

What the user's voice tells us about UX

- Analysing parameters of the voice as indicators of the User Experience of the usage of intelligent voice assistants

David Obremski, Carolin Wienrich, Astrid Carolus

Motivation

- Steigende Beliebtheit von Smart Speakern¹
- User Experience von Smart Speakern wichtig für Entwicklung
- Meist erhoben durch explizite Maße (Fragebögen)
 - Unterbricht die Nutzung, Bias durch z.B. soziale Erwünschtheit



¹Intelligente Lautsprecher - Absatz weltweit bis Q4 2019. <https://de.statista.com/statistik/daten/studie/818982/umfrage/absatz-von-intelligenten-lautsprechern-weltweit-pro-quartal/Library>

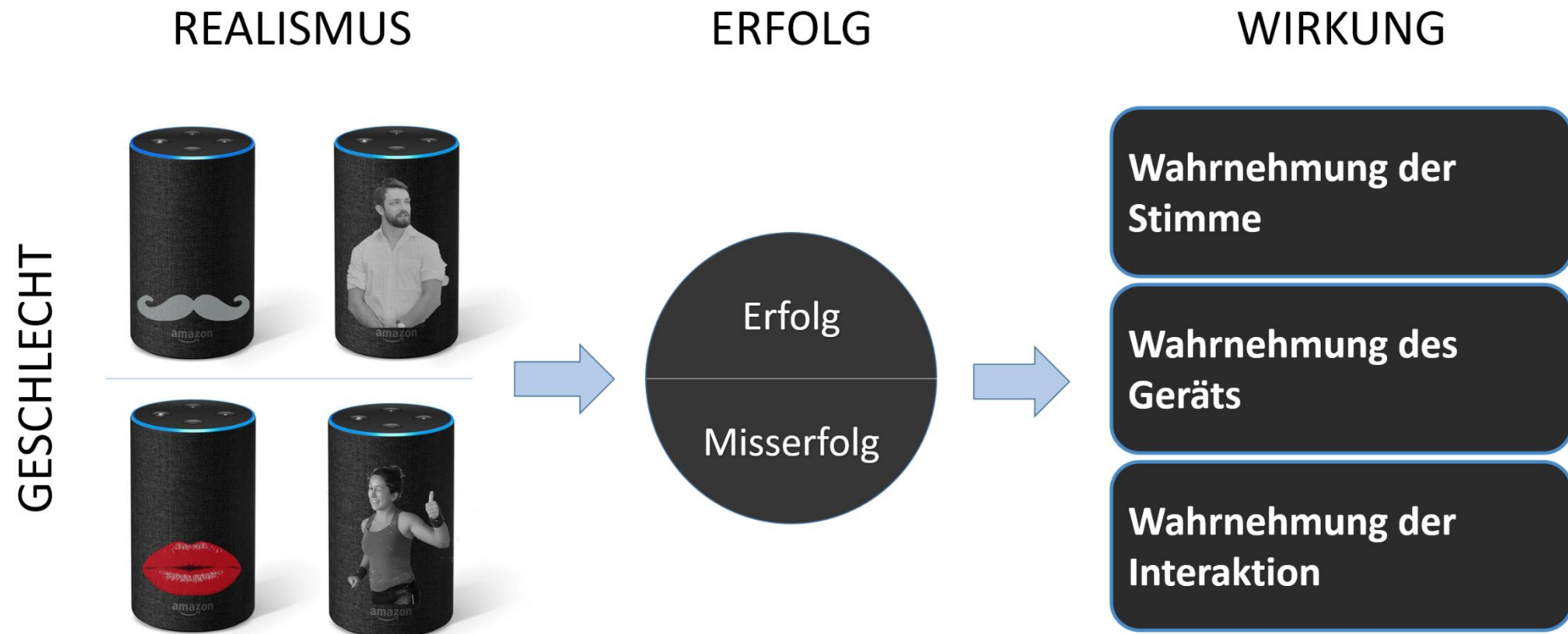
Motivation

- Tonhöhe und Lautstärke der Sprache des Users kann Aufschluss über Emotionen geben¹
- Kann während der Nutzung erhoben werden
- Keine Sentiment- oder semantische Analyse, sondern die Analyse der akustischen Daten

¹Rainer Banse and Klaus R Scherer. 1996. Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology* 70, 3 (1996), 614.



