

---

SOFTWARE METAPAPER

# webMUSHRA — A Comprehensive Framework for Web-based Listening Tests

Michael Schoeffler, Sarah Bartoschek, Fabian-Robert Stöter, Marlene Roess, Susanne Westphal, Bernd Edler and Jürgen Herre

International Audio Laboratories Erlangen, DE

Corresponding author: Michael Schoeffler, Project lead and development ([mail@mschoeffler.de](mailto:mail@mschoeffler.de))

---

For a long time, many popular listening test methods, such as ITU-R BS.1534 (MUSHRA), could not be carried out as web-based listening tests, since established web standards did not support all required audio processing features. With the standardization of the Web Audio API, the required features became available and, therefore, also the possibility to implement a wide range of established methods as web-based listening tests. In order to simplify the implementation of MUSHRA listening tests, the development of webMUSHRA was started. By utilizing webMUSHRA, experimenters can configure web-based MUSHRA listening tests without the need of web programming expertise. Today, webMUSHRA supports many more listening test methods, such as ITU-R BS.1116 and forced-choice procedures. Moreover, webMUSHRA is highly customizable and has been used in many auditory studies for different purposes.

---

**Keywords:** web-based; listening tests; auditory experiments; MUSHRA; Web Audio API

---

## (1) Overview

### Introduction

In audio research, listening test methods are often utilized to investigate the auditory system of humans or to evaluate audio systems. On the one hand, when the auditory system is the subject of interest, experimenters make use of all kinds of psychophysical methods, such as the forced-choice method (AB) or the unforced-choice method (ABN) [1], in their auditory experiments. On the other hand, when the main subject of interest is to subjectively evaluate audio systems by listening tests, experimenters make often use of standardized or formalized evaluation methods, such as ITU-R BS.1534 (MUSHRA) [2] or ITU-R BS.1116 [3]. In many scenarios, especially if no specific audio hardware is required, both types of listening tests can be carried out over the Internet using web-based listening tests. The advantage of web-based listening tests is that they can be accessed by everyone who has a device with a compatible web browser and an Internet connection. Moreover, participants can participate from almost anywhere at anytime and, in most cases, without the need to install additional software. It has been shown by multiple researchers that there exist only minor differences between listening tests carried out in an laboratory environment and those carried out over the Internet, if the experiment was properly designed [4, 5]. For a long time, not all types of listening test methods could be implemented with existing web standards, as, e.g., sample-by-sample manipulation is required by some

methods but has not been supported by any accepted web standard. Since the release of the Web Audio API standard, many browsers started to implement the standard and thereby support audio processing features required by many listening test methods. One of these methods is MUSHRA, which requires sample-by-sample manipulation to implement a standard-compliant fade-out and fade-in. Hence, the webMUSHRA project has been initiated [6]. The goal of the project has been to establish a framework which assists experimenters to carry out listening tests without the need to program web applications. At first, webMUSHRA was specifically designed to support only MUSHRA-compliant listening tests. In the meanwhile, webMUSHRA supports many more methods, such as ITU-R BS.1116, forced and unforced methods, single-stimulus and multi-stimulus procedures, subjective evaluation based on overall listening experience, as well as 2D- and 3D-based graphical localization methods. This paper introduces the main features of webMUSHRA, including an overview of studies that used webMUSHRA in various contexts.

Besides webMUSHRA, there exist other frameworks that can be utilized to conduct web-based listening tests: BeaqleJS is a framework that supports ABX- and MUSHRA-based listening tests [7]. Moreover, the Web Audio Evaluation Tool supports a wide range of methods, such as MUSHRA and ITU-R BS.1116 [8]. In comparison to these frameworks, webMUSHRA supports some additional methods, such as reporting methods for spatial

attributes. For example, these reporting methods enable experimenters to conduct sound localization tests or evaluations of spatial audio quality provided by auditory virtual environments (AVEs) [9].

