Bringing Confounding Variables to a HALT: Suggestions for the Controlled Use of Playback Devices in Internet Experiments

Yves Wycisk (1), Reinhard Kopiez (1), Kilian Sander (1), Jakob Bergner (2), Stephan Preihs (2), Jürgen Peissig (2), Friedrich Platz (3)

(1) Hanover University of Music, Drama and Media, Hanover, Germany
(2) Leibniz University, Hanover, Germany
(3) State University of Music and Performing Arts, Stuttgart, Germany

Background

In internet experiments on auditory perception, playback devices may be a confounding variable that are difficult to control. Responses of participants may be influenced by the following factors:

- playback volume (which can influence the pleasurable sensation of self-motion, Todd & Cody, 2000),
- headphone playback (which can cause higher mean velocities of head and body movements, Zelechowska et al., 2020).

Although promising approaches for playback device screening tests exist (Milne et al., 2020), a comprehensive test procedure to remotely control characteristics of sound transducers is currently non-existent.

Aims

Our main research aim was to develop a reliable, objective, and efficient *Headphone and Loudspeaker Test* (HALT) to remotely control loudness adjustments, stereo/mono playback, headphone/loudspeaker playback, and cut-off frequencies. Additionally, we wanted to provide an online tool to make the setup and use of the test as easy as possible.

Methods

As a main paradigm, counting tasks for acoustical events were used to set up a HALT procedure. In a laboratory study (N = 40), the procedure was evaluated using four different playback devices: two kinds of headphones (circumaural/intra-aural), loudspeakers, and a laptop (built-in speakers). Acoustical features of the devices were measured. After each loudness adjustment, participant-selected dBFS values were documented and later transformed into dBSPL values. In an internet-based study (N = 211), HALT was validated

using various control procedures (e.g., cover story, timeout, and checking participants' attention).

Results

Based on a music example, participants (N = 35, non-impaired hearing) adjusted an average sound pressure level (SPL) of 61.99 dBA (*min* = 42.30 dBA, *max* = 82.20 dBA, *SD* = 8.65) as the preferred loudness level. In contrast, if loudness was adjusted by a HALT inherent loop procedure based on a counting task of noise events, the selected average SPL increased slightly to 67.77 dBA (*min* = 59.50 dBA, *max* = 82.600 dBA, *SD* = 4.29). Regarding the A-weighted dBSPL, the coefficient of variation (v) for the music stimulus was higher (v_C = 0.140) when compared to the HALT condition (v_E = 0.063), indicating a higher degree of unintended variability when music was used for the loudness adjustment. To check for test-retest reliability, we conducted a Bayesian correlation analysis with uninformed priors (N = 160 responses of 40 participants in four conditions). As an outcome, the loop procedure can be considered highly reliable (r_{tt} = .899, 95% CI [.862, .924], BF_{+0} = 1.458e+55).

Additionally, HALT correctly detected stereo playback in 95.6% of the cases. Again, a Bayesian Pearson correlation showed a high test-retest reliability (r_{tt} = .792, 95% CI [.722, .842], BF_{+0} = 5.704e+32).

A headphone screening based on interaural time differences achieved the highest test parameters: sensitivity = 80% (95% CI [70.0%, 87.3%]); specificity = 83.2% (95 % CI [75.9%, 88.6%]).

Conclusion and Implications

Although previous test procedures for the control of playback devices have been suggested, HALT is the first approach to a comprehensive procedure for remotely controlling playback devices and listening conditions. Both the sensitivity and specificity of our developed headphone screening can be further increased in the future by combining the test with other procedures (e.g., Milne et al., 2020). As the prevalence of playback devices should be taken into account to select a suitable test to check for headphone playback, we have provided an online calculator (http://testing.musikpsychologie.de/HALTConfig/) to simplify the selection and combination of different screening methods and to estimate the data quality after using a screening. Additionally, the calculator is an easy-to-use method to set up HALT for use with *psychTestR* (Harrison, 2020). We suggest that HALT is a useful tool when relocating laboratory experiments to the internet as it enables researchers to maintain control over confounding variables.

References

Harrison, P. (2020). psychTestR: An R package for designing and conducting behavioural psychological experiments. *Journal of Open Source Software*, *5*(49), 2088. https://doi.org/10.21105/joss.02088

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2020). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01514-0

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

Zelechowska, 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