

Press release**Universität des Saarlandes****Claudia Ehrlich**

12/13/2012

<http://idw-online.de/en/news512013>Research projects
Information technology, Language / literature
transregional, national**Berechenbare Sprache – Wie künstliche Stimmen menschlicher klingen**

Was macht eine menschliche Stimme aus? Bernd Möbius entschlüsselt die gesprochene Sprache, um herauszufinden, wie künstliche Stimmen mehr Charakter, mehr „menschlichen Touch“ bekommen. Hierzu übersetzt der Saarbrücker Professor für Phonetik und Phonologie Laute in Zahlen und sucht und glättet Störstellen mit einem Rechenverfahren. Ziel ist ein selbstlernendes mathematisches Modell der menschlichen Sprache, das es möglich macht, jedem beliebigen Gegenstand jede beliebige Stimme zu verleihen – ohne künstlich zu klingen.

„Den habe ich mir ganz anders vorgestellt“ - das Phänomen tritt zu Tage bei Radiomoderatoren oder Leuten, die bisher nur am Telefon miteinander zu tun hatten: Wer Menschen ausschließlich von ihrer Stimme her kennt, macht sich ein bestimmtes Bild. Da kann eine junge Frau älter wirken, ein kräftiger Mann dünner oder ein blonder Haarträger eher dunkelhaarig. Menschliche Stimmen wecken die Phantasie. „Betrachtet man die Hirntätigkeit beim Hören, ruft eine natürliche Stimme Aktivität in Arealen hervor, die für Gefühle und Assoziationen verantwortlich sind. Bei der klassischen Computerstimme ist das anders. Hört der Mensch eine künstliche Stimme bleiben diese Areale stumm“, erläutert Professor Bernd Möbius, Saarbrücker Experte für Sprachproduktion.

Zwar sind moderne Computerstimmen in Auskunfts- oder Dialogsystemen vom blechernen Klang der abgehackten Wörter ohne Betonung bereits weit entfernt. Trotzdem hört das verwöhnte und seit jeher auf Stimmen spezialisierte menschliche Ohr die feinen Unterschiede, ob Mensch oder Maschine spricht. Werden Sätze aus Laut- und Wortschnipseln zusammengesetzt, entlarvt es selbst feinste Sprünge sofort. Fließende Sprachmelodien und sonstige Eigenheiten machen den besonderen Charakter natürlicher Stimmen aus. Fehlen sie, klingt die Stimme künstlich – und sie weckt auch keinerlei Gefühl.

Bernd Möbius forscht daran, diese Charakteristika der menschlichen Stimme herauszufinden, um sie in künstliche Stimmen hineinzurechnen und Sprungstellen und Störfaktoren aus ihnen herauszuholen. „Der Hörer soll sich eine Person hinter der Stimme vorstellen“, sagt er.

Hierzu begibt sich der Forscher mit seinem Team gewissermaßen auf die mikroskopische Ebene und betrachtet die Sprache in ihren kleinsten Einzelteilen. Zugrunde liegt ein digitalisierter Textkorpus, den ein Sprecher im Tonstudio eingesprochen hat. Die Phonetiker verwenden unter anderem die so genannte „Diphonsynthese“. Ein Diphon ist ein kurzer Sprach-Abschnitt, der in der Mitte eines Lautsegments beginnt und in der Mitte des folgenden Lautsegments endet. „Unsere Sprache kennt 45 Laute und etwa 2000 Diphone, jedes davon ist etwa 100 Millisekunden lang. Mit diesem Instrumentarium können wir auf lautlicher Ebene die gesamte Sprache abdecken“, erläutert Möbius.

In den Diphonen liegen die größten Probleme der künstlichen Stimmen verborgen: Sie enthalten etwa den Übergang zwischen den Lauten – winzige Schallsegmente, die bei der Verknüpfung der Sprachbausteine die verräterischen Sprünge hinterlassen. Diese Übergänge verkettet Möbius neu und glättet sie auf diese Weise, wodurch unstete Holperer und Sprungstellen aus der Computersprache verschwinden. „Anzahl und Häufigkeit der Übergänge lassen sich außerdem verringern, wenn es gelingt, längere Bausteine wie Silben oder ganze Wörter, die in den Sprachaufnahmen bereits verfügbar sind, wieder zu verwenden“, erklärt er. Die optimierten Sprachbausteine lassen sich in allen

erdenklichen Kombinationsmöglichkeiten völlig neu zusammensetzen. Mit dieser künstlichen, aber natürlich klingenden Sprache lassen sich beliebige Äußerungen mit unbeschränktem Wortschatz erzeugen.

Das mathematische Sprachsynthese-Modell ist unabhängig von der Stimme des ursprünglichen Sprechers – dadurch ist es auf jede beliebige Stimme übertragbar. Hieraus ergeben sich in Zukunft neben den üblichen Anwendungen in Dialog- oder Auskunftssystemen auch neue Möglichkeiten in der Medizin: „Menschen, die ihre Stimme etwa durch eine Kehlkopfoperation verlieren, könnten so in nicht ferner Zukunft mit einer künstlich erzeugten Stimme sprechen, die wie ihre eigene, natürliche klingt“, stellt Möbius in Aussicht. Die Betroffenen müssten hierfür lediglich ihre Stimme im Tonstudio konservieren, wobei bereits relativ wenig „Sprach-Material“ ausreichen würde – das System könnte den Rest berechnen.

Bei seiner Forschung arbeitet Möbius an der Universität des Saarlandes unter anderem mit Computer- und Psycholinguisten sowie mit Informatikern im Exzellenzcluster „Multimodal Computing and Interaction“ und am Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) zusammen.

Prof. Dr. Bernd Möbius: Tel.: 0681 / 302-4500; E-Mail: moebius@coli.uni-saarland.de
www.coli.uni-saarland.de/~moebius/

Hinweis für Hörfunk-Journalisten: Sie können Telefoninterviews in Studioqualität mit Wissenschaftlern führen, über Rundfunk-Codec (IP-Verbindung). Interviewwünsche bitte an die Pressestelle (0681/302-2601) richten.

URL for press release: <http://www.coli.uni-saarland.de/~moebius/>