



Wen täuscht die KI?

Der Einfluss von Sprecher- und Hörergeschlecht auf die Erkennung synthetischer und menschlicher Stimmen

Mia Fabienne Babik, Marie Kenzler, Nele Knoppik, Catharina Lohmann, Hannah Naubert, Emma Wittig

Leitung: Dr. Christine Nussbaum



Theoretischer Hintergrund

- synthetische Stimmen werden zunehmend in Sprachassistenzsystemen und KI-basierten Anwendungen eingesetzt
- Fortschritte in der Sprachsynthese erschweren die Unterscheidung zwischen menschlichen und synthetischen Stimmen
- Merkmale wie Geschlecht können die Wahrnehmung und Natürlichkeit von Stimmen beeinflussen [1]
- bislang ist unklar, ob das Geschlecht von Sprechern und Hörern die Diskriminationsleistung menschlicher und synthetischer Stimmen beeinflusst

Methode

Stimuli

- 3 einfache deutsche Sätze:
„Der Fahrer lenkt den Wagen“, „Der Maler streicht die Wohnung“, „Die Eltern kaufen das Spielzeug“
(semantische Bedeutung war für die Aufgabe irrelevant)
- 3 Dimensionen: Art (menschlich vs. synthetisch)
Geschlecht (männlich vs. weiblich)
Alter (jung vs. alt)

Design

- *Experiment*: Teilnehmende hörten 120 zufällig präsentierte Sprachaufnahmen (aufgeteilt in 4 Blöcke à 30 Trials)
→ Entscheidung per Tastendruck nach jeder Aufnahme, ob die Stimme menschlich oder synthetisch war
- anschließend Fragebogen zur Nutzung von und Einstellung zu KI
- ANOVA - 2x2-Design:
abhängige Variable: Diskriminationsleistung (d')
unabhängige Variablen: Sprecher-geschlecht, Hörer-geschlecht

Stichprobe

- Deutsch als Muttersprache
- normales Hörvermögen

	männlich	weiblich
18 - 35 Jahre	21	16
60 - 80 Jahre	15	15

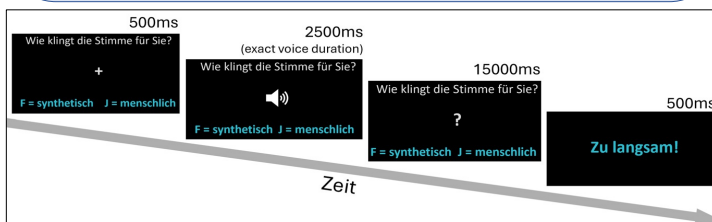


Abb. 1: Illustration der Prozedur eines Durchgangs

Ergebnisse

- **signifikanter Haupteffekt des Sprecher-geschlechts**: weibliche Stimmen wurden besser unterschieden als männliche Stimmen, ($F(1,67) = 12.63$, $p = .001$)
- höhere Diskriminationsleistung für weibliche Stimmen ($d' = 2.45$) als für männliche Stimmen ($d = 2.13$)
- kein Einfluss des Hörer-geschlechts auf die Diskriminationsleistung ($p = .562$)
- keine Interaktion zwischen Sprecher- und Hörer-geschlecht ($p = .673$)

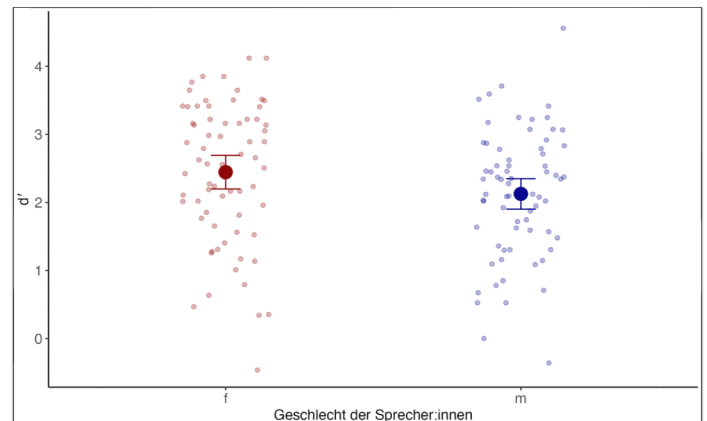


Abb. 2: Verteilung der d' -Werte nach Sprecher-geschlecht mit Mittelwerten und 95%-Konfidenzintervallen

Diskussion

Interpretation:

1. höhere Vertrautheit mit synthetischen weiblichen Stimmen durch häufige Nutzung in Sprachassistenten und Navigationssystemen
→ *zukünftige Forschung*: Untersuchung der individuellen Nutzung von Sprachassistenten als möglicher Moderator der Diskriminationsleistung
2. Unterschiede in Natürlichkeit und Qualität synthetischer Stimmen können auf variierende Trainingsdaten der Sprachsynthesysteme zurückzuführen sein
→ *zukünftige Forschung*: Ergänzung zusätzlicher abhängiger Variablen (z.B. Natürlichkeits- oder Vertrauenswürdigkeitstrainings) zur Bewertung der Stimulusqualität [1]

Limitationen:

- individuell per Prompt erzeugte synthetische Stimuli könnten Unterschiede in Natürlichkeit bzw. Qualität zwischen den Bedingungen verursacht und die Vergleichbarkeit beeinflusst haben
- online-Datenerhebung mit eingeschränkter Kontrolle der Testbedingungen
→ mögliche Erhöhung der Fehlervarianz