Reproducibility of an ABX-Listening Test using Electronic Documentation and Automation: A case study

Max Schröder, Wiebke Hahne, Frank Krüger, and Sascha Spors
Institute of Communications Engineering, University of Rostock, Germany
E-Mail: {max.schroeder, wiebke.hahne, frank.krueger, sascha.spors}@uni-rostock.de

1 Introduction

Psychoacoustic experiments investigate the relation between a physical sound and its perception by humans. However, the experimental details are often insufficiently documented in scientific publications [1]. Reproducing such investigations thus, often requires much effort due to missing details in the documentation of prerequisites and experimental set-up as well as unpublished research data and statistical analysis [1]. By means of an ABX-listening test detecting the perceptual difference of correlated and uncorrelated measurement noise in binaural room impulse responses, we analyse how documentation and reproducibility of such experiments can be improved by technical approaches and methodological aspects. In particular, the documentation of the experiment including the hypothesis, the preparation, and the analysis is realized in an Electronic Laboratory Notebook (ELN). Before the listening test started, a pre-processing script is implemented and automated processes are set-up in order to evaluate new measurements. Finally, we discuss the publication of the data and the data analysis scripts including their documentation with respect to the FAIR principles to ensure their re-usability.

Use Case: ABX-Listening Test The study of Hahne and Spors [2] entitled "On the perceptual difference of correlated and uncorrelated measurement noise in binaural room impulse responses" was employed to demonstrate the proposed approach.[1] In particular, the study is based on an ABX-listening test that was subsequently conducted by 13 subjects in the audio lab of the Institute of Communications Engineering at the University of Rostock over a period of four days.

A listening test typically starts with the formulation of a hypothesis, followed by several preparation steps and the data collection, which will be analysed and scientifically published afterwards. With respect to reproducibility, it is crucial to both: (1) document sufficient details along these activities, and (2) store all necessary data, scripts and other research objects that were produced. Considering these two aspects, we can derive several requirements for potential workflow support:

R1 **Prepare data processing in advance** in order to avoid p-Hacking [3] or HARKing [4]. While this does not exclude exploratory data analysis, it avoids that researchers search for significant correlations in their data that can not be explained by the original measurements. Furthermore, the study and data analysis design can be pre-registered in order to ensure high quality by reviewing of the hypothesis, method and details of the study from the scientific community. Thus, further mistakes such as a too small number of participants or the selection of a wrong study design with respect to the hypothesis might be avoided.

R2 **Reduce media disruption** to avoid errors introduced from the transferring of information. As an example, the answers of a digital questionnaire should be kept digital and not printed in order to be digitized again in a spreadsheet.

R3 **Share information in consortia** in order to allow members of the research group respectively the project partners in other institutions to get all information about the study. This increases transparency and fosters collaboration while at the same time access control can be applied.

R4 **Store versions of data** in order to track changes of the data with respect to questions of accountability but also to restore previous versions. Furthermore, together with R3, this prevents questions about the latest version of the data.

R5 **Reduce manual interactions** to prevent errors as well as to make the research study more transparent. During the handling of data, e. g., during the data analysis, manual interactions such as in the form of click-and-point software are difficult to document so that the measurement data can be re-analysed with the same results. Furthermore, such interactions may introduce errors from mistakes such as pasting information into a wrong location or by accidentally selecting the wrong column in a spreadsheet.

