



# Auditive Szenenanalyse auf der Cocktailparty

Luise Bock, Merija Gühring, Luise Hillebrand, Alina Schönherr, Marco Schumacher, Vera Wolfertetter  
Leitung: PD Dr. Romi Zäske ([romi.zaeske@uni-jena.de](mailto:romi.zaeske@uni-jena.de))  
EmPra-Kongress: 15. Juli 2022

## Einleitung

Unser alltägliches Leben setzt sich aus einer Vielzahl von Geräuschen zusammen. In sog. Cocktailparty-Situationen sind mehrere Menschen in Gespräche aus unterschiedlichen Richtungen involviert, und gleichzeitig versucht man seinem eigenen Gesprächspartner zu folgen (Abb. 1). Auditive Szenenanalyse wurde bislang v.a. für Sprachverstehen untersucht<sup>1</sup>, selten jedoch für die Segregation von Sprecheridentitäten innerhalb sog. Multitalker-Stimuli. Dabei wurde gezeigt, dass es Hörern mit steigender Sprecheranzahl schwerer fällt, die Sprecher zu zählen und damit zu segregieren<sup>2,3</sup>. Wie können wir die Segregation von mehreren Sprechern in Cocktailparty-Situationen unterstützen, wenn eine zusätzliche aufgabenirrelevante Stimme im Hintergrund spricht? Ausgehend von Studien zum Sprachverstehen sollte eine Beeinflussung höherer Verarbeitungsprozesse bei der Segregation der Sprecher helfen<sup>1</sup>. Dazu zählt u.a. das Lenken des Aufmerksamkeitsfokus auf das relevante Sprachsignal (Multitalker) und das Entfernen aufgabenirrelevanter semantischer Information aus der Hintergrundsprache. Hierfür hinterlegten wir die aufgabenrelevanten Multitalker-Stimuli, welche einen Vokal im linken oder rechten Ohr sprachen, mit einer binauralen irrelevanten Stimme. Diese äußerte einen Satz. Manipuliert wurde die Sinnhaftigkeit des irrelevanten Satzes (vorwärts vs. rückwärts) und das Vorhandensein eines Cues, der den Ort des Multitalker-Stimulus ankündigt. Unsere Hypothesen sind in Abb. 2 schematisch dargestellt.

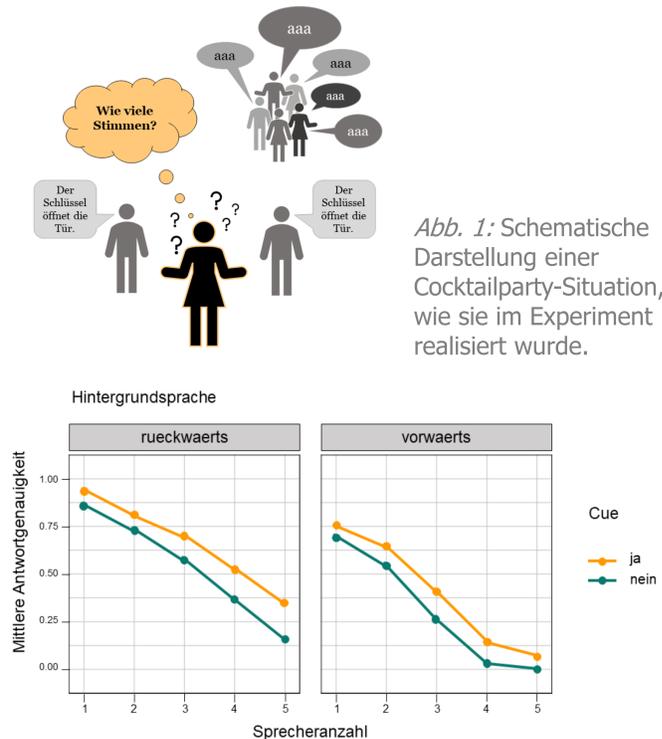


Abb. 1: Schematische Darstellung einer Cocktailparty-Situation, wie sie im Experiment realisiert wurde.

