

Der Headphone and Loudspeaker Test [HALT]: Vorschläge für die kontrollierte Verwendung von Wiedergabegeräten in Internet-Experimenten

Yves Wycisk¹, Kilian Sander¹, Reinhard Kopiez¹, Benedetto Manca², Friedrich Platz³

¹Hanover University of Music, Drama and Media, Hanover, Germany

²University of Cagliari, Cagliari, Italy

³State University of Music and Performing Arts, Stuttgart, Germany



1 Hintergrund

- In Internet-Experimenten zur Hörwahrnehmung können Wiedergabegeräte als Störvariable betrachtet werden.
- Wiedergabelautstärke (beeinflusst Gefühl „self-motion“; Todd & Cody, 2000)
- Kopfhörerwiedergabe (höhere mittlere Geschwindigkeit der Kopf- und Körperbewegung; Zalechowska et al., 2020).
- Ein umfassendes Verfahren zur Kontrolle existiert nach unserem Kenntnisstand nicht.

2 Ziele

- Entwicklung eines zuverlässigen, objektiven und effizienten Verfahrens (HALT) zur Erfassung der Eigenschaften von Wiedergabegeräten und Abhörbedingungen in Internetexperimenten.
- Ermittlung der Prävalenz für Kopfhörernutzung, um bei der Anwendung von Screening-Verfahren prädiktive Werte berechnen zu können.

3 Methodik

1. Schritt: Labor-Studie

Akquise

- Mailing-Listen, Facebook, Poster

Probanden

N = 40, weiblich: 25, männlich: 15
Alter: M = 31.83, SD = 13.48

Basierend auf rosa Rauschen und Sinustönen wurden Höraufgaben konstruiert, um Wiedergabeeigenschaften objektiv zu erfassen.

Stimuli und Höraufgaben

Lautstärke-Einstellung: laute und leise Rauschabschnitte zählen

Stereo/Mono: alternierende Rauschabschnitte auf einer Lautsprecher-/Kopfhörerseite zählen

Untere Grenzfrequenz: Sinustöne der Frequenzen 20 Hz, 60 Hz, 100 Hz und 140 Hz in separaten Aufgaben zählen

2. Schritt: Internet-Studie

Akquise

- Panel: mo'web GmbH

Probanden

N = 211, weiblich: 117, männlich: 94
Alter: M = 42.4, SD = 11.35

In der Internet-Studie wurden Daten zur Beurteilung der Screening-Tests gesammelt.

