Headphone and Loudspeaker Screening for Web-Based Auditory Experiments: Suggestions for a Reliable Estimation of Data Quality and Sample Size

Kilian Sander[1], Yves Wycisk[1], Reinhard Kopiez[1], Benedetto Manca[2], Friedrich Platz[3]

[1]*Hanover University of Music, Drama and Media, Hanover, Germany*
[2]*University of Cagliari, Cagliari, Italy*
[3]*State University of Music and Performing Arts, Stuttgart, Germany*

ICMPC ESCOM 2021
Connectivity and diversity in music cognition.

HANOVER MUSIC LAB

hmtmh Hochschule für Musik Theater und Medien Hannover
HANOVER UNIVERSITY OF MUSIC, DRAMA AND MEDIA

1 Background

- Suggestions for screening methods already exist (Woods et al., 2017) ➔ Test C in Methods.
- Their practical application does not consider the prevalence of playback devices.
- The proportion of headphones to loudspeakers in web-based experiments seems to be unknown.
- In the current state of research, the assessment of correctly identified playback devices (data quality) based on screening methods is unclear.

2 Aims

- To develop a reliable screening method for detecting headphones and loudspeakers as playback devices.
- To determine the screening method's metrics sensitivity (ability to correctly identify headphone users) and specificity (ability to correctly identify loudspeaker users) analogous to epidemiology.
- To provide an online tool which calculates application-oriented data quality and the required sample size for web-based surveys, that also considers both the prevalence of playback devices and the test procedure metrics.

3 Method

Pre-Study

Acquisition
- mailing lists, Facebook, Posters

Participants
N = 40, female: 25, male: 15
Age: *M* = 31.83, *SD* = 13.48

pre-study - laboratory - 4 items per screening test:
General viability of Test A and B was examined

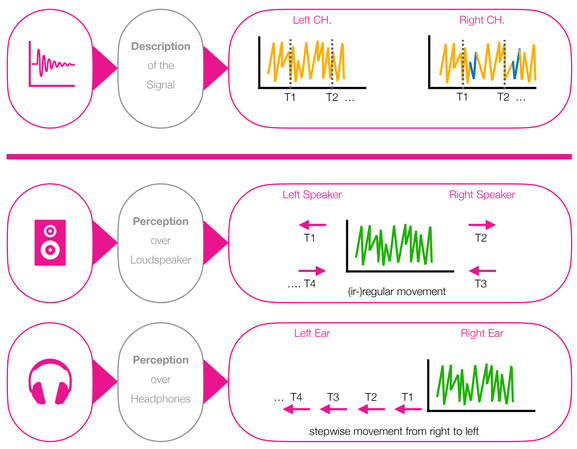
main study - online - 6 items per test:
Three of the stimuli used for Test A and B are exactly as described in the respective figures below whereas the channels are reversed for the remaining three. 0-6 correct answers for each screening test (A, B, and C).

Main Study

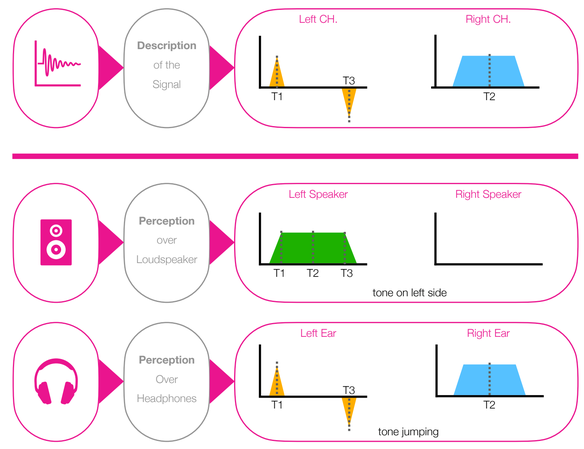
Acquisition
- panel by *mo'web GmbH*

Participants
N = 211, female: 117, male: 94
Age: *M* = 42.4, *SD* = 11.35

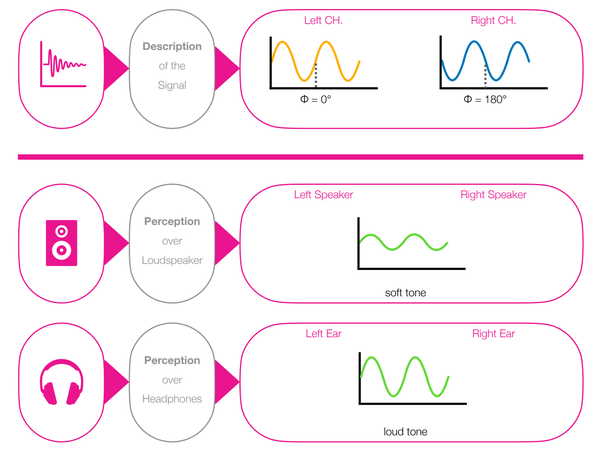
Test A (Interaural Time Difference)
(Bilsen & Raatgever, 2002)