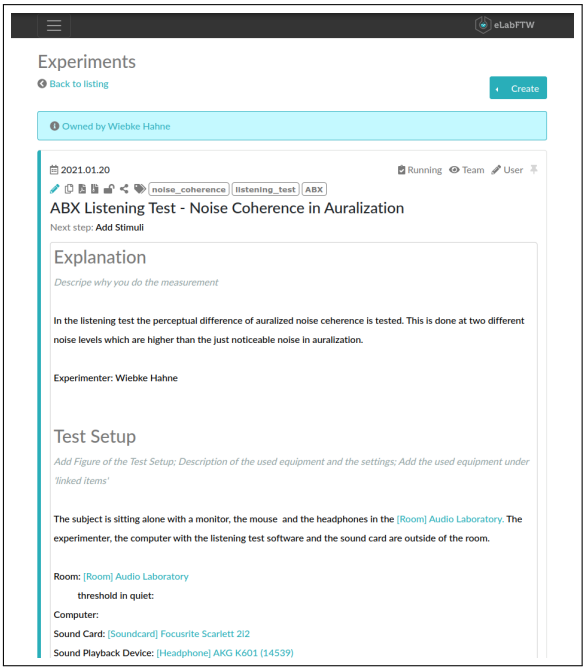
R6 **Enable publication of research objects** to support open science and reproducibility. In particular, the research objects produced in the study as well as their documentation should be publishable. As an example, tools that do not allow to export the produced research data in an community standard format should be avoided.

Note that reproducibility of scientific investigations involves a multitude of other aspects that cannot be discussed to the full-extent here. Next, we discuss our two main aspects: documentation and data storage as well as automation with respect to these requirements.

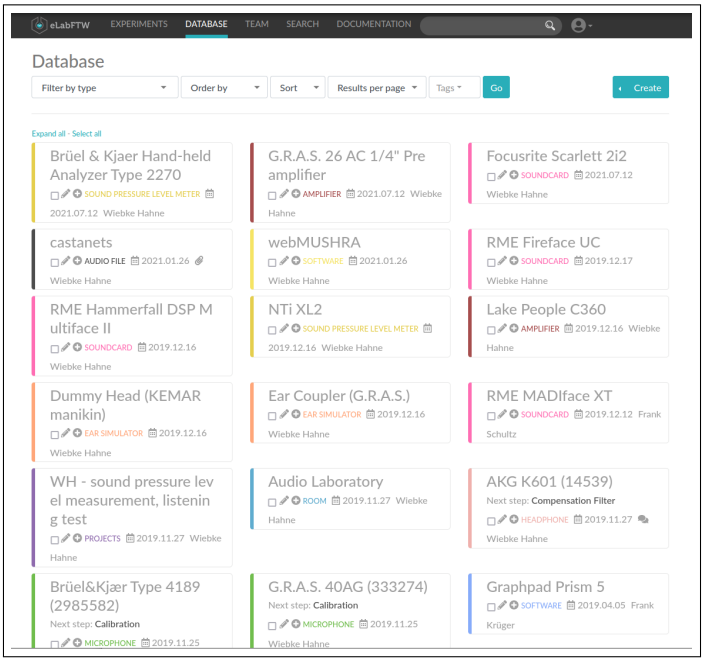
2 Electronic Documentation

Documenting research processes is essential for the comprehensibility and reproducibility of the research results

[1] Note, we only take this study as an example to analyse the reproducibility aspects of such listening tests. All details of this particular study are discussed in [2].



(a) Protocol documenting the listening test including setup, participants, and data storage



(b) Overview of database items centrally documenting software, hardware, and lab details

Figure 1: The documentation of the listening test is performed in an ELN by writing a protocol for each listening test and connecting the used resources to it. Resources are centrally managed by database items the reduce the documentation burden.

and findings. While information about the processes might be documented in analogue notes, errors might be introduced when digitizing them in order to use these information during the data analysis (R2). Furthermore, analogue information can not be shared within the consortia (R3). On the other hand, there exists a huge number of digital alternatives that satisfy these requirements. From text processing tools such as LibreOffice Writer[2], to note management systems such as Microsoft OneNote[3], to project management systems such as Redmine[4], to wiki systems such as MediaWiki[5]. While text processing tools are very easy to handle they do not provide a simple mechanism to store and connect multiple documents such as listening tests of the same kind using the same hardware. Furthermore, sharing (R3) of the information is not possible except from sending files to the project consortia. This is where note management and project management but also wiki systems are better suited. While the documentation is assigned to particular projects in the project management system, wiki and note management systems do not have a predefined structure.

