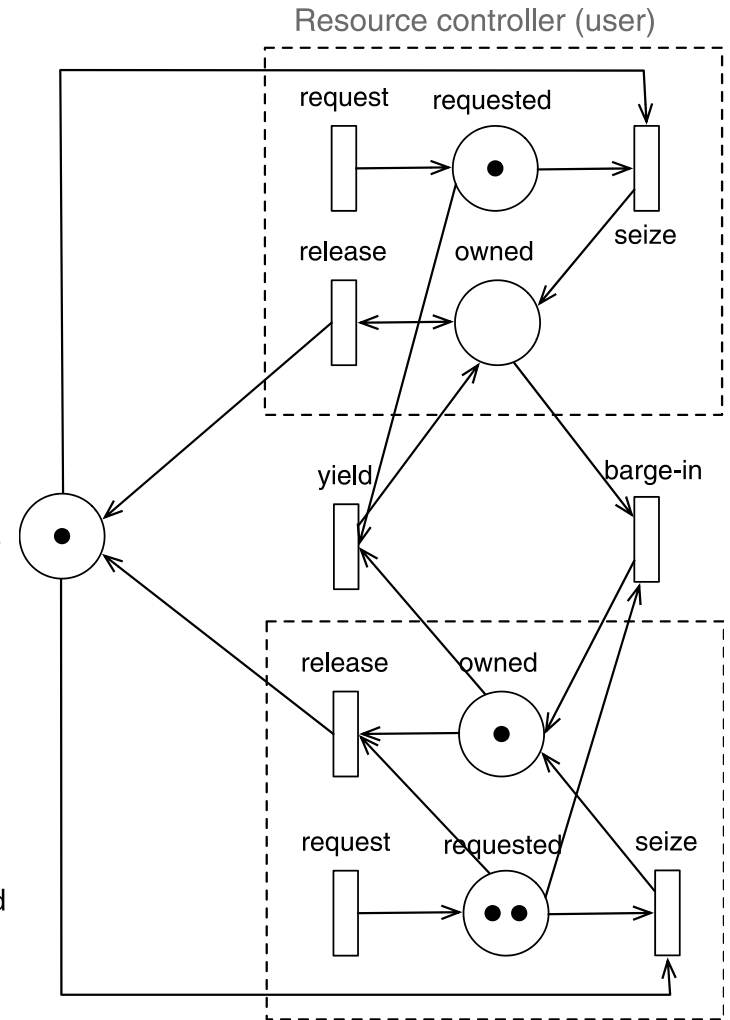
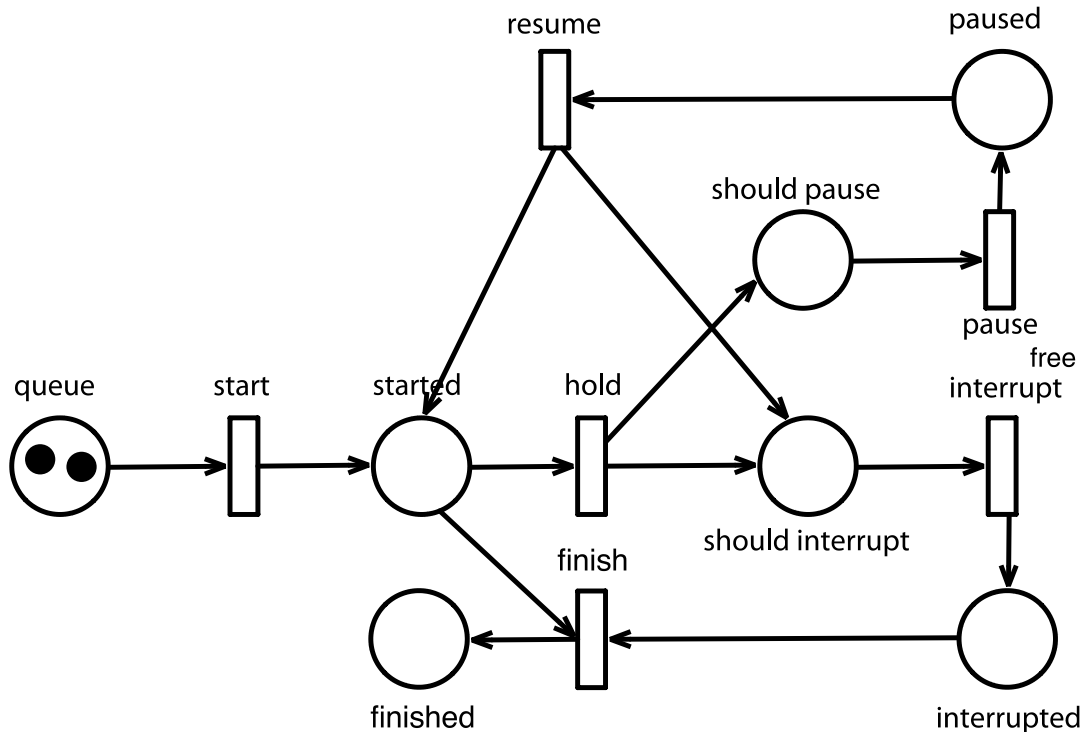
probably culture-dependent as well