### Implementation and architecture

The basic software architecture of webMUSHRA is based on *pages*. The idea is that an experiment consists of pages. On each page, something is presented to the participants, e.g., a play-back button to play back a stimuli or a rating scale in order to obtain responses. A single scene of a listening test method, e.g., rating screen in MUSHRA, is implemented as such a page. An experimenter configures the whole procedure of the experiment based on such pages and stores the configuration in a YAML file. When an experiment is started, pages are created based on the configuration and, then, handed over to the *page manager*. During an experiment, the page manager holds the data of all pages, including the given responses of the listener. Moreover, if a listener has finished a page, the page manager takes care of showing the next page until the whole experiment is completed. As a final step, the page manager transfers the responses to a PHP script (path: “service/write.php”) that stores the listener’s responses in a CSV file. In **Figure 1**, a schematic of the software architecture is shown. If a developer wants to add support for a new listening test method, he or she has to create a new page type by implementing the page interface. The page interface is represented by a JavaScript class having six methods (path: “lib/webmushra/pages/Page.js”):

**getName()** Returns the name of the page type. Each new page type requires a unique page name.

**init(\_callbackError)** This method is called when the page is requested to be initialized. The method’s

parameter is a callback function that has to be called if an error occurs.

**render(\_parent)** This method is called when the page has to render its graphical elements. The parameter is the parent DOM element where the page can attach its elements.

**load()** This method is called after the render-method. The purpose of this method is to load default values or saved values for the graphical elements.

**next()** This method is called if the listener proceeds to a new page. The method can be used to save entered data, e.g., if the page is called a second time.

**store(\_reponsesStorage)** If a listening session is completed, this method is called to allow all pages to store their collected responses into a so-called response storage. The response storage is utilized to create a CSV file containing the listening test results.

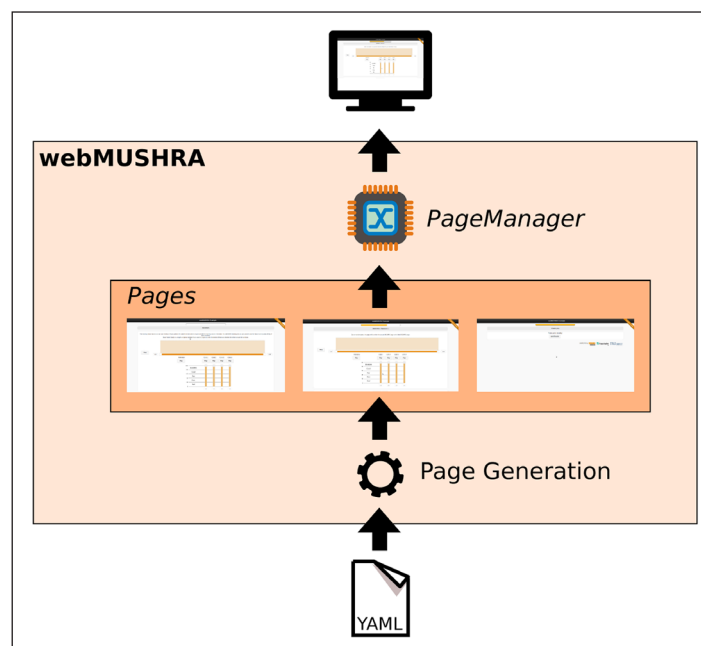
Moreover, the developer must register the new source file (to index.html) and page type (to startup.js). If the new page type should store responses of listeners to the CSV file, the PHP script (path: “service/write.php”) must be extended by a storage function of the new page type.

### Selection of supported methods

In this section, a selection of listening test methods are briefly introduced that are supported by webMUSHRA.

#### ITU-R BS.1116

In ITU-R BS.1116, a listening test method is described to assess the basic audio quality (BAQ) of audio systems which introduce small impairments to an original audio [3]. BAQ is defined as “this single, global attribute is used to judge any and all detected differences between the reference



**Figure 1:** High-level schematic of the software architecture. First, based on YAML-based configuration files, *pages* are generated that describe what is shown to a listener. Then, the *page manager* presents the pages in the configured order to the listener.

