

# Phonetic Adaptation Module for Spoken Dialogue Systems

Eran Raveh<sup>1,2</sup>, Ingmar Steiner<sup>1-3</sup>

<sup>1</sup>Multimodal Computing and Interaction, Saarland University

<sup>2</sup>Computational Linguistics & Phonetics, Saarland University

<sup>3</sup>DFKI Language Technology Lab, Saarbrücken

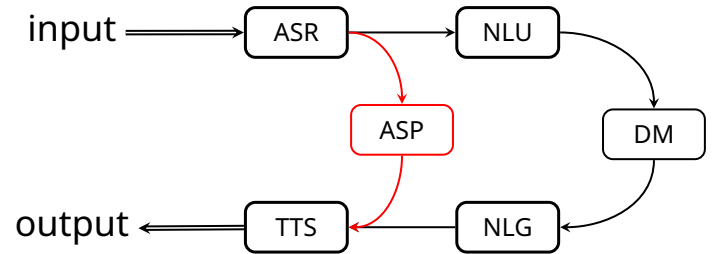
## 1 Phonetic Convergence

- Increase in segmental and suprasegmental similarities between two speakers [1]
- Occurs naturally, like gesture and posture assimilation
- Found in a shadowing experiment with natural and synthetic stimuli [2]

### Current research

- Can dialogue systems simulate convergence behavior?
- First integration of convergence model [3]
- Will this lead to an easier and more fluent interaction?

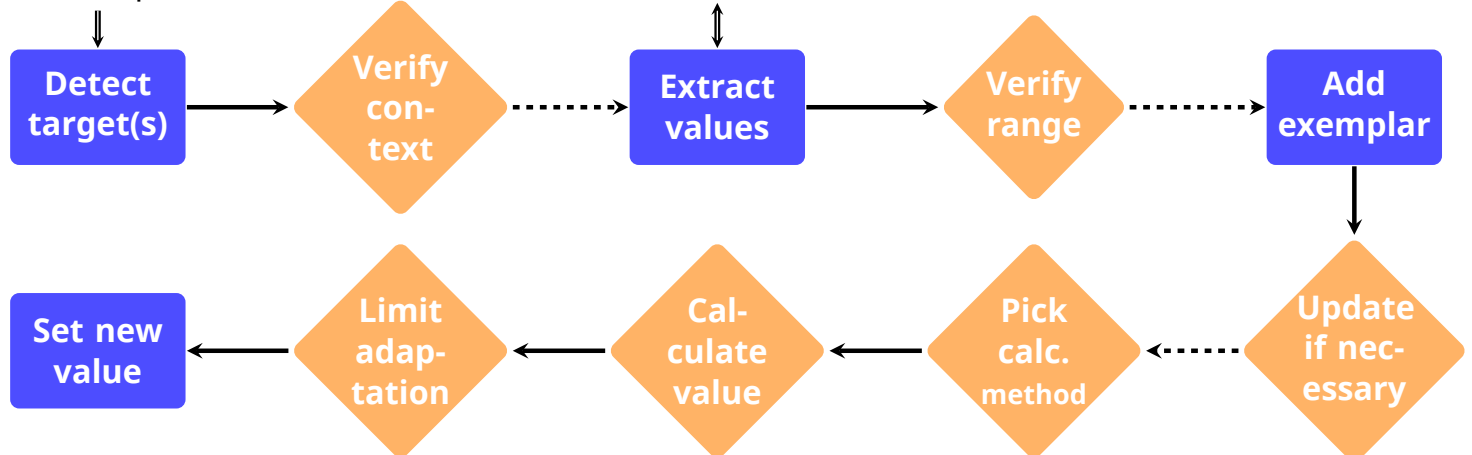
## 2 Architecture



The ASP module creates a direct link between the ASR and the TTS modules, so that not only plain text output is used.

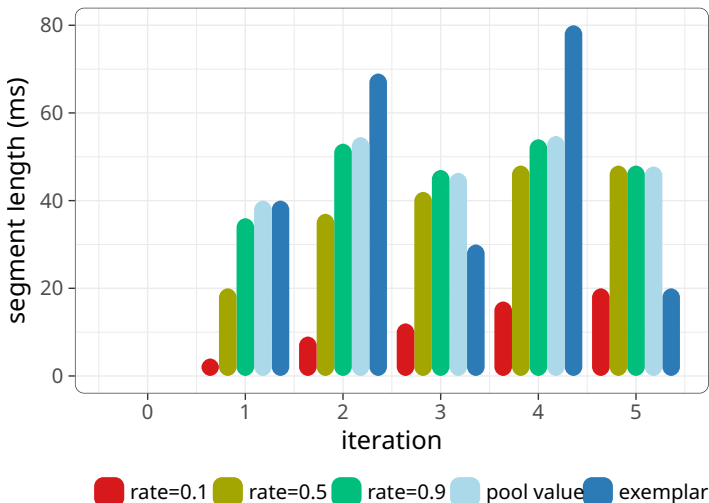
## 3 The ASP Module

Audio input



Overview of the adaptation pipeline integrated into the additional speech processing (ASP) module, with Praat as the signal processing back-end. Mandatory, fixed steps are marked by blue rectangles and parameterized steps by orange diamonds. Dashed arrows mark conditional transitions that terminate the process if they are not fulfilled.

## 4 Evaluation



## 5 Future Work

- More segmental and suprasegmental features
- Better synthesis technique
- User Evaluation in a task-specific system

## References

- [1] J. S. Pardo, "On phonetic convergence during conversational interaction," *Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2382–2393, Apr. 2006.
- [2] I. Gessinger, E. Raveh, S. Le Maguer, B. Möbius, and I. Steiner, "Shadowing synthesized speech – segmental analysis of phonetic convergence," in *Interspeech*, Stockholm, Sweden, Aug. 2017.
- [3] E. Raveh, I. Steiner, and B. Möbius, "A computational model for phonetically responsive spoken dialogue systems," in *Interspeech*, Stockholm, Sweden, Aug. 2017.