Der Faktor Mensch in der Mensch-Maschine-Interaktion

Daniel Duran und **Sarah Warchhold**

– Germanistische Linguistik, Albert-Ludwigs-Universität Freiburg – Bertoldstraße 14, 79108 Freiburg
– daniel.duran@germanistik.uni-freiburg.de | sarah.warchhold@germanistik.uni-freiburg.de

Forschungshintergrund

Verbessert eine natürliche, spontansprachliche Ausdrucksweise der Maschine die Mensch-Maschine-Interaktion (MMI)?

- kognitive Beanspruchung durch (nicht-) Akkomodation?
- Ziel: Nachahmung von menschlichem Kommunikations- und Interaktionsverhalten (Standardsprache)
- technische Aspekte der Entwicklung von Sprachassistenzsystemen (SAS) oft im Vordergrund
- Vorbild *child-directed*, *pet-directed* oder *machine-directed speech*?

Wizard-of-Oz-Experiment

- spielbasiert Vertrauen in die Stimme einer KI/eines Agenten testen
- kollaborative Aufgabenstellung: Kombination von Quiz- und Navigationsaufgabe
- Indiz für Vertrauen: Befolgung von Ratschlägen

Hypothesen

I Anpassung der Maschine an die Varianz in der alltäglichen Sprachnutzung (zB. Dialekte, Akzente) von Menschen kann die MMI im Sinne einer effizienteren und natürlicheren Interaktion verbessern.

II Nicht-standardsprachliche Stimmen einer Maschine brechen die Erwartungshaltung des Menschen und wirken sich dadurch auf die kognitive Beanspruchung der Benutzer_innen aus.

III (Vorstudie) Sprachliche Variation in der MMI kann Irritation hervorrufen und führt daher zu verlängerten Reaktionszeiten und Problemen in der Einordnung von Sprachbeispielen in natürliche oder synthetische Stimmen.

Vorstudie

☞ Welche sprachlichen Eigenschaften werden als synthetisch wahrgenommen und welche nicht?

Design

Bewertung
- von standardsprachlich eingelesenen Nonsense-Sätzen: *„Die rote Giraffe betrachtet kritisch den Aufzug in die Hölle.“*
- auf einer 5-Punkte-Skala von *eher natürlich* bis *eher synthetisch* in Psychopy [4]

Stimuli

- insgesamt 288 Stimuli zur Bewertung (12 Sätze, 3 Sprecher_innen, 8 Manipulationsebenen- und -stufen)
- Manipulation erfolgte auf 3 Ebenen mit PRAAT [2]:

Spektral: Spektrale Diskontinuitäten gefiltert und unterschiedlich eingegliedert (**2** Stufen)

Segmental: Zerschnittenes Sprachsignal: *Schwa* und Artikel *der, die, das* aus anderem Satz extrahiert und in unterschiedlichem Ausmaß im Stimulus ersetzt (**2** Stufen)

Pitch: Resynthese (1.) leicht abgeflachte Intonation (2.) und monotone Intonationskontur (3.) (**3** Stufen)

Proband_innen

- 27 deutsche Muttersprachler_innen (20 w, 6 m, 1 d)
- Bildungshintergrund: 17 Akademiker_innen, 7 Personen mit Abitur, 2 mit Real- und 1 mit Hauptschulabschluss
- unterschiedliche Nutzung von SAS (nach Altersgruppen in Jahren)

	nie	selten	manchmal	täglich
<20	–	2	–	–
20–25	–	4	1	2
26–30	2	6	2	–
31–35	–	5	2	–
>35	–	1	–	–

Erste Ergebnisse

- spektrale Manipulationen der Frikative eher *natürlich* bewertet
- segmentale Manipulationen eher *synthetisch*
- Manipulation der Intonationskontur mit der größten Bandbreite

Linear mixed model

- finales Modell mit *Backwards Elimination*:
  response ∼ speaker + manipulation + RT rel + gender + SAS + (1|participant) + (1|sentence) + (1|trial) + speaker:gender
- positive Estimates für alle Manipulationsstufen
- Kinderstimme tendenziell eher als natürlich bewertet
- erste Intonationsmanipulation wurde nicht *synthetisch* bewertet, die dritte Manipulationsstufe hauptsächlich *synthetisch*

Fazit

- Bewertung von SAS durch *realistischere*, alltagsnahe Aufgabe statt passiver Perzeptionsaufgaben (siehe [5])
- Fokus auf *intelligibility*, aber zunehmend auch psycholinguistische und kognitive Aspekte (zB. kognitive Beanspruchung) (siehe [6])
- *Natürlichkeit* der Interaktion ebenfalls wichtig (siehe [1, 7])

Ausblick

- Vorbild für MMI könnte nach Moore [3] eher Interaktion zwischen Nichtmuttersprachler_innen oder die zwischen Menschen und weniger intelligenten Interaktionspartnern wie Tieren sein
- Basis für Grundlagenforschung und Anwendungsentwicklung
- nächster Schritt: Testen von Vertrauen in MMI in spielbasiertem Wizard-of-Oz-Experiment (unter Einbezug der Ergebnisse der Vorstudien zur Wahrnehmung manipulierter natürlicher Stimmen)

Wir freuen uns über Feedback, Anregungen und Fragen!

Literatur

[1] Štefan Beňuš, Marian Trnka, Eduard Kuric, Lukáš Marták, Agustín Gravano, Julia Hirschberg, and Rivka Levitan. Prosodic entrainment and trust in human-computer interaction. In *9th International Conference on Speech Prosody 2018*, pages 220–224, Poznan, Poland, June 2018. ISCA.

[2] Paul Boersma and David Weenink. Praat: doing phonetics by computer [computer program]. version 6.1.24, 2020. Available at `http://www.praat.org/`.

[3] Roger K. Moore. Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction. In Kristiina Jokinen and Graham Wilcock, editors, *Dialogues with Social Robots*, pages 281–291. Springer Singapore, 2017.

[4] Jonathan Peirce, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1):195–203, 2019.

[5] David P. Pisoni. Perception of synthetic speech. In Jan P. H. van Santen, Richard W. Sproat, Joseph P. Olive, and Julia Hirschberg, editors, *Progress in Speech Synthesis*, pages 541–560. Springer-Verlag, New York, 1997.

[6] David L. Strayer, Jonna Turrill, Joel M. Cooper, James R. Coleman, Nathan Medeiros-Ward, and Francesco Biondi. Assessing cognitive distraction in the automobile. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(8):1300–1324, 2015.

[7] Jürgen Trouvain and Bernd Möbius. Zu Mustern der Pausengestaltung in natürlicher und synthetischer Lesesprache. In André Berton, Udo Haiber, and Wolfgang Minker, editors, *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2018*, pages 334–341. TUDpress, Dresden, 2018.

Diese Arbeit ist Teil eines durch die **Vector Stiftung** geförderten Projekts `https://vector-stiftung.de` — PI: Daniel Duran

Weitere Informationen auf researchgate:
`https://www.researchgate.net/project/Der-Faktor-Mensch-in-der-Mensch-Maschine-Interaktion`