Inspired by their use for the documentation of wet-lab experiments, we want to introduce another category of software: Electronic Laboratory Notebooks (ELNs) are specifically tailored for the documentation of experiments in the wet-lab often providing a central database that can be used to centrally maintain information of e. g., devices, software, and materials. This allows the reuse of existing information and thus reduces the potential for errors.

All described tools can be used for the documentation of ABX-listening tests. In our use case we decided to use the ELN tool *elabFTW* [5] as it provides a user-friendly interface and integrates well into the existing infrastructure. Figure 1a shows an example protocol containing an excerpt of the listening test documentation: a general part including the hypothesis, the test setup including details about the room and the hardware as well as the participants and the data storage. Furthermore, a central database was used to maintain information about soundcards, amplifier, microphones, and any other device or software. Figure 1b contains an overview of these database items defined in our ELN–each of them contains an extended description that is not shown in the figure. To reduce the documentation effort for similar experiments, templates might be created.

3 Automation

Besides the documentation of the listening test, the management of the research data and its analysis is the second important aspect to foster reproducibility. While the planning of this data management for an entire research project can be documented in so-called Data Management Plans (DMPs) [7], we consider their implementation for our use case here. First of all, the research data needs to be stored in a way that enables their sharing within the consortia (R3) as well as their versioning (R4). A lot of specifically tailored tools exist that can be employed for this purpose: versioning systems such as Git

[2]see `https://www.libreoffice.org/`
[3]see `https://www.onenote.com/`
[4]see `https://www.redmine.org/`
[5]see `https://www.mediawiki.org/`

