EXEMPLARBASIERTE SPRACHPRODUKTION UND UNIT SELECTION-SYNTHESE

Bernd Möbius

FR 4.7, Phonetik, Universität des Saarlandes
moebius@coli.uni-saarland.de

Sprachsynthese nach der Methode der Unit Selection wählt aus einem Korpus akustische Einheiten aus, die nach Verkettung eine gegebene Zieläußerung optimal repräsentieren. Bei der Auswahl werden die akustischen Einheiten bewertet, ob sie einerseits gute Repräsentanten der jeweiligen Zieleinheit sind und sich zugleich gut mit den benachbarten Einheiten verketten lassen. Wissenschaftshistorisch gesehen ist die Unit Selection-Methode nicht durch Modelle der menschlichen Sprachproduktion inspiriert. Exemplarbasierte Modelle der Sprachproduktion, die in den letzten Jahren zunehmend Aufmerksamkeit finden, legen jedoch eine Analogie zwischen Modellen der Sprachproduktion und Verfahren der Sprachsynthese nahe.

Im Context Sequence Model (CSM, Wade et al. 2010) [1] werden akustische Ziele der Sprachproduktion durch die Auswahl von Einheiten aus einem Gedächtnisspeicher bestimmt. Der Speicher enthält eine große Anzahl früher wahrgenommener (oder produzierter) sprachlicher Einheiten, die sowohl abstrakt (phonologisch) indiziert als auch entlang vieler akustischer Dimensionen spezifiziert sind. Die Signale im Gedächtnisspeicher entsprechen langen Sequenzen kontinuierlicher Sprache, so dass individuelle Sprachlaute immer in einem größeren Kontext auftreten. Eine zentrale Eigenschaft des Modells besteht darin, dass die Auswahl von Exemplaren für die Produktion auf einer Bewertung der Ähnlichkeit zwischen dem Kontext, in dem sie ursprünglich aufgetreten waren, und dem aktuellen Produktionskontext beruht.

Simulationen der Sprachproduktion mit realistischen akustischen Daten zeigen, dass die optimale Auswahl von kontextadäquaten Einheiten auf Lautebene die Berücksichtigung eines linken und rechten Kontextes von etwa 0,5 Sekunden erfordert. Die Ergebnisse legen außerdem die Interpretation nahe, dass die kontextabhängige Produktion auf der Lautebene einige Effekte der Auftretenshäufigkeit bedingt, die bislang als Effekte auf Silben-,Wort- und anderen höheren Ebenen der sprachlichen Organisation betrachtet wurden.

Dieser Beitrag widmet sich der Frage, ob die Analogie zwischen Modellen der Sprachproduktion und Verfahren der Sprachsynthese oberflächlicher Art ist oder ob die komputationelle Simulation von Prozessen der menschlichen Sprachverarbeitung die Implementierung sprachtechnologischer Verfahren informieren kann.

Literatur

[1] Wade, T., G. Dogil, H. Schütze, M. Walsh und B. Möbius: Syllable frequency effects in a context-sensitive segment production model. Journal of Phonetics, 38(2):905–945, 2010.