Abb. 2: Hypothesengeleitete, erwartete Ergebnisse: Mittlere Antwortgenauigkeiten in der Zählungsaufgabe in Abh. von experimentellen Bedingungen (Hintergrundsprache und Cueing), sowie getrennt nach Sprecheranzahl.

## Methode

### Stichprobe

- 23 Probanden (15 weiblich, 7 männlich, 1 binär; Alter:  $M = 22.1$   $SD = 3.13$ , Range: 19-31 Jahre)

### Stimuli

- Multitalker-Stimuli: 5 Gehaltene Vokale (1500 ms) von 31 Sprechern (14♀) aus der JESS-Datenbank<sup>4</sup>, monoaural
- Hintergrundsprecher: Sätze von 20 weiteren Sprechern (10♀) in 2 Versionen (vorwärts und rückwärts), binaural
- Normalisierung und Kombination der Multitalker- und Hintergrundsprecher zu 500 Stimuli (wav-Format, 16-bit, 44.1 kHz)

### Prozedur

- Online-Studie mit PsyToolkit<sup>5,6</sup> über Kopfhörer
- 1-5 gleichgeschlechtliche Sprecher wurden gleichzeitig monoaural (links/rechts) abgespielt (Multitalker)
- Aufgabe: Schätzung der Sprecheranzahl (1-5)
- Gleichzeitig sprach ein irrelevanter Sprecher (binaural)

2 x 2 Innersubjektdesign:

- Hintergrundsprache: vorwärts vs. rückwärts
- Cueing: Ort der aufgabenrelevanten Multitalker wird von einem Signalton (Cue) angekündigt oder nicht (2 Blöcke)
- 80 randomisierte Trials, siehe Abb. 3 für Trialablauf

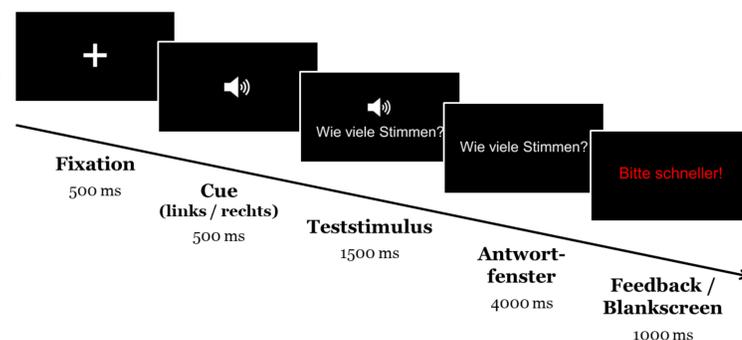


Abb. 3: Beispielhafter Ablauf für einen Trial in der Cueing-Bedingung mit Cue.

## Ergebnisse

### Mittlere Antwortgenauigkeiten (Abb. 4)

Zweifaktorielle ANOVA mit Messwiederholung auf beiden Faktoren

- Kein sign. HE von Hintergrundsprache  $F(1, 22) = 1.308, p = .265, \eta_p^2 = .06$
- Kein sign. HE von ortsbasiertem Cueing  $F(1, 22) < 1$
- Keine sign. Interaktion  $F(1, 22) < 1$

### Reaktionszeiten

- Keine sign. HE oder Interaktion  $0.08 \leq F(1, 22) \leq 1.84, .189 \leq p \leq .777$