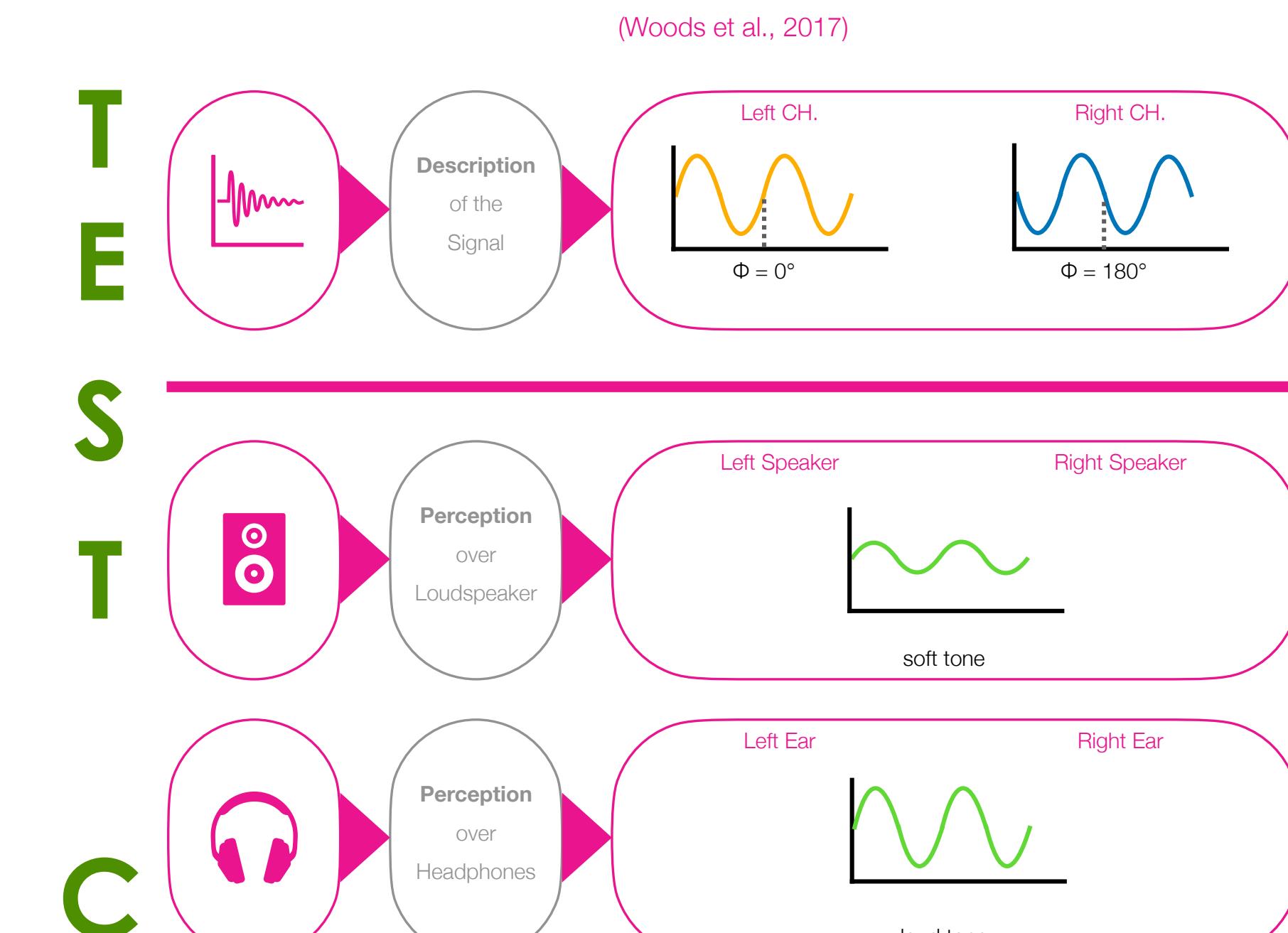
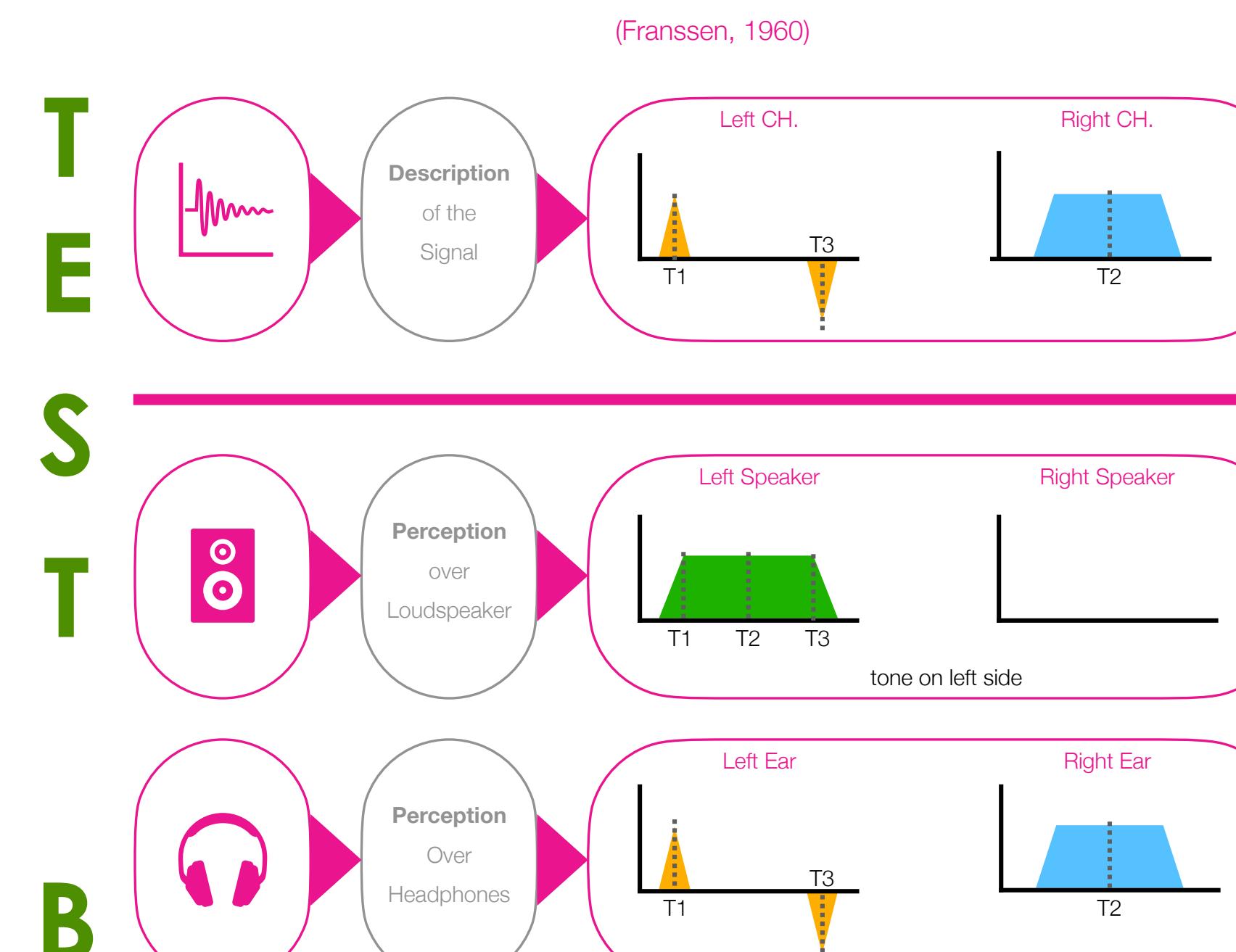
Stimuli und Höraufgaben

