Open-ended, extensible system utterances are preferred, even if they require filled pauses

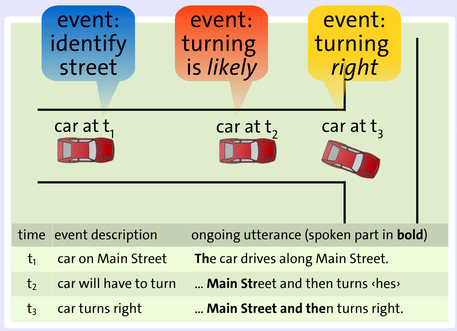
Timo Baumann • Natural Language Systems Division • Department of Informatics • Universität Hamburg • Germany • baumann@informatik.uni-hamburg.de
David Schlangen • Dialogue Systems Group • Faculty of Linguistics and Literature • Bielefeld University • Germany • david.schlangen@uni-bielefeld.de

Abstract

As situations evolve, speakers are able to start talking without having prepared their full utterance. This enables speakers to start commenting about events for which the *outcome is not yet known*, resulting in time pressure for the generation of a completion of the partial utterance. *Temporal overcommitment* may occur, which can be (somewhat) resolved by introducing a hesitation.

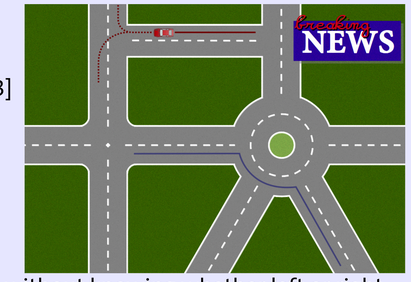
We investigate the impact of incremental spoken output [1] in a domain where only few utterances could be finished as planned due to the high rate of change.
We find that users prefer this system over baseline behaviour, even in cases when the system has to use a hesitation to resolve temporal overcommitment.

Incremental Commenting in a **Highly Dynamic** Environment



time	event description	ongoing utterance (spoken part in **bold**)
t_1	car on Main Street	**The** car drives along Main Street.
t_2	car will have to turn	... **Main Str**eet and then turns ‹hes›
t_3	car turns right	... **Main Street and the**n turns right.

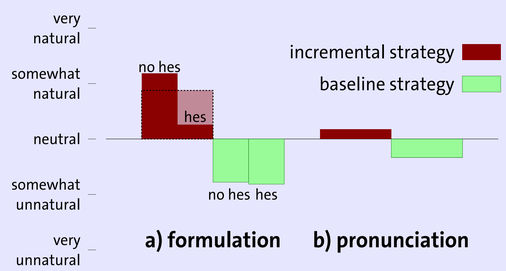
- highly dynamic ≙ **high event density**
  - most utterances could not be finished as planned
  - e.g. sports commentary [2], **route descriptions**/navigation [3]
- Conventional: talk about individual events
  - in order to not accumulate delays, need to abort/skip utterances
- **Our Strategy**: *combine events* to complex descriptions
  - start early, extend later (using *incremental synthesis*)
  - make use of *underspecified events*: assume the car will turn, without knowing whether left or right
  - risk *temporal overcommitment*: the car may go more slowly than anticipated



Experiment and **Results:**

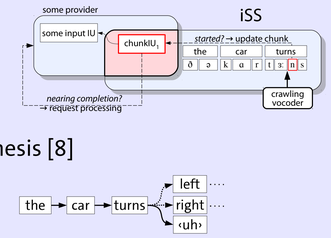
incrementally expanded utt's. (imperceptibly) > incrementally expanded utt's. (w/ hesitations) > short utt's, one per event (standard)