Test B (Franssen Effect)
(Franssen, 1960)



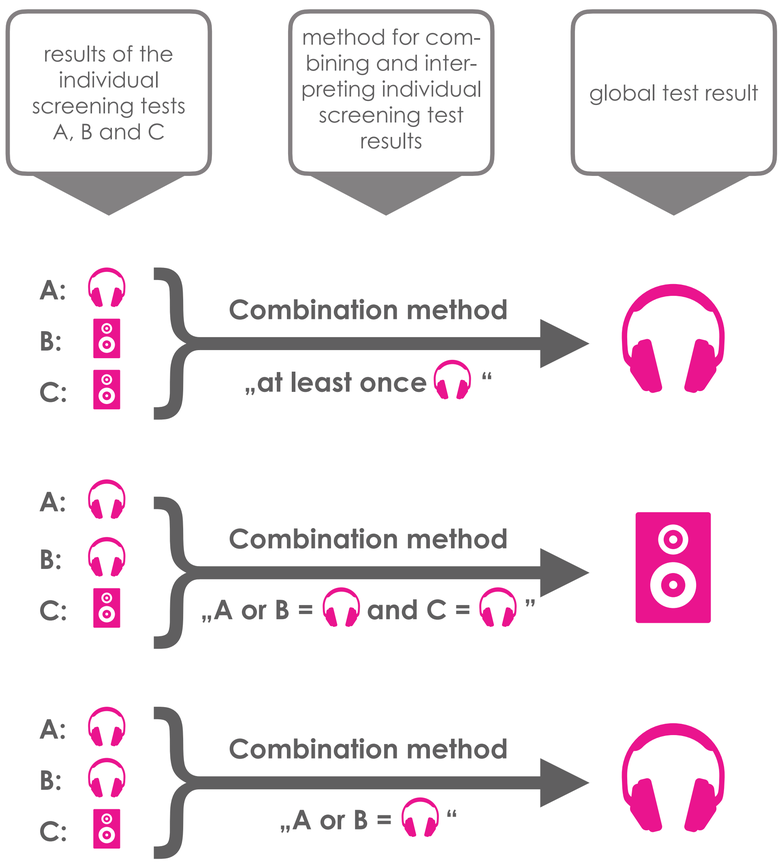
Test C (Destructive Interference)
(Woods et al., 2017)



4 Results

Combination of Screening Tests

Three different examples for the combination of the screening tests and their corresponding global results:



For different minimum numbers of correct answers (threshold), and 18 different methods, the sensitivity and specificity were estimated.
For a given base rate/prevalence for headphone users the overall utility (Treat & Viken, 2012) can be calculated to choose the optimal test combination and its thresholds.

Prevalence

In the trusted **unfiltered** sample (*N* = 1,194) *n* = 211 participants used headphones. This corresponds to a prevalence A of 17.67%, 95% CI [15.6%, 19.9%].
In the trusted **filtered** sample (no smartphones, tablets, monitors/ TVs; *N* = 211) *n* = 80 participants used headphones. This corresponds to a prevalence B of 37.92%, 95% CI [31.6%, 44.6%].
Prevalence A reflects the headphone usage in the studied population whereas prevalence B applies to a subset.

Sample Size Estimation

Model: The number of true headphone users *H* in a sample of *n* participants with a headphone test result can be conceptualized as random variable following a Binomial distribution. *p* is the probability that a participant used headphones given their headphone test result calculated from the prevalence and the test's sensitivity and specificity. ϑ is the probability of at least *k* true headphone users in the sample.

$$\vartheta := \mathbb{P}(H \geq k) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

Approximation: According to the De Moivre-Laplace theorem a Binomial distribution can be approximated by a Normal distribution (with Φ denoting the cumulative distribution function of the standard normal distribution). With a continuity correction we yield:

$$\vartheta \approx 1 - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

Calculating *n*: From the approximation follows for $\vartheta > 0.5$

$$n \approx -\frac{a}{2} + \sqrt{\left(\frac{a}{2}\right)^2 - b}$$

with

$$a = -\frac{1}{p}\left(2k - 1 + (1-p)\left(\Phi^{-1}(1-\vartheta)\right)^2\right)$$

$$b = \left(\frac{k - \frac{1}{2}}{p}\right)^2$$

Online Tool

- Determination of test combinations and thresholds in regard to the sample size estimation
- onfiguration of the Headphone And Loudspeaker Test [HALT] (R package)

QR code to Online Tool (click for link):

QR code to HALT R package (click for link):

QR code to HALT demo (click for link):

5 Conclusion

The low prevalence of headphone usage in web-based experiments indicates the central role of highly sensitive and specific screening methods. Considering the standards of signal detection theory (Treat & Viken, 2012) and epidemiology (Ahrens & Pigeot, 2014), it is insufficient to focus solely on sensitivity and specificity without obtaining information on device prevalence. Our findings can contribute to improve the data quality and efficiency of future online studies.

References

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.
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