Rahmenbedingungen der Studie



Probleme/Anforderungen

- Auswertung muss automatisiert werden
- Bei nicht intrusiver Aufzeichnung der Stimme:
 - Sowohl User- als auch Gerätsprache auf der Aufnahme
 - Trennung der beiden Signale erforderlich
- Baseline Aufnahme erforderlich zur Ermittlung der relativen Veränderung der Sprache

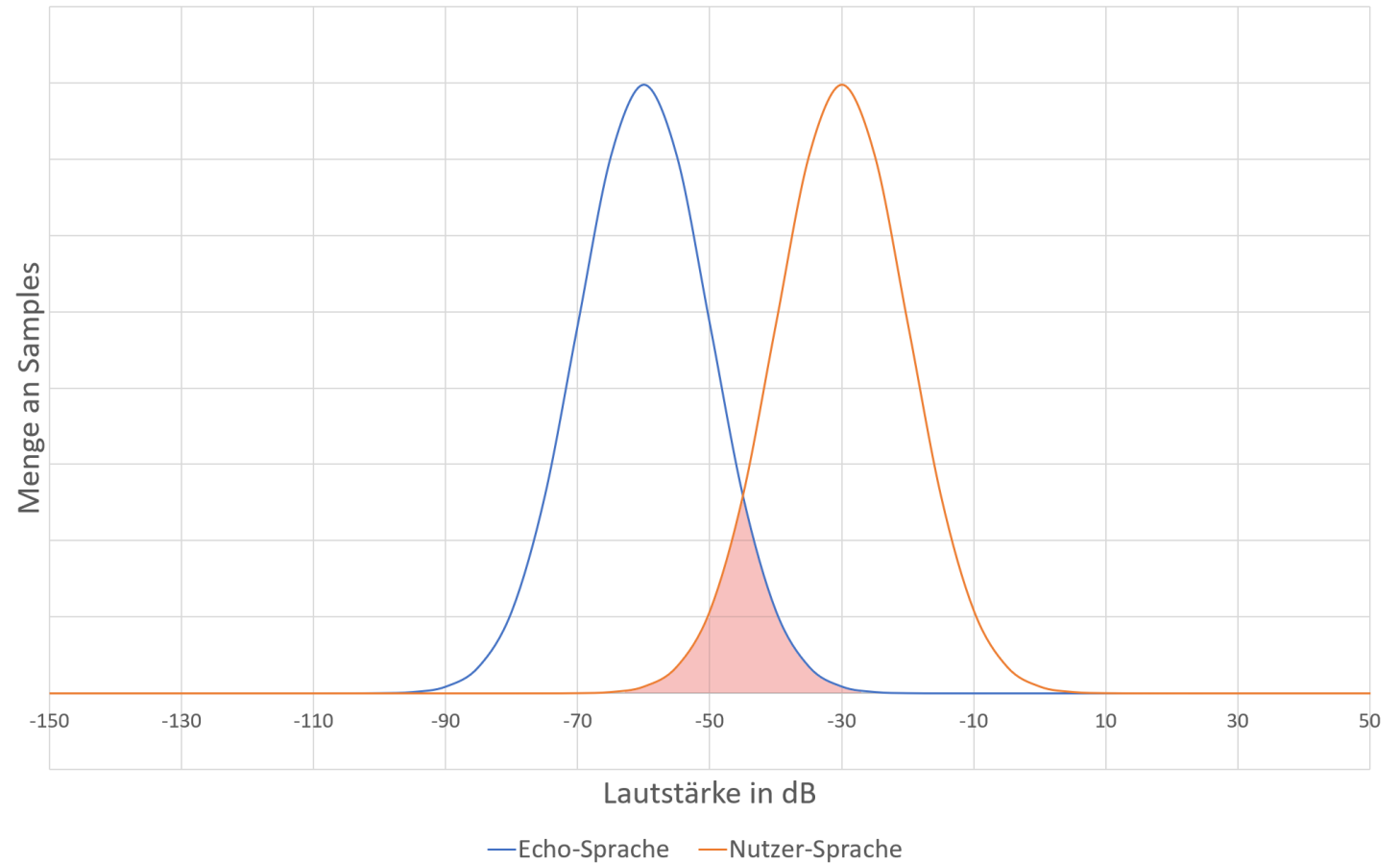
Entwicklung des Tools

- Java Application