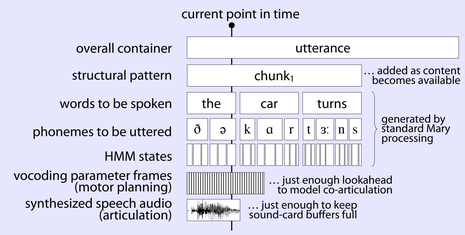
- two settings:
  - baseline (individual events, skip/abort optional utterances in favour of non-optional utterances)
  - incrementally extended utterances, using underspecified prediction events
- 9 scenarios (4 configurations × timing variations)
  - temporal overcommitment in 3 scenarios
- rating (on 5-point Likert scale)
  - naturalness of formulation
  - naturalness of pronunciation
- 9 subjects
  - a total of 81 paired samples for each question
- Results:
  - preference for incremental formulations
  - preference even in cases of temporal overcommitment
- even a preference in pronunciation
  - likely a carry-over effect from formulation
  - *interactional adequacy* may be more inportant than raw synthesis quality

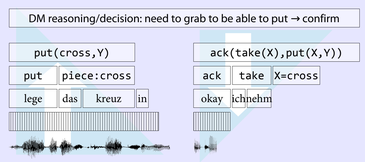


Implementation and Integration

- Incremental Processors in the IU model [4]
- utterance plans [5] extended and changed until immediately before the fact
- uses Inpro_iSS [1] based on InproTK [6] and MaryTTS [7] with incremental HMM synthesis [8]
- iSS provides:
  - extension/change/abortion of ongoing synthesis
  - feedback on delivery,
  - automatically aborted hesitations, and
  - adaptation of prosody in the vicinity of hesitations
- uses *just-in-time* processing in a *triangular data model*:
  - detailed, concrete data for the near future
  - underspecified, abstract data further ahead







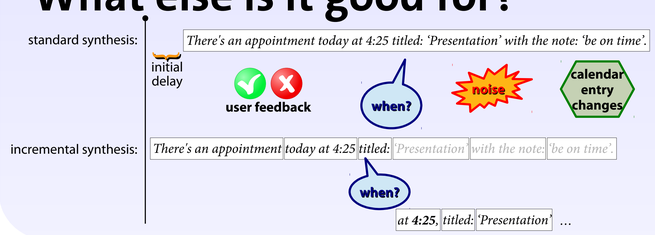
Open Source!

Our software for incremental dialogue processing is available as open source:

- **inprotk.sf.net** for the source code and documentation
- **www.inpro.tk** for more information on the Inpro project

We value your feedback to inprotk-devel@lists.sourceforge.net !

Incremental Speech Synthesis: **What else is it good for?**



References:

[1] T. Baumann and D. Schlangen: „Inpro_iSS: A component for just-in-time incremental speech synthesis," in *Procs. of ACL System Demonstrations*, Jeju, Korea, 2012.
[2] D. Chen and R. Mooney: „Learning to sportscast: A test of grounded language acquisition," in *Proceedings of ICML*, Helsinki, Finland, 2008.
[3] K. Lohmann, O. Eichhorn and T. Baumann: „Generating situated assisting utterances to facilitate tactile-map understanding," in *Procs. of SLPAT*, Montréal, 2012.
[4] D. Schlangen and G. Skantze: „A general, abstract model of incremental dialogue processing," in *Proceedings of EACL*, Athens, Greece, 2009.
[5] G. Skantze and A. Hjalmarsson: „Towards incremental speech generation in dialogue systems," in *Proceedings of SIGDial*, Tokyo, Japan, 2010.
[6] T. Baumann and D. Schlangen: „The InproTK 2012 release," in *Proceedings of SDCTD*, Montréal, Canada, 2012.
[7] M. Schröder and J. Trouvain: „The German Text-to-Speech synthesis system MARY: A tool for research, develeopment and teaching," *Int. J. of Speech Tech*, **6**(3), 2003.
[8] T. Dutoit, M. Astrinaki, O. Babacan, N. d'Alessandro and B. Picart: „pHTS for Max/MSP: A streaming architecture for statistical arametric speech synthesis," numediart Research Program on Digital Art Technologies, Tech. Rep. 1, Mons University, 2011.