### Deskriptiv

Abnahme der Genauigkeit mit steigender Sprecheranzahl

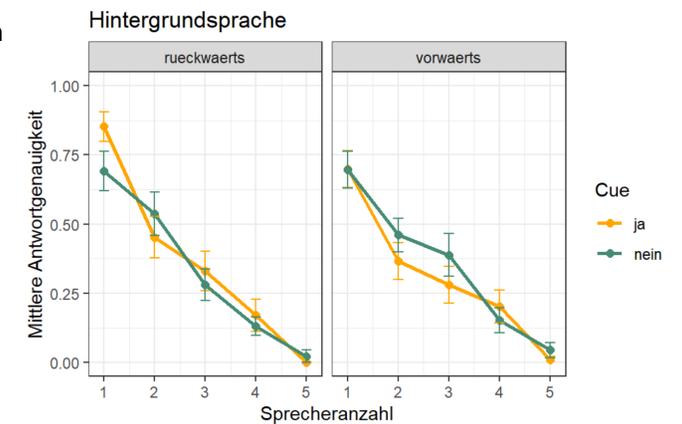


Abb. 4: Mittlere Antwortgenauigkeiten in der Zählungsaufgabe in Abh. von experimentellen Bedingungen (Hintergrundsprache und Cueing), sowie getrennt nach Sprecheranzahl. Fehlerbalken sind Standardfehler des Mittelwerts (SEM).

## Diskussion

Entgegen unserer Hypothesen und der Befunde zur Sprachverarbeitung<sup>1</sup> wurde die Segregierbarkeit von Multitalker-Stimuli weder durch die Sinnhaftigkeit irrelevanter Hintergrundsprache noch durch ortsbasiertes Cueing moduliert.

### Mögliche Gründe

- Das Segregieren von Sprechern anhand ihrer Klangfarbe findet im Ggs. zu linguistischer Verarbeitung auf relativ frühen Verarbeitungsstufen statt. Eine Interferenz durch Hintergrundsprache fand mglw. nicht statt, weil die Zählungsaufgabe keiner Sprachverarbeitung bedarf. Überprüfbar wäre dies durch Änderung der Aufgabe (z.B. Klassifikation von Vokalen).
- Ausbleiben eines Cueing-Effekts könnte an generell hoher Aufgabenschwierigkeit liegen, die auch durch einen Orts-Cue nicht erleichtert wird. Die Aufgabe könnte erleichtert werden durch geschlechtsheterogene Multitalker-Stimuli bzw. komplexere Sätze statt gehaltener Vokale<sup>7</sup>.
- Evtl. kam es zu Verfälschung durch übliche Störvariablen in Online-Studien, weshalb wir eine Replikation unter kontrollierten Laborbedingungen empfehlen.
- Geringe Power unserer relativ kleinen Stichprobe könnte mit Erhöhung der Probandenzahl begegnet werden<sup>8</sup>.

### Ausblick

Insgesamt ist das Zählparadigma ein nützliches Instrument zur sprachfreien Untersuchung von Fähigkeiten der auditiven Szenenanalyse und erscheint damit geeignet für den Vergleich von Gruppen mit unterschiedlichem Sprachhintergrund oder verschiedenen Sprachfähigkeiten.

## Literatur

- Wang, X. & Xu, L. (2021). Speech perception in noise: Masking and unmasking. *Journal of Otology*, 16, 109-119.
- Bock, L., Gühring, M., Hillebrand, L., Schönherr, A., Schumacher, M., Wolfertetter, V. (2022). Forschungsbericht zum Empiriepraktikum „Auditive Szenenanalyse“ im WiSe. Lehrstuhl für Allgemeine Psychologie 1, Friedrich-Schiller-Universität Jena.
- Kwak, C., Han, W. (2020). Towards Size of Scene in Auditory Scene Analysis: A Systematic Review. *Journal of Audiology & Otology*, 24(1), 1-9.
- Zäske, R., Skuk, V. G., Golle, J., & Schweinberger, S. R. (2020). The Jena Speaker Set (JESS) - A database of voice stimuli from unfamiliar young and old adult speakers. *Behavior Research Methods*, 52(3), 990-1007.
- Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4), 1096-1104.
- Stoet, G. (2010). PsyToolkit - A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24-31.
- Kawashima, T., & Sato, T. (2015). Perceptual limits in a simulated "Cocktail party". *Attention Perception & Psychophysics*, 77(6), 2108-2120. doi:10.3758/s13414-015-0910-9
- Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.

Besonderer Dank gilt Anna Knosp und Manuel Pöhlmann, die uns bei der Programmierung der Experimente unterstützend zur Seite standen.