- Einbindung des Tarsos DSP Frameworks¹ zur akustischen Analyse
- Trennung der beiden Signale an bestimmten Filterwerten
- Analyse der Baseline-Aufnahme und der Experimental-Aufnahme
- Ermittlung der relativen Änderung von Lautstärke und Tonhöhe

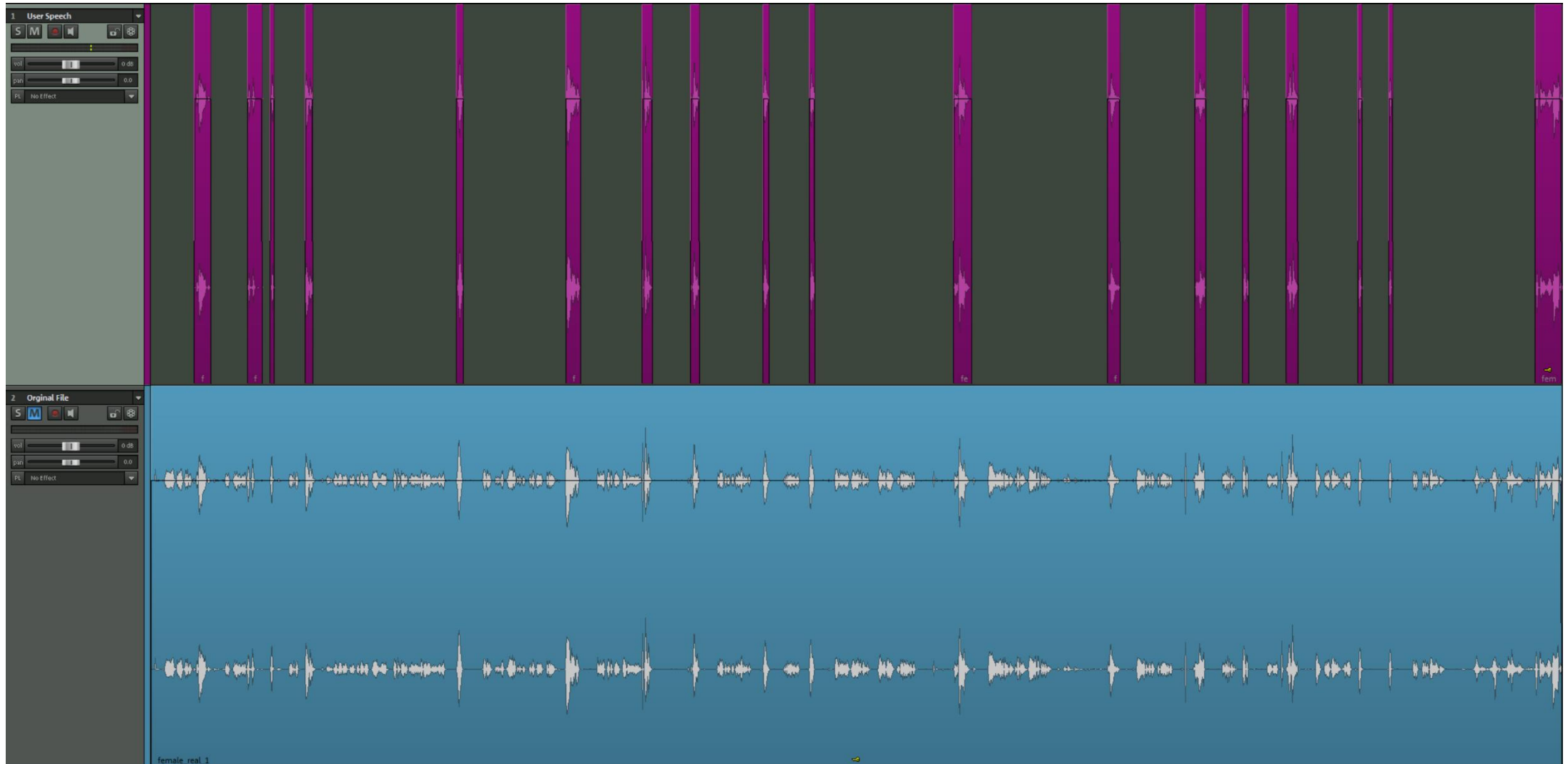
¹Joren Six, Olmo Cornelis, and Marc Leman. 2014. Tarsos DSP, a Real-Time Audio Processing Framework in Java. In Proceedings of the 53rd AES Conference (AES53rd).

