Automatic Analysis of Segmental Features in a Real-Time Phonetic Convergence Pipeline

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

\texttt{(raveh|steiner)@coli.uni-saarland.de}

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

1 Phonetic Convergence

- Increase in segmental and suprasegmental similarities between two speakers [1]
- Found in a shadowing experiment with natural and synthetic stimuli [2]

Current research

- Integrating convergence behavior into spoken dialogue systems, based on a computational model [3]
- Does Praat suffice as an analysis back-end?
- Does Praat suffice as a synthesis engine?

2 Architecture

\begin{tikzpicture}[auto, node distance=2cm, >=latex, every node/.style={align=center}]

  \node (input) {input};
  \node (asr) [right of=input] {Automatic speech recognition (ASR)};
  \node (nlu) [right of=asr] {Natural language understanding};
  \node (asp) [right of=nlu] {Additional speech processing (ASP)};
  \node (di) [right of=asp] {Dialogue manager};
  \node (tts) [below of=di] {Text-to-speech (TTS)};
  \node (nlg) [right of=tts] {Natural language generation};

  \draw[->] (input) -- (asr);
  \draw[->] (asr) -- (nlu);
  \draw[->] (nlu) -- (asp);
  \draw[->] (asp) -- (di);
  \draw[->] (asp) -- (tts);
  \draw[->] (tts) -- (nlg);

\end{tikzpicture}

The ASP module creates a direct link between the ASR and the TTS modules, so that not only plain text output is used. ASP internally uses Praat [4] for signal processing.

3 ASP Module Pipeline Core

\begin{tikzpicture}[node distance=2cm, >=latex, every node/.style={align=center}]

  \node (input) {Audio input};
  \node (scripts) [above of=input] {scripts};
  \node (pd) [below of=scripts] {Phoneme detection};
  \node (vpc) [right of=pd] {Verify phonetic context};
  \node (ev) [right of=vpc] {Extract values};
  \node (vr) [right of=ev] {Verify range};
  \node (anv) [right of=vr] {Apply new value};

  \draw[->] (input) -- (scripts);
  \draw[->] (scripts) -- (pd);
  \draw[->] (pd) -- (vpc);
  \draw[->] (vpc) -- (ev);
  \draw[->] (ev) -- (vr);
  \draw[->] (vr) -- (anv);

\end{tikzpicture}

The ASP module pipeline core includes audio input, phoneme detection, verify phonetic context, extract values, verify range, and apply new value.

4 Discussion

- Sufficiently accurate analysis \hspace{1cm} \checkmark \hspace{1cm} \text{good for analysis!}
- Good performance in real time \hspace{1cm} \checkmark
- Simple script implementation \hspace{1cm} \checkmark
- Outdated, local synthesis \hspace{1cm} \xmark
- Based on existing files \hspace{1cm} \xmark
- Requires external processing and temporary files \hspace{1cm} \xmark

5 Outlook

- More segmental and suprasegmental features
- Better synthesis technique
- Integration into an end-to-end system
- User evaluation in a task-specific system

References