Phonetic Adaptation Module for Spoken Dialogue Systems

Eran Raveh\textsuperscript{1,2}, Ingmar Steiner\textsuperscript{1–3}

\textsuperscript{1}Multimodal Computing and Interaction, Saarland University \textsuperscript{2}Computational Linguistics & Phonetics, Saarland University \textsuperscript{3}DFKI Language Technology Lab, Saarbrücken

1 Phonetic Convergence

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

Current research

\rightarrow Can dialogue systems simulate convergence behavior?
\rightarrow First integration of convergence model \cite{3}
\rightarrow 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

\[
\begin{align*}
\text{Detect target(s)} & \quad \rightarrow \quad \text{Verify context} \\
\downarrow & \quad \rightarrow \quad \text{Extract values} \\
\rightarrow & \quad \rightarrow \quad \text{Verify range} \\
\downarrow & \quad \rightarrow \quad \text{Add exemplar} \\
\rightarrow & \quad \rightarrow \quad \text{Set new value} \\
\rightarrow & \quad \rightarrow \quad \text{Limit adaptation} \\
\rightarrow & \quad \rightarrow \quad \text{Calculate value} \\
\rightarrow & \quad \rightarrow \quad \text{Pick calc. method} \\
\rightarrow & \quad \rightarrow \quad \text{Update if necessary}
\end{align*}
\]

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

\[
\begin{align*}
\text{Segment length (ms)} & \quad \text{iteration} \\
\text{rate=0.1} & \quad \text{rate=0.5} \\
\text{rate=0.9} & \quad \text{pool value} \\
\text{exemplar}
\end{align*}
\]

5 Future Work

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

References