Auswahl der Filterschwellenwerte

- Analyse verschiedener Gerätestimmen hinsichtlich Lautstärke
- Ermittlung der passenden Filterschwellenwerte ($M_{\text{Lautstärke}} + 1 \text{ SD}$)
- Wenn Lautstärke länger als $2s < \text{Filterschwellenwert}$ = Gerätsprache
- Sonst Usersprache

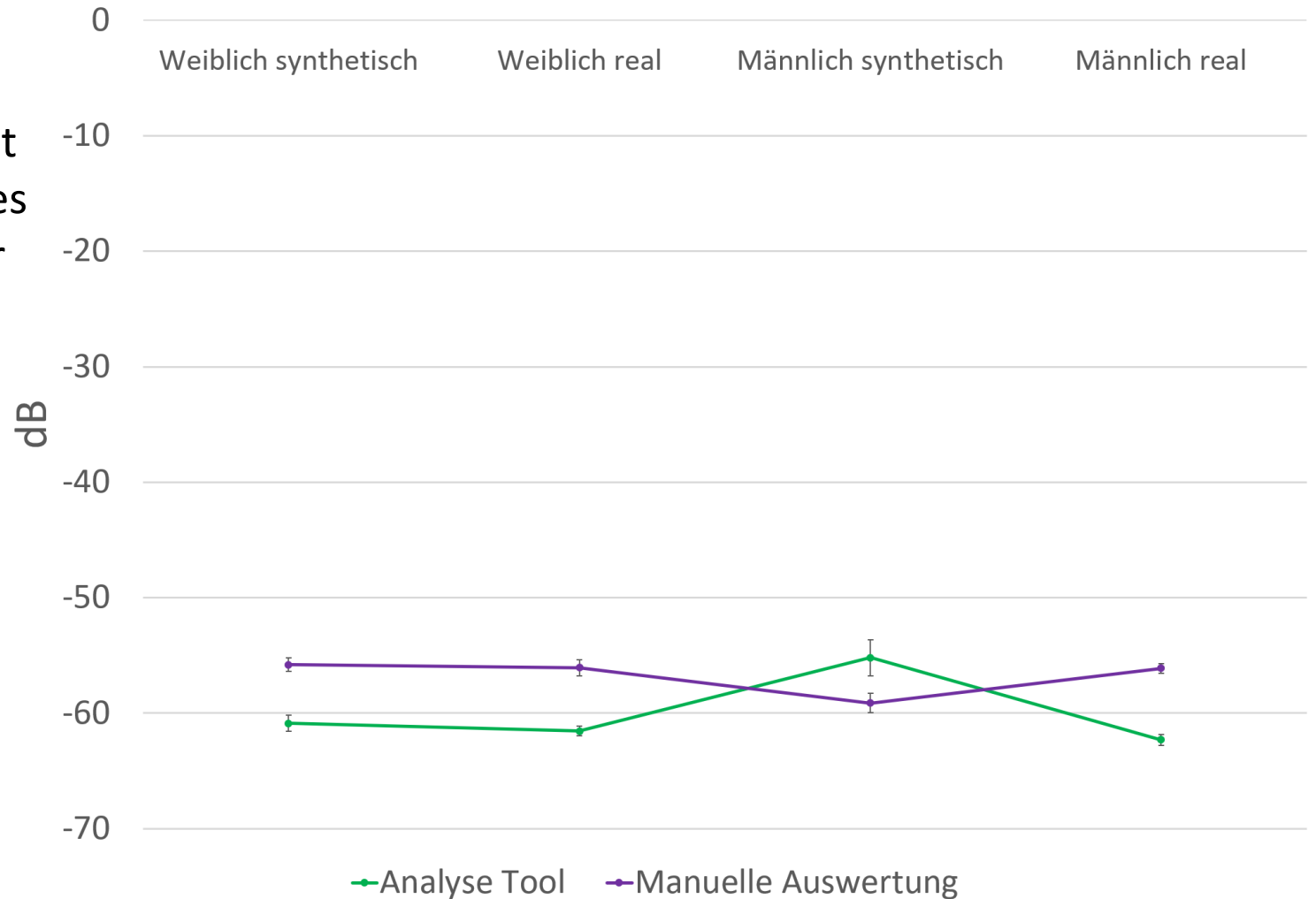


1. Validierung



1. Validierung (Lautstärke)

- Vergleich von $M_{\text{Lautstärke}}$ der händisch geschnittenen Files mit den vom Tool geschnittenen Files
- Analyse Tool immer leiser außer bei männlich synthetisch
 - Analyse Tool wertet auch Gerätesprache aus



1. Validierung (Tonhöhe)

- Vergleich von $M_{\text{Tonhöhe}}$ der händisch geschnittenen Files mit den vom Tool geschnittenen Files
- Tarsos DSP kann nicht zwischen weiblicher und männlicher Stimme trennen
 - Pitch Analyse unbrauchbar

Condition	Pitch
Female synthetic	M=98.72, SD=2.12
Female real	M=128.88, SD=16.12
Male synthetic	M=128.23, SD=160.93
Male real	M=115.41, SD=35.68

2. Validierung

- Laborstudie mit N=121 Probanden
- Between Subjects Design
- **UV:** Interaktion mit Smart Speaker **erfolgreich** vs. **nicht erfolgreich**
- **AVs:**
 - **Implizit:** $M_{\text{lautstärke}}$
 - **Explizit:** QUESI, PANAS

2. Validierung Ergebnisse

- Explizit:
 - Der Gesamt-QUEST-Wert war in der Erfolgs-Bedingung **signifikant höher** als in der Misserfolgs-Bedingung
 - Der positive Affekt war in der Erfolgs-Bedingung **signifikant höher** als in der Misserfolgs-Bedingung
- Implizit:
 - Kein signifikanter Unterschied von $M_{\text{Lautstärke}}$ in den beiden Bedingungen
 - Trend zeigt aber: Höhere Lautstärke in der Erfolgsbedingung

Limitierungen & zukünftige Arbeit

- Tarsos DSP Framework nicht geeignet für Tonhöhenanalyse der Sprache
- Filterschwellenwerte sollten genauer ausgewählt werden
- Eventuell Machine Learning basierte Verfahren zur Signaltrennung