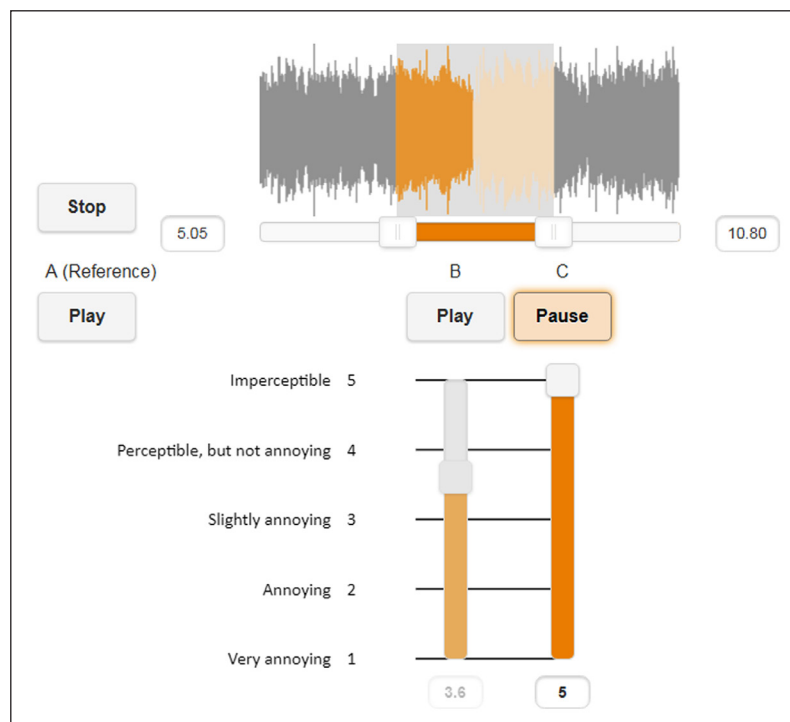
and the object". Here, the reference is often an undistorted original of an auditory stimulus, whereas the object is a degraded (or encoded) version of the same auditory stimulus. In an ITU-R BS.1116 listening test, the actual assessment is based on a "double-blind triple-stimulus with hidden reference" method. The test method involves three kinds of stimuli: reference (undistorted original), hidden reference (copy of the reference), and stimuli processed by systems under test (known as condition or object). The listeners are presented with three stimuli which are labeled as "A", "B", and "C". Stimulus "A" is always the reference, which is known to the listener. The hidden reference and the condition are randomly assigned to "B" and "C". Thereby, the listeners do not know which kind of stimulus is behind which label. The listeners are asked to assess the impairments on "B" compared to "A", and "C" compared to "A", according to the continuous quality scale (CQS) [2]. The grading must reflect "B" and "C"'s provided BAQ compared to "A". Due to the definition of BAQ, any perceived differences between the reference and the other stimuli must be interpreted as an impairment. The continuous quality scale ranges from 1.0 to 5.0 and has five anchor points on each whole number which are labeled with "Very annoying", "Annoying", "Slightly annoying", "Perceptible, but not annoying", and "Imperceptible" [10]. **Figure 2** shows a screenshot of a ITU-R BS.1116 listening test designed with webMUSHRA.

#### ITU-R BS.1534 (MUSHRA)

In contrast to Recommendation ITU-R BS.1116, Recommendation ITU-R BS.1534, better known as MUSHRA (Multi-Stimulus Test with Hidden Reference and Anchor), is a more recent description of a listening test method for evaluating audio systems that introduce

intermediate impairments [2]. Furthermore, MUSHRA is based on a multi-stimulus comparison, meaning the reference and more than one condition can be accessed freely at random. In particular, listeners are presented with a (open) reference and multiple conditions. Among the conditions are the stimuli processed by the systems under test, a hidden reference, and two anchor stimuli. The two anchors, so-called low-quality anchor and mid-quality anchor, are low-pass-filtered versions of the reference stimulus and were introduced to make the ratings of different assessors and labs more comparable [11, 12]. The low-quality anchor has a cut-off frequency of 3.5 kHz and the mid-quality anchor has a cut-off frequency of 7 kHz. These two types of anchors can be automatically generated by webMUSHRA if configured by the experimenter. During a MUSHRA listening test, the conditions are presented in random order without any information that would identify the condition being an audio system under test, the hidden reference, or an anchor. Like in an ITU-R BS.1116 test, listeners can instantaneously switch between the reference and the conditions when listening. Further, listeners are asked to rate the BAQ of the stimuli as well. In a MUSHRA test, a different continuous quality scale is used for giving grades. The scale ranges from 0 to 100 and is divided into five equal intervals with the adjectives "Bad", "Poor", "Fair", "Good", and "Excellent". Although multiple conditions are rated on the same page, the wide scale range makes it still possible to rate very small differences.