Jeder der Tests A, B und C besteht aus sechs Items. Dadurch ergeben sich für jeden Test null bis sechs mögliche korrekte Antworten.

Test A: basiert auf interauralen Zeitdifferenzen

Test B: basiert auf dem Franssen-Effekt

Test C: basiert auf destruktiven Interferenzen



4 Ergebnisse

Lautstärke-Einstellung

Wenn die Lautstärke mit der **traditionellen Aufforderung** „Bitte stellen Sie eine angenehme Wiedergabelautstärke ein“ von den Probanden eingestellt wurde, ergaben sich die folgenden Werte für den Schalldruckpegel:

Level: M = 61.99 dBa (min = 42.30 dBa, max = 82.20 dBa, **SD = 8.65**)

Wenn die Aufgabe zur Lautstärke-Einstellung aus der **HALT-Prozedur** genutzt wurde, ergaben sich die folgenden Werte für den Schalldruckpegel:

Level: M = 67.77 dBa (min = 59.50 dBa, max = 82.60 dBa, **SD = 4.29**)

Reliabilität: N = 160, $r_{tt} = .899$, 95% CI [.862, .924], BF+0 = 1.458e+55

Ein Levene's-Test zeigte, dass die **HALT-Prozedur** die Heterogenität der Lautstärke-Einstellungen im Vergleich zur **traditionellen Aufforderung** signifikant reduzierte. Die Test-Retest-Reliabilität der Aufgabe kann mit .899 als hoch bezeichnet werden.

Stereo/Mono

Genauigkeit: 95.6% korrekte Detektionen

Reliabilität: N = 160, $r_{tt} = .792$, 95% CI [.722, .842], BF+0 = 5.704e+32

Die Test-Retest-Reliabilität der Aufgabe kann mit .792 als hoch bezeichnet werden.

Untere Grenzfrequenz

Es werden Frequenzen erkannt, die 40 dB unter dem spektralen Median des Wiedergabegeräts liegen.

