

# Evaluation eines Sprechers für schnell gesprochene Sprache in der Unit-Selection basierten Sprachsynthese

Donata Moers<sup>1</sup>, Petra Wagner<sup>2</sup>

<sup>1</sup>IfK, Sprache und Kommunikation, Universität Bonn

<sup>2</sup>Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld

Die hier vorgestellte Sprecherevaluation ist Teil eines Projekts zur Modellierung hoher Sprechgeschwindigkeit in der Unit-Selection basierten Sprachsynthese. Insbesondere Blinde und Sehbehinderte bevorzugen bei häufiger Anwendung von Sprachsynthesystemen eine hohe Sprechrate. Diese wird in Unit-Selection basierten Synthesystemen bisher aber nur unzureichend modelliert. Um für die Modellierung schnell gesprochener Sprache ein möglichst hochwertiges Syntheseinventar erstellen zu können, wurde unter Vorgabe bestimmter Kriterien ein geeigneter Sprecher gesucht. Nach der Durchführung einer Vorauswahl sollte die tatsächliche Eignung des Sprechers zum Aufsprechen eines schnell gesprochenen Bausteininventars mittels eines Perzeptionstests evaluiert und bestätigt werden.

Je schneller ein Sprecher spricht, desto unverständlicher werden seine Äußerungen meist. Dies liegt unter anderem daran, dass die einzelnen Laute bei hoher Sprechgeschwindigkeit stärker koartikuliert und reduziert werden, und sich so ihre charakteristischen akustischen Merkmale verändern. Auch größere Einheiten wie Silben oder Intonationsphrasen sind in schnell gesprochener Sprache von Veränderungen betroffen; so werden Silben verkürzt, die Anzahl und Stärke der Phrasengrenzen nimmt ab und die Grundfrequenz weist einen flacheren Verlauf auf. Bei der Modellierung schneller Sprache in der Unit-Selection basierten Sprachsynthese sind diese Phänomene größtenteils unerwünscht, da sie nicht nur die Verständlichkeit der erzeugten Sprache negativ beeinflussen, sondern auch die Definition der Auswahlseinheiten wesentlich erschweren.

Es hat sich immer wieder gezeigt, dass die Qualität synthetischer Sprache zum Großteil vom Sprecher des Bausteininventars determiniert wird. Ausgebildete Sprecher, die über einen längeren Zeitraum mit gleich bleibender Stimmqualität und hoher artikulatorischer Präzision zu sprechen gelernt haben, produzieren in der Regel hochwertigere Synthesebausteine. Bauen die Inventare auf schnell gesprochener Sprache auf, verschärfen sich die möglicherweise auftretenden Probleme, da zunächst davon auszugehen ist, dass ein Sprecher bei diesem Sprechstil seine Artikulationsgenauigkeit der Sprachökonomie zumindest teilweise opfern wird. Eine mögliche Antwort auf die Frage, ob es möglich ist, das Auftreten unerwünschter Phänomene in schneller Sprache weitestgehend zu vermeiden, liefert die von Lindblom aufgestellte H&H-Theorie. Sie besagt, dass trotz des kontinuierlichen Sprachverlaufs und dadurch bedingter Koartikulation eine ausreichende Kontrastierung akustischer Signale für den Sprecher möglich ist. Macht das Kommunikationsziel es erforderlich, könnte es einem geeigneten Sprecher somit durchaus möglich sein, sehr schnell und trotzdem deutlich zu sprechen.

Mittels eines globalen Präferenztests wurde aus einem Pool von 9 potentiellen Sprechern derjenige ausgewählt, dessen schnelle Sprache am verständlichsten und dabei noch am natürlichsten war. Um vor der Erstellung eines umfangreicheren Bausteininventars für die Sprachsynthese zu verifizieren, dass die ausgewählte Sprecherin den genannten Kriterien tatsächlich genügt und sie insbesondere in der Lage ist, bei Bedarf auch bei hoher Sprechgeschwindigkeit deutlich zu sprechen, wurden weitere Aufnahmen in normaler sowie in intendiert zunehmender Sprechgeschwindigkeit gemacht; außerdem wurden 3 schnelle Versionen mit intendiert zunehmender Sprechgeschwindigkeit erstellt, bei denen die Sprecherin mehrfach aufgefordert wurde, bewusst den Artikulationsaufwand zu erhöhen und besonders deutlich zu sprechen.

Für die perzeptive Evaluation wurde ein Experiment erstellt, das aus insgesamt 9 Teilexperimenten bestand. In jedem Teilexperiment wurde derselbe Ausschnitt der unterschiedlich schnell und deutlich gesprochenen Versionen in einer paarweisen Gegenüberstellung bezüglich seiner Deutlichkeit beurteilt. Es war davon auszugehen, dass die bewusst deutlicher artikulierte Versionen bei ähnlicher Sprechgeschwindigkeit über alle Ausschnitte als besser bewertet werden würden. Aufgrund der vorgestellten Ergebnisse der perzeptiven Evaluation wird deutlich, dass es der ausgewählten Sprecherin tatsächlich möglich ist, Koartikulation und übermäßige Reduktion in bewusst deutlich artikulierter, schnell gesprochener Sprache zu vermeiden und sie somit als Sprecherin für das Aufsprechen eines schnell gesprochenen Bausteininventars für die Unit-Selection basierte Sprachsynthese geeignet ist.

- [1] Lindblom, B. (1990): Explaining phonetic variation: A sketch of the H&H-Theory. In *Hardcastle, W.J.; Marchal, A.: Speech Production and Speech Modelling*. Dordrecht: Kluwer. S. 403 – 439.
- [2] Moers, D., & Wagner, P. (2008): Evaluation eines Sprechers für schnell gesprochene Sprache in der Unit-Selection basierten Sprachsynthese. In *ITG-Fachtagung Sprachkommunikation*. Aachen.