Although MUSHRA has been designed for audio systems introducing intermediate impairments, many so-called "MUSHRA-like" listening tests have been carried out to assess all kinds of audio systems. For example, the open reference or the anchors are often left out for various



**Figure 2:** ITU-R BS.1116 listening test.

reasons. The webMUSHRA software can easily be modified and is, therefore, especially helpful to experimenters who have basic knowledge about web development and plan to carry out such MUSHRA-like experiments. **Figure 3** shows a screenshot of a MUSHRA listening test designed with webMUSHRA.

**Likert scale/Evaluation of overall listening experience**

The software supports full customization of Likert scales. Likert scales are widely used in research and can be found in all kinds and variations. Likert scales consist of so-called Likert items, which are typically horizontally aligned. Likert items correspond to a participant's statement that he or she is asked to evaluate by giving it a quantitative value. Usually, this quantity is given as level of agreement/disagreement. When configuring webMUSHRA, the experimenter can add as many Likert scales as desired. Moreover, it is possible to assign images to Likert items. Therefore, it is possible to create, e.g., five-star Likert scales.

Five-star Likert scales can be used for evaluating the perceived overall listening experience (OLE) when assessing audio systems [13]. In contrast to BAQ-based methods, participants are asked to rate the stimuli according to how much they like, enjoy, or feel pleased when listening to the stimuli. Thereby, participants are allowed to involve affective aspects, like emotional or individual aspects. Typically, five-star Likert scales are utilized in this type of evaluation, since they are known to lead to consistent ratings [14, 15]. The method comprises two phases: In the first phase, participants rate stimuli that (if possible) have not been processed by any audio system under test in a multi-stimulus procedure. These ratings, called "basic item ratings", are expected to predominantly reflect how much the content of the stimuli are liked by a listener. In the second phase, participants rate all stimuli processed by the audio systems under test in a single-stimulus procedure. These ratings are called "item ratings".

As described in our previous paper [16], having these two rating types has several advantages: The distribution of the ratings given in the first phase indicate whether the content of stimuli was well balanced regarding the ratings. For example, if the content per se is not liked by the participants, this will result in a large percentage of very low ratings, also if the stimuli were processed by the audio systems under test. As a consequence, in many cases, the results analysis might not be suited for evaluating a desired research question due to the low variances between ratings. With this procedure, experimenters can test early on whether the likening of the stimuli is different across the participants.

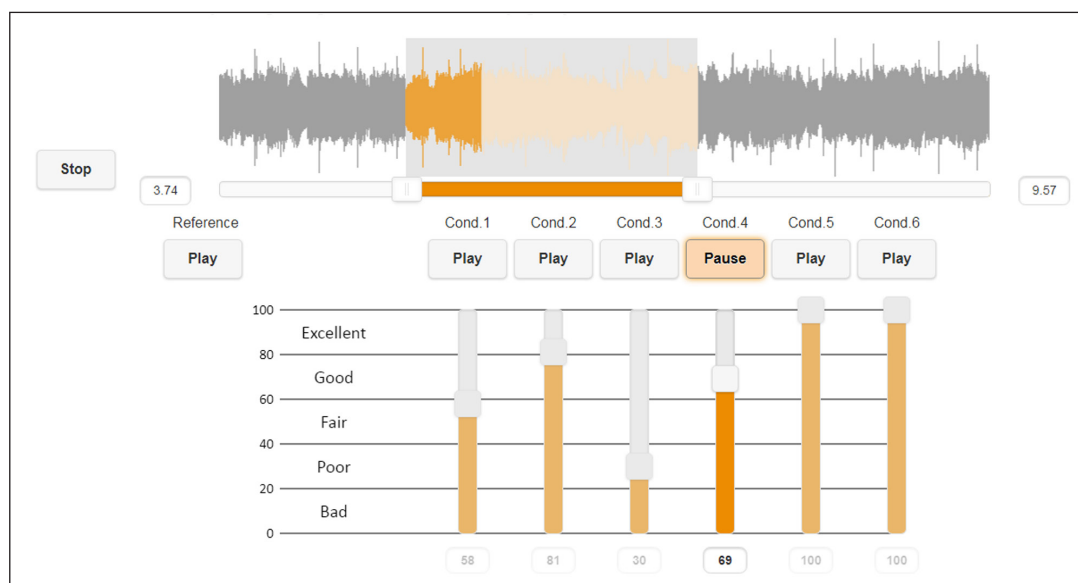
As webMUSHRA allows to configure five-star Likert scales and uses them in a single-stimulus procedure as well as in a multi-stimulus procedure, setting up OLE evaluations is very simple.