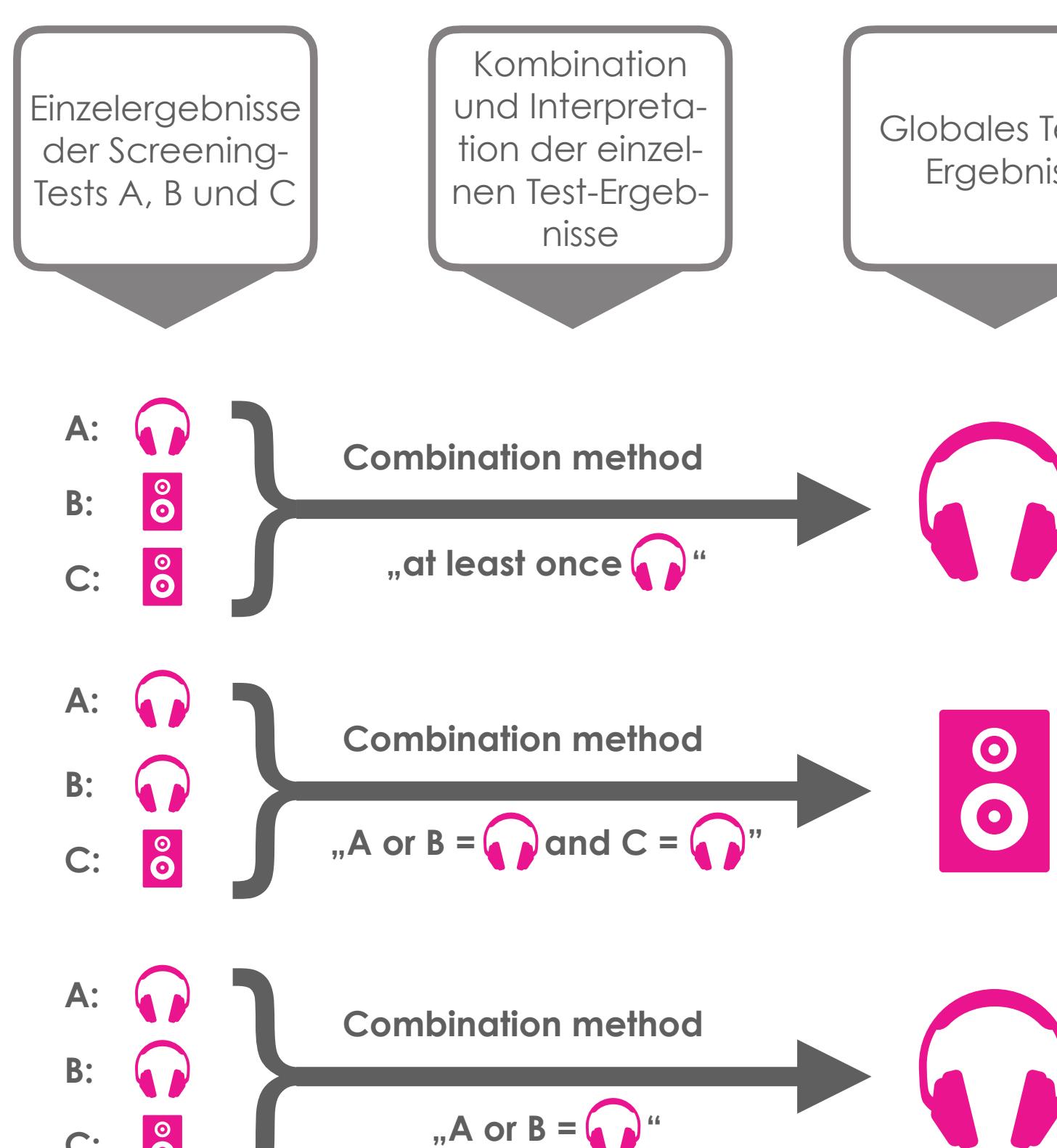
Reliabilität: N = 640, $r_{tt} = .821$, 95% CI [.793, .844], BF+0 = 5.524e+153

Die Test-Retest-Reliabilität der Aufgabe kann mit .821 als hoch bezeichnet werden.