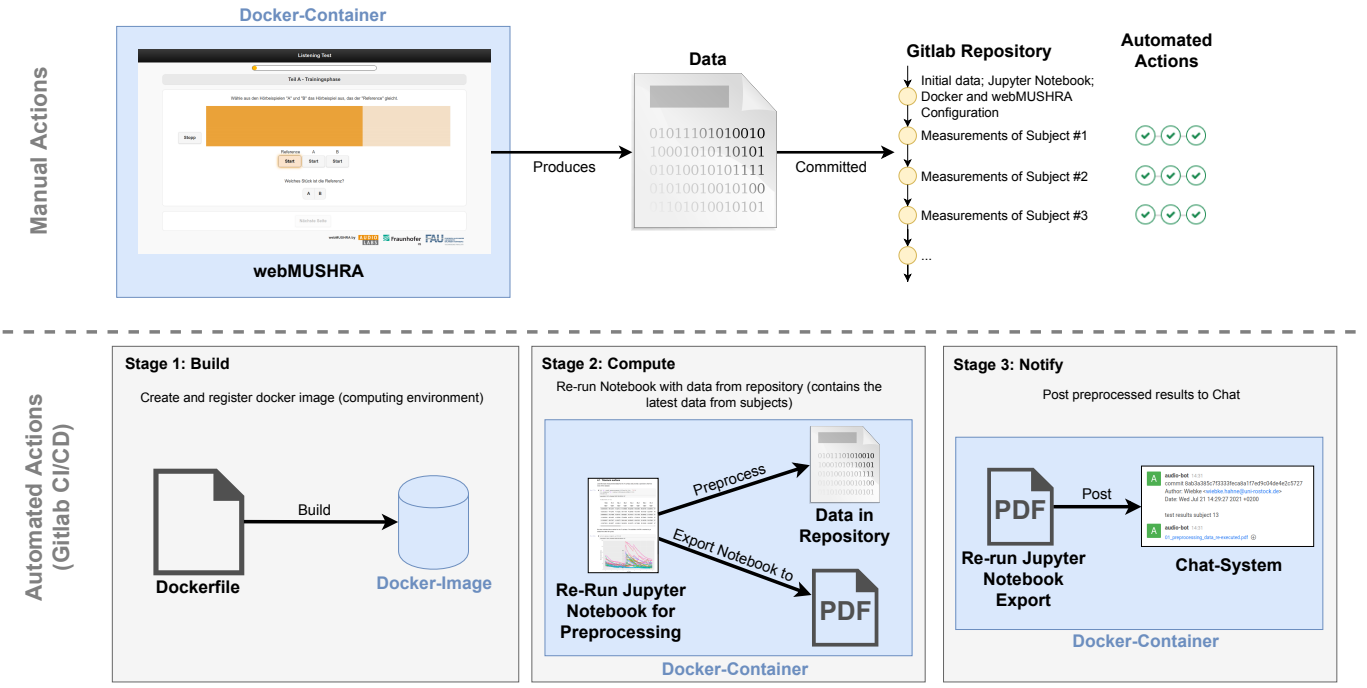


Figure 2: Implementation of the automated pre-processing and notification system. The upper part illustrates the data flow initiated by the listening test: The participant controls the webMUSHRA software [6] and the resulting data is committed into a GIT repository by the operator afterwards. Every such commit initiates the automated workflow in the lower part that (1) Builds a read-to-use Docker environment for the analysis, (2) Re-runs a Jupyter notebook performing the pre-processing of the repository data, and (3) Pushes the resulting PDF into a chat system in order to notify the operator about the results.

and SVN; project management systems such as Redmine and Microsoft Sharepoint; or repository software such as Dataverse[6], DSpace[7], and CKAN[8]. Schröder et al. [8] compare several tools with respect to the requirements of data management in large research consortia. They, furthermore, conclude that general repository software is best suited to fulfill these requirements. In our use case, the main focus is not only on the sharing of research data within the consortia, where the data that will be shared must be curated with additional meta information in order to enable others to understand the data. Furthermore, we aim at automation of the data handling and their analyses in order to reduce manual interactions (R5).

In the field of software development, build server such as Jenkins[9] or Continuous Integration and Development (CI/CD) server such as Github Actions[10] or Gitlab CI/CD[11] are used to process automated tasks such as testing the current software and deploying it to the production server. Similarly in our use case, we want to process the current research data by checking if the data recording worked without any problems and by pre-processing the data. Given the particular requirements and the domain specific background of the participating researcher, we decided to use a Gitlab versioning system including its CI/CD server to perform the data storage and automation tasks. However, we see the use of an institutional repository as a subsequent step for the sharing of the data in larger research consortia.

Figure 2 illustrates the workflow in our use case[12]. Particularly, after each participant performed the listening test, the new data will be committed as an individual version to the Git repository. Each of these commits invoke three automation steps that build the computing environment, runs the pre-processing (R1), and notifies the operator of the listening test with the results (see Figure 2 for details). Especially the building of the computing environment using a containerization solution is essential in order to ensure the reproducibility of the data analysis. Thus, errors from different library versions or difficulties in installing the software dependencies can be avoided.

4 Publication

After all data are collected and analysed, a potential next step could be their scientific publication. Part of this is also the publication of the data as well as the source code itself in order to enable reproduction and re-use [1, 9]. Recommendations for the publication of research data have been proposed in the so-called FAIR principles [10] which corresponds to: Findable, Accessible, Interoperable, and Re-usable. Together with the domain-

[6] see `https://dataverse.org/`
[7] see `https://duraspace.org/dspace/`
[8] see `https://ckan.org/`
[9] see `https://www.jenkins.io/`
[10] see `https://github.com/features/actions`
[11] see `https://docs.gitlab.com/ee/ci/`

[12] This workflow is adapted from `https://github.com/m6121/Workshop-ORDS-Reproducible-Data-Science`

independent implementation recommendations of Jacobsen et al. [11], they build a framework for research data publication.