When assessing audio systems, OLE-based evaluations are a useful addition to BAQ-based evaluations [16]. As BAQ-based evaluations predominantly reveal the rather technical differences between the audio systems under test, they do not give sufficient indications whether a technical superiority of an audio system will be appreciated in the end-user scenario.

Also in early phases of an audio system development, decision makers can utilize results of an OLE-based evaluation in order to get insights whether the gain of audio quality is also reflected by a gain in the perceived overall listening experience by end-users.

**Reporting method for spatial attributes**

Today, there exists a wide range of graphical user interfaces (GUIs) for localization tests in which participants are asked to report the spatial location of auditory stimuli. Unfortunately, these GUIs are often limited to the horizontal plane. As a consequence, these GUIs are not perfectly suited for evaluating advanced multi-channels formats, in which listeners are fully surrounded by



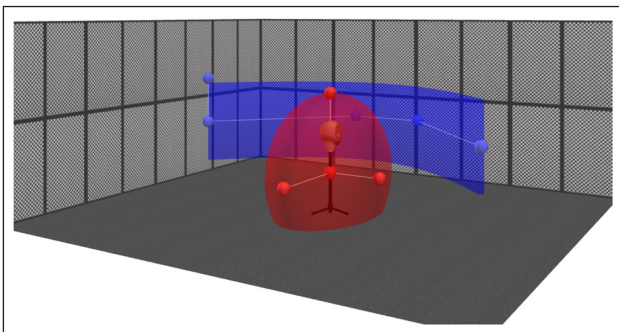
**Figure 3:** ITU-R BS.1534 (MUSHRA) listening test GUI.

loudspeakers. Since these formats ask for evaluation methods which support reporting spatial attributes in all three dimensions. To this end, webMUSHRA features a reporting method for three-dimensional spatial attributes in listening tests [17]. The reporting method enables to report width, height, depth, and location of stimuli. In addition, reporting the apparent/auditory source width (ASW) and listener envelopment (LEV) is also supported. The ASW is defined by Morimoto as the width of the sound image fused temporally and spatially with a direct sound's image [18]. LEV is defined by Norcross et al. as the listener's sense of being surrounded or enveloped by sound [19]. Dependent on the author, envelopment is sometimes interpreted as surrounded only on the horizontal plane. In these cases, the term engulfment is sometimes used to express being covered by sound [20].

The reporting method supports a 2D- and 3D-based graphical user interface (GUI) for reporting the perception regarding the spatial attributes. In **Figure 4**, the 3D-based GUI is shown. Although reporting the perceived location of sound sources by pointing (with or without the extension of a body part) has been found to be the most accurate method [21], webMUSHRA's reporting method for spatial attributes is still valuable in many scenarios. For example, if an experimental setup or required devices for pointing are not available or if a time-efficient method is needed. Using such 3D- and 2D-based GUIs for localization listening tests has been evaluated in [22].

### Quality control

The software aims at providing established listening tests as well as new concepts of listening methods to a wide range of experimenters with various backgrounds. Due to its background in research, also experimental features and methods are included in the main version of webMUSHRA. For this and other reasons, webMUSHRA should not be seen as a perfectly reliable evaluation tool that can be used for judging audio systems in official performance evaluations. For webMUSHRA, it is not possible to guarantee that the audio processing works correctly, due to its web-based nature and, therefore, its lack of control of the underlying browser software, operation system, audio driver, and audio periphery. Nonetheless, webMUSHRA has already been used in a wide range of listening tests



**Figure 4:** The reporting method's 3D view showing a scene with a blue extruded curve (for reporting width, height, and depth) and a red ellipsoid (for reporting LEV).

(see Section 3) and has proven to be a reliable framework for all kinds of listening tests.

