

Incremental, Adaptive and Interruptive Speech Realization for Fluent Conversation with ECAs

Herwin van Welbergen¹, Timo Baumann², Stefan Kopp¹,
and David Schlangen^{3,*}

¹ Sociable Agents Group, CITEC, Fac. of Technology, Bielefeld University

² Department of Informatics, Natural Language Systems Division,
Universität Hamburg

³ Dialogue Systems Group, Faculty of Linguistics, Bielefeld University

1 Introduction

Human conversations are highly dynamic, responsive interactions. In such interactions, utterances are produced incrementally, subject to on-the-fly adaptation (e.g. speaking louder to keep a challenged turn) and (self) interruptions. While listening, plans for next speaking contributions are constructed, allowing very rapid turn transitions. To enable such fluent interaction in Embodied Conversational Agents (ECAs) we must steer away from the traditional turn-based non-incremental interaction paradigm in which the ECA first fully analyzes user contributions and subsequently fully plans its contribution, which is then executed entirely ballistically (providing no adaptation in nor interruption of ongoing behavior). Recently, several systems have done exactly this and introduce one or more aspects of incrementality, interruptibility, or adaptivity. Their focus is mostly on behavior planning and they introduce a limited set of behavior realization capabilities only where it helps illustrate their flexible planning strategies. Furthermore, many of these systems can be characterized as proof-of-concepts, designed for a single purpose, domain (for example, only generating backchannel feedback) or set of experiments. The focus of our ongoing work is to provide a comprehensive architecture that unifies the fluent *behavior realization* functionality of these more experimental systems and additionally can reproduce other important phenomena that occur in fluent dialog. Our architecture serves 1) as a platform for those experimenting with the ‘best’ way to deploy fluent behavior realization strategies or those researching social effects of certain deployment strategies and 2) as a building block (specifically the Behavior Realizer) in a ECA architecture that supports fluent human-ECA interaction.

To this end, we provide behavior planners and human authors of behavior realization strategies with a language for specifying behavior realization plans that allow fluent interaction on an ECA. Our specification language and realizer implementation provides:

* This research and development project is funded by the German Federal Ministry of Education and Research (BMBF) within the Leading-Edge Cluster Competition and managed by the Project Management Agency Karlsruhe (PTKA). The author is responsible for the contents of this publication.

1. Incremental construction of behavior.
2. Smooth (prosodic) concatenation of the increments (e.g. successive phrases in speech).
3. Graceful, any-time interruption of ongoing behavior.
4. Mechanisms to adapt certain parameters in ongoing behavior (e.g. pitch, loudness) at fine granularity (at least at syllable level).
5. Mechanisms to preplan behavior that can potentially be executed instantly later on. Multiple execution alternatives can be preplanned concurrently.
6. The realization of apositional beginnings (e.g. uhm) to keep or take the turn without having a plan at hand.

2 Implementation

The requirements for fluent dialog realization are satisfied by combining the flexible BML Realizer *AsapRealizer* [1] (and its specification language BMLA) with the incremental TTS system *Inpro_iSS* [2] and by extending both systems with new features. Requirement 1 is partly covered by BMLA: behavior plans can be composed by specifying successor relations between blocks. BMLA does however not provide the specification mechanisms required to postpone ongoing behavior. In addition to allowing one to specify that a BML block should be inserted *after* blocks already in the plan, we therefore contribute a specification mechanism that allow us to plan BML blocks *before* specific other BML blocks in a plan. This is especially useful to plan short contributions (e.g. repairs) *before* some already planned behavior and then continue with the existing plan. Requirement 2) is satisfied by *Inpro_iSS*'s capability to provide smooth prosodic connections between speech increments. To satisfy requirement 3, we combine BMLA's specification mechanisms for interruption with *Inpro_iSS* capability to interrupt speech at phoneme level. BMLA's parameterization mechanisms are flexible enough to be directly applicable for speech parameterization to satisfy requirement 4. We have implemented adaptable parameters in speech (e.g. loudness, tempo, pitch) using *Inpro_iSS*'s speech adaptation mechanisms. BMLA's preplanning satisfies requirement 5 and fits well with parallel plan construction. To realize requirement 6, we contribute the specification and automatic realization of speech fillers.

We have thus contributed to both the specification and implementation of the realization of incremental, adaptive and interruptive behavior. This contribution is implemented in *AsapRealizer*, allowing it to support the flexible realization of (synchronized) gesture and speech.

References

1. van Welbergen, H., Reidsma, D., Kopp, S.: An incremental multimodal realizer for behavior co-articulation and coordination. In: Nakano, Y., Neff, M., Paiva, A., Walker, M. (eds.) IVA 2012. LNCS, vol. 7502, pp. 175–188. Springer, Heidelberg (2012)
2. Baumann, T., Schlangen, D.: *Inpro_iSS*: A component for just-in-time incremental speech synthesis. In: ACL System Demonstrations, pp. 103–108 (2012)