A workflow for publication of our ABX-listening test data and analysis source code can be implemented by pushing the data to public Git repository services such as Github[13] or Gitlab[14] (R6). However, with respect to the FAIR principles it is not sufficient to create a repository, commit the data and reference this repository in the scientific publication. Instead, the research data should be curated with meta information such as author, title and a detailed description of the study, the data and the source code. Furthermore, a license should be specified as well as a persistent identifier such as a DOI that points to the particular version of the repository data that is used. To enable this archival, the Zenodo[15] platform can be used. By connecting the Github account to Zenodo, a new DOI can be automatically created for each release in the Github repository. Afterwards, the meta data can be adjusted in order to acknowledge e. g., authorship and funding. Recently, both platforms added support for the Citation File Format (CFF) [12], that provides a standard to define meta information in the repository. By providing these information, software can be unambiguously cited which also fosters reproducibility. Apart from the particular platforms, these aspects are included in the FAIR recommendations that contain also further improvements of the publication such as the use of ontologies for meta data terms and the use of standards that should also be implemented.

5 Conclusion

In this paper we analysed the requirements of ABX-listening tests using the example of Hahne and Spors [2] and discussed how reproducibilty of such studies can be fostered by electronic documentation and automation. In particular we discussed advantages and disadvantages of different software tools for documentation as well as data management and automation tasks with respect to this use case. In the future, reproducibility of acoustic experiments and the re-use of data can be further improved if additional standards for meta data, exchange formats, and repositories are developed.

Acknowledgement

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1270/2 - 299150580.

References

[1] M. Baker. "1,500 scientists lift the lid on reproducibility". In: *Nature* 533.7604 (May 2016), pp. 452–454. DOI: 10.1038/533452a.

[2] W. Hahne and S. Spors. "On the perceptual difference of correlated and uncorrelated measurement noise in binaural room impulse response". In: *German Annual Conference on Acoustics (DAGA)*. Aug. 2021.

[3] M. L. Head, L. Holman, R. Lanfear, A. T. Kahn, and M. D. Jennions. "The Extent and Consequences of P-Hacking in Science". In: *PLOS Biology* 13.3 (Mar. 2015), e1002106. DOI: 10.1371/journal.pbio.1002106.

[4] N. L. Kerr. "HARKing: Hypothesizing After the Results are Known". In: *Personality and Social Psychology Review* 2.3 (Aug. 1998), pp. 196–217. DOI: 10.1207/s15327957pspr0203_4.

[5] N. CARPi, A. Minges, and M. Piel. "eLabFTW: An open source laboratory notebook for research labs". In: *The Journal of Open Source Software* 2.12 (Apr. 2017), p. 146. DOI: 10.21105/joss.00146.

[6] M. Schoeffler, S. Bartoschek, F.-R. Stöter, M. Roess, S. Westphal, B. Edler, and J. Herre. "webMUSHRA — A Comprehensive Framework for Web-based Listening Tests". In: *Journal of Open Research Software* 6 (2018). DOI: 10.5334/jors.187.

[7] W. K. Michener. "Ten Simple Rules for Creating a Good Data Management Plan". In: *PLOS Computational Biology* 11.10 (Oct. 2015). Ed. by P. E. Bourne, e1004525. DOI: 10.1371/journal.pcbi.1004525.

[8] M. Schröder, H. LeBlanc, S. Spors, and F. Krüger. "Intra-consortia data sharing platforms for interdisciplinary collaborative research projects". In: *it - Information Technology* 62.1 (Feb. 2020), pp. 19–28. DOI: 10.1515/itit-2019-0039.

[9] R. D. Peng. "Reproducible Research in Computational Science". In: *Science* 334.6060 (Dec. 2011), pp. 1226–1227. DOI: 10.1126/science.1213847.

[10] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, and et al. "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific Data* 3.1 (Mar. 2016). DOI: 10.1038/sdata.2016.18.

[11] A. Jacobsen, R. de Miranda Azevedo, N. Juty, D. Batista, and et al. "FAIR Principles: Interpretations and Implementation Considerations". In: *Data Intelligence* 2.1-2 (Jan. 2020), pp. 10–29. DOI: 10.1162/dint_r_00024.

[12] S. Druskat et al. *Citation File Format*. en. 2021. DOI: 10.5281/ZENODO.5171937.

[13] see https://github.com/

[14] see https://gitlab.com/

[15] see https://zenodo.org/