Test	Sensitivität für 5 von 6 korrekte Antworten (95% CI)	Spezifität für 5 von 6 korrekte Antworten (95% CI)	AUC
A	80.0% (.700, .873)	68.7% (.603, .760)	.768
B	80.0% (.700, .873)	83.2% (.759, .886)	.844
C	92.5% (.846, .965)	58.0% (.494, .661)	.807

Kombination von Screening-Tests

Die Ergebnisse der Tests A, B und C können einzeln genutzt werden. Durch die Kombination der Tests sind höhere Erkennungsleistungen möglich als bei der Nutzung eines einzelnen Tests. Drei verschiedene Beispiele für die Kombination der Screening-Tests und die daraus resultierenden globalen Test-Ergebnisse:



Für verschiedene Mindestanzahlen richtiger Antworten (Schwellenwert) und 18 verschiedene Kombinations-Methoden wurden die Sensitivität und Spezifität geschätzt.

Für eine gegebene Prävalenz für Kopfhörernutzer kann die **overall utility** (Treat & Viken, 2012) berechnet werden, um eine optimale Testkombination und die Schwellenwerte der einzelnen Tests auszuwählen.

Prävalenz

In der vertrauenswürdigen **ungefilterten** Stichprobe (N = 1.194) benutzten n = 211 Teilnehmer Kopfhörer. Dies entspricht einer Prävalenz von 17,67%, 95% CI [15,6%, 19,9%].

Schätzung der Stichprobengröße

Die Anzahl der wahren Kopfhörernutzer in einer Stichprobe von Teilnehmern mit einem Kopfhörertestergebnis kann als binomialverteilte Zufallsvariable konzeptualisiert werden. Die Wahrscheinlichkeit, dass ein Teilnehmer angesichts seines Kopfhörertestergebnisses Kopfhörer benutzt hat, berechnet sich aus der Prävalenz und der Genauigkeit (Sensitivität und Spezifität) eines Tests bzw. einer Test-Kombination. Für eine Wahrscheinlichkeit einer Mindestanzahl an wahren Kopfhörernutzern – ähnlich der Testpower – kann mithilfe einer Normalapproximation der Binomialverteilung die benötigte Stichprobengröße bestimmt werden.

Online Tool

- Bestimmung von Testkombinationen und Schwellenwerten im Hinblick auf die Schätzung des Stichprobenumfangs.
- Konfiguration des Kopfhörer- und Lautsprechertests [HALT] (R-Paket).

HALT-Demo:



Online Tool:



HALT-R-Paket:



5 Fazit

Im Gegensatz zu existierenden Tests zur Kontrolle von Wiedergabegeräten (Woods et al. 2020), ist der HALT ein umfassendes Verfahren zur Überprüfung der Wiedergabe. In Übereinstimmung mit den Standards der Epidemiologie sollte die ermittelte Kopfhörer-Prävalenz dafür genutzt werden die Datenqualität nach einem Kopfhörer-Screening einzuschätzen. Der HALT kann einen nützlichen Beitrag dazu leisten, um Laborexperimente ins Internet zu verlagern, während die Kontrolle über Störvariablen erhalten bleibt.

Referenzen

- Ahrens, W., & Pigeot, I. (Eds.). (2014). *Handbook of epidemiology* (Second edition). Springer Reference.
- Bilsen, F., & Raatgever, J. (2002). *Demonstrations of dichotic pitch* [CD].
- Franssen, N. V. (1960). *Some considerations on the mechanism of directional hearing* [Doctoral dissertation]. Technische Hogeschool.
- Todd, N. P. M., & Cody, F. W. (2000). Vestibular responses to loud dance music: A physiological basis of the “rock and roll threshold”? *The Journal of the Acoustical Society of America*, 107(1), 496–500. <https://doi.org/10.1121/1.428317>
- Treat, T. A., & Viken, R. J. (2012). Measuring test performance with signal detection theory techniques. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology* (Vol 1: Foundations, planning, measures, and psychometrics, pp. 723–744). American Psychological Association. <https://doi.org/10.1037/13619-038>
- Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, 79(7). <https://doi.org/10.3758/s13414-017-1361-2>
- Zalechowska, A., Gonzalez-Sanchez, V. E., Laeng, B., & Jensenius, A. R. (2020). Headphones or speakers? An exploratory study of their effects on spontaneous body movement to rhythmic music. *Frontiers in Psychology*, 11, 698. <https://doi.org/10.3389/fpsyg.2020.00698>