Don't stop talking then!

Turn-Taking Implementations

(Timed) Petri Nets.



Warning: This differs from the Petri nets formalism taught in Hamburg!

Turn-Taking Implementations

Decision Trees.

The user pauses ...

here:

mainly **prosodic** features
& content-word duration

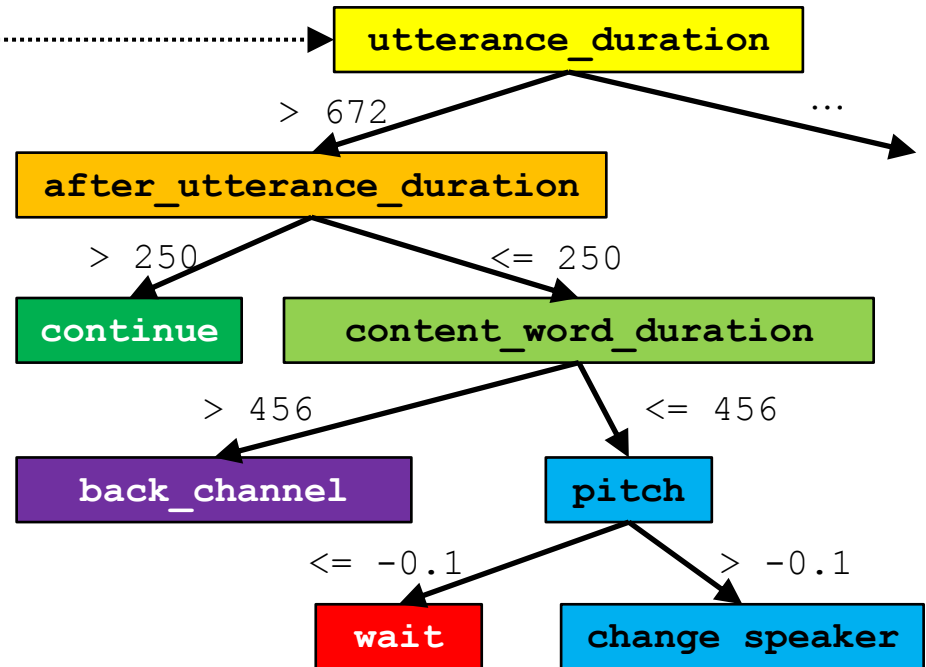
hardly used:

linguistic information

another paper that used decision trees:

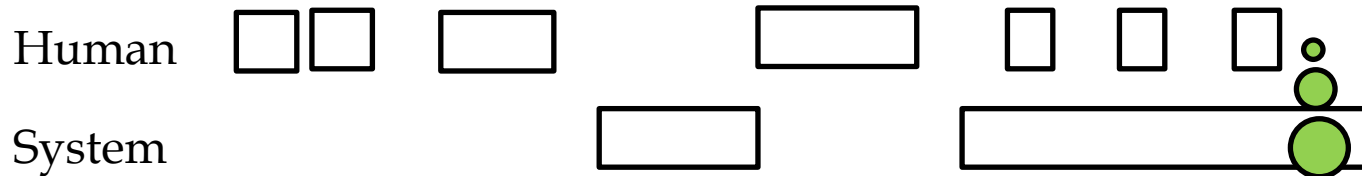
'understanding state'

syntactic and **semantic** categories of words



prosody was unimportant...

Implementation Evaluation



Problem: Speech recognition still too bad.

'The system doesn't understand me!'
= 'The system is bad.'

Sufficiently simple tasks required ...
... so only simple turn-taking can be tested.



Possible Solution (but different experimental set-up):

Use non-speech.

(Paper by Timo Baumann, 2008)

„Was soll dieser Mist?“ → „ba ba baba ba?“

Also suited for data acquisition.

This slide was not part of the presentation.
It just contains the references it is based on.

- Stivers et al., 2009: *Universals and cultural variation in turn-taking in conversation*
<http://www.pnas.org/content/106/26/10587.full>
- Schegloff, 2000: *Overlapping talk and the organization of turn-taking for conversation*, *Language in Society*, Vol. 29
<http://sites.google.com/site/llewellynnick/SchegloffOverlap.pdf>
- Auer, 1991: *Vom Ende deutscher Sätze*, *Zeitschrift für germanistische Linguistik*, Band 19
<http://paul.igl.uni-freiburg.de/auer/userfiles/downloads/Vom%20Ende%20deutscher%20Saetze.pdf>
- Reinelt, 2008: *Interaktionsorganisation: Sprecherwechsel im Deutschen*
https://www.jstage.jst.go.jp/article/dokubun1947/88/0/88_0_101/_pdf
- Sato et al., 2002: *Learning decision trees to determine turn-taking by spoken dialogue systems*, ICSLP2002
https://www.researchgate.net/profile/Mikio_Nakano/publication/221480953_Learning_decision_tree_to_determine_turn-taking_by_spoken_dialogue_systems/links/0912f50f1233c403c7000000.pdf
- Gravano & Hirschberg, 2011: *Turn-taking cues in task-oriented dialogue*, *Computer Speech & Language* 25(3)
https://www.researchgate.net/profile/Agustin_Gravano/publication/220629598_Turn-taking_cues_in_task-oriented_dialogue/links/02e7e52b25d3ba7471000000.pdf
- Edlund et al., 2005: *Utterance segmentation and turn-taking in spoken dialogue systems*, in: Fisseni, Schmitz, Schröder & Wagner (eds.), *Computer Studies in Language and Speech*
https://www.researchgate.net/profile/Jens_Edlund/publication/228094864_Utterance_segmentation_and_turn-taking_in_spoken_dialogue_systems/links/0fcfd506ad6c798b9b000000.pdf
- Takeuchi et al., 2004: *Timing Detection for Realtime Dialog Systems Using Prosodic and Linguistic Information*
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.113.2474&rep=rep1&type=pdf>
- Baumann, 2008: *Simulating spoken dialogue with a focus on realistic turn-taking*
http://edoc.sub.uni-hamburg.de/informatik/volltexte/2013/185/pdf/timo_baumann_bababa.pdf
- Chao & Thomaz, 2016: *Timed Petri nets for fluent turn-taking over multimodal interaction resources in human-robot collaboration*, *The International Journal of Robotics Research* 1–24
<http://ijr.sagepub.com/content/early/2016/03/09/0278364915627291.abstract>
full text accessible via VPN, I think

I do not guarantee that none of the above web links will be dead/broken.