Timo Baumann*, David Schlangen** | * University of Hamburg, Germany | ** Bielefeld University, Germany

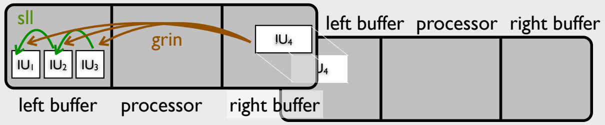
the INPROTK 2012 release

Abstract We describe the 2012 release of our "incremental processing toolkit" (inproTK) which combines a powerful and extensible architecture for incremental processing with components for incremental speech recognition and, new to this release, incremental speech synthesis. More domain-specific components such as NLU and DM are provided as example implementations.

The toolkit realises the IU-Model of incremental processing of (Schlangen & Skantze 2009, 2011).

The IU-Model (S&S'09,'11) Basic notions:

- IU: Incremental Unit, minimal unit of information to be passed around between *modules* of processing system.
- Module: Consists of *left buffer, processor,* and *right buffer.*
- Operations:
  - add: new IUs are posted by one module for next one.
  - purge / revoke: IUs that were wrongly hypothesized are "taken back". (E.g., "four" becomes "fourty".)
  - commit: module signals that it will not revoke IU.



Architecture

- data stored as network of incremental units (IUs)
- shared memory, single process (bridges possible to external processes)
- base IU class implements basic "housekeeping" functionality and abstract methods
- derived IU types implement specific IU types (e.g., wordIU, semIU, daIU, etc.)

- enables different *schemes* of incremental processing:
  - asynchronous, concurrent processors that use efficient update method (IUs are added / revoked)
  - update listeners that register with IUs, more fine-grained event-based processing

Developer Support

- easy configuration of module network via config file
- viewer for incremental data
- transparently swap in text in/output for debugging

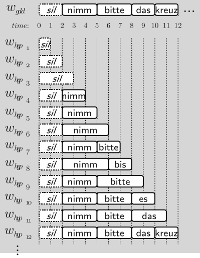
Benefits to Community

- (relatively) easy to set up and get started
- comes with example systems highlighting certain capabilities
- open source, documented (...), actively maintained
- components to benchmark against

- beyond code, conceptual framework may be helpful
- validated evaluation metrics for incremental processing (Baumann *et al.* 2011)

ASR

- based on Sphinx-4
- supports all Sphinx-4 acoustic and language models
- updating results every 10ms
- configurable filtering methods that control trade-off between timeliness and hypothesis stability (Baumann *et al.* 2009)
- often word is recognised while still being spoken

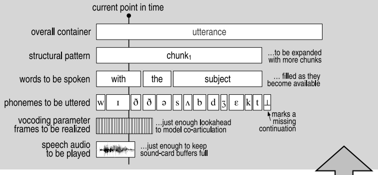

w_{gld}
sil
nimm
bitte
das
kreuz
...
time: 0 1 2 3 4 5 6 7 8 9 10 11 12
$w_{hyp\ 1}$ sil
$w_{hyp\ 2}$ sil
$w_{hyp\ 3}$ sil
$w_{hyp\ 4}$ sil nimm
$w_{hyp\ 5}$ sil nimm
$w_{hyp\ 6}$ sil nimm
$w_{hyp\ 7}$ sil nimm bitte
$w_{hyp\ 8}$ sil nimm bis
$w_{hyp\ 9}$ sil nimm bitte
$w_{hyp\ 10}$ sil nimm bitte es
$w_{hyp\ 11}$ sil nimm bitte das
$w_{hyp\ 12}$ sil nimm bitte das kreuz
⋮

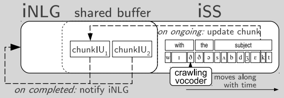
NLU various example implementations

- simple concept spotter
- incremental parser, building up fully connected RMRS representations (Peldszus *et al.* 2012)

iTTS

- based on MaryTTS
- incremental, just-in-time speech synthesis
- makes possible changes to ongoing utterances
- gives feedback on utterance progress
- reduces utterance-initial delay to <200ms, while still employing optimization over full input string


current point in time
overall container
utterance
structural pattern
chunk₁
...to be expanded with more chunks
words to be spoken
with
the
subject
... filled as they become available
phonemes to be uttered
w ɪ ð ð ə s ʌ b d ʒ ɛ k t ⊥
marks a missing continuation
vocoding parameter frames to be realized
...just enough lookahead to model co-articulation
speech audio to be played
...just enough to keep sound-card buffers full



(Buschmeier *et al.* 2012)

DM various example implementations

- QUD-based DM (Buß *et al.* 2010)
- DM based on IU operations (Buß & Schlangen 2011)

References

- Hendrik Buschmeier, Timo Baumann, Benjamin Dosch, Stefan Kopp, and David Schlangen. Combining incremental language generation and in- cremental speech synthesis for adaptive information presentation. In Proceedings of SIGdial 2012, Seoul, South Korea, July 2012
- Andreas Peldszus, Okko Buß, Timo Baumann, and David Schlangen. Joint satisfaction of syntactic and pragmatic constraints improves incremental spoken language understanding. In Proceedings of EACL 2012, Avignon, France, April 2012
- Okko Buß and David Schlangen. DIUM – an incremental dialogue manager that can produce self-corrections. In Proceedings of semdial 2011 (Los Angelogue), Los Angeles, CA, USA, September 2011
- Okko Buß, Timo Baumann, and David Schlangen. Collaborating on utterances with a spoken dialogue system using an ISU-based approach to incremental dialogue management. In Proceedings of SIGdial 2010, pages 233–236, Tokyo, Japan, September 2010
- Timo Baumann, Okko Buß, and David Schlangen. Evaluation and optimization of incremental processors. Dialogue and Discourse, 2(1):113–141, 2011
- Timo Baumann, Michaela Atterer, and David Schlangen. Assessing and improving the performance of speech recognition for incremental systems. In Proceedings of NAACL/HLT 2009, Boulder, Colorado, USA, May 2009

http://sourceforge.net/projects/inprotk

http://inpro.tk

This work was generously supported by the DFG in the Emmy Noether Programme.