Wie relevant ist die Phonetik noch für die Sprachsynthese?

Petra Wagner

Universität Bielefeld
petra.wagner@uni-bielefeld.de

Kurzfassung: Die sprachtechnologischen Entwicklungen der letzten Jahre haben die Notwendigkeit für traditionelle phonetische Systemanteile, z.B. Modelle für die Bestimmung akustischer Parameter, immer geringer werden lassen. Auch in der Evaluation ersetzen objektive Messverfahren mittlerweile immer mehr aufwändige Perzeptionstests. Dieser Trend spiegelt sich auch in den Veröffentlichungen zur Sprachsynthese der letzten Jahre wider. Es stellt sich daher die Frage, welche Rolle die Phonetik jenseits der Erstellung von Aussprachelexika und Trainingsdatenbanken in Zukunft überhaupt spielen kann bzw. sollte.

Die Antwort ergibt sich direkt aus einer Analyse der gegenwärtigen Schwächen synthetischer Sprache. Genau da, wo die Synthesesysteme derzeit an ihre Grenzen stoßen, kann die Phonetik möglicherweise nach wie vor wichtige Impulse liefern. Eine ähnliche Einsicht setzt sich derzeit in der Spracherkennungsforschung durch, wo verstärkt menschliche Sprachentwicklungs- und Lernmodelle berücksichtigt werden.

Die aktuellen Synthesen liefern qualitativ sehr verständliche, weitgehend natürlich – im Sinne von menschlich – klingende Sprache und werden immer besser in der Generierung verschiedener Sprechstile und unterschiedlicher Sprecher, z.B. durch *Voice Morphing*. Trotz dieser Möglichkeiten sind die bestehenden Systeme in aller Regel wenig adaptiv, und scheinen im Regelfall auf eine/n Durchschnittshörer/in zugeschnitten. Dieser Ansatz ignoriert, dass jeder Mensch gewissermaßen ein *Special Listener* ist, also jemand mit situations- und personenspezifisch besonderen Bedürfnissen an den Synthesestil. Eine solche Betrachtung geht weit über die Unterscheidung zwischen "normalen" und "speziellen" Hörergruppen wie Blinde, Fremdsprachenlerner etc. hinaus. Menschen sind hingegen in der Regel sehr gut in der Lage, ihre Sprechstile situations-, hörer- und textspezifisch angemessen anzupassen. Ein weiterer Punkt ist die Frage nach dem Adäquatheit bestimmter Sprechstile im Kontext einer Sprachsyntheseanwendung. Zwischenmenschliche Kommunikationsstrategien können ein Vorbild sein, wir müssen aber damit rechnen, dass sich die Hörererwartungen an Synthese sich von denen an menschliche Interaktionspartner stark unterscheiden. Wir benötigen daher eine weitergehende Erforschung zwischenmenschlicher sprachlicher und phonetischer Anpassungsprozesse, hörerspezifischer Erwartungen an synthetische Sprache sowie nutzerzentrierter Evaluationstechniken. Eine entsprechende Forschungsausrichtung könnte im hohen Maße von den technologischen Neuentwicklungen der letzten Jahre profitieren.