In order to prevent compile time errors, the continuous integration system Travis CI has been integrated to the development process. If new commits are pushed to the source code repository, Travis CI will build the whole project and check for compile errors. Moreover, to retain the stability and reliability of webMUSHRA, unit testing with QUnit has also been integrated into the development process. At the time of writing, the use of unit tests is voluntary for developers.

Furthermore, webMUSHRA is maintained by a well-known institution for audio research, the International Audio Laboratories Erlangen (AudioLabs), and, therefore, the quality control does not rely on a single maintainer. Within the AudioLabs, the development on webMUSHRA started in 2012 and since then has reached a mature state. The software has frequently been used internally for many types of experiments, which alone results in an active user base.

## (2) Availability

### Operating system

Any operating system on which a browser with Web Audio API support is available.

### Programming language

HTML5, PHP and JavaScript.

### Additional system requirements

Web server with PHP support.

### Dependencies

jQuery  $\geq 2.1.1$   
 jQuery Mobile  $\geq 1.4.4$   
 mousetrap  $\geq 1.4.6$   
 noUiSlider  $\geq 9.2.0$   
 yaml.js  $\geq 0.2.8$   
 three.js  $\geq r71$   
 QUnit  $\geq 1.17.1$

### List of contributors

- Michael Schoeffler (International Audio Laboratories Erlangen): Project lead and development.
- Sarah Bartoschek (International Audio Laboratories Erlangen): Development.
- Fabian-Robert Stöter (International Audio Laboratories Erlangen): Development.
- Marlene Roess (International Audio Laboratories Erlangen): Development.
- Susanne Westphal (International Audio Laboratories Erlangen): Development.
- Bernd Edler (International Audio Laboratories Erlangen): Consulting.
- Jürgen Herre (International Audio Laboratories Erlangen): Consulting.

### Software location

#### Archive

**Name:** Zenodo

**Persistent identifier:** <https://doi.org/10.5281/zenodo.1069840>

**Licence:** Software License for the webMUSHRA.js Software

**Publisher:** International Audio Laboratories Erlangen

**Version published:** 1.4.0

**Date published:** 08/06/2017

#### Code repository

**Name:** Github

**Persistent identifier:** <https://github.com/audiolabs/webMUSHRA>

**Licence:** Software License for the webMUSHRA.js Software

**Date published:** 24/03/2017

#### Language

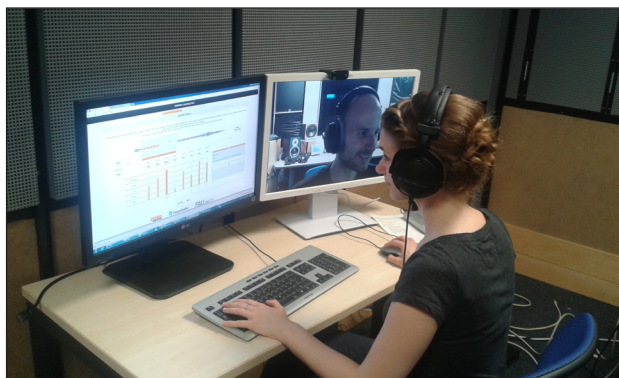
HTML5, PHP, and JavaScript

### (3) Reuse potential

Instead of having a well-known license approved by the Open Source Initiative, webMUSHRA has a customized (MIT-like) open source license. Nonetheless, redistribution and use of webMUSHRA in source and binary forms, with or without modification, are permitted without payment of copyright license fees provided that the license's conditions are satisfied.

On the one hand, since webMUSHRA supports many established methods for subjective evaluations, such as ITU-R BS.1534 and ITU-R BS.1116, anyone who wants to evaluate audio systems can make use of it. On the other hand, webMUSHRA is well suited for auditory experiments, since fundamental methods, such as forced-choice, unforced-choice, single-stimulus, and multi-stimulus are supported.

Within the last years, pre-release versions of webMUSHRA have been used in a wide range of experiments with different contexts. Next, a selection of these experiments is presented in order to demonstrate the capabilities and possibilities of the software. An early version of webMUSHRA has been used in a psychoacoustically-motivated auditory experiment [23]. In this experiment, listeners were asked to estimate the number of instrumental voices in short music recordings. Here, webMUSHRA served as a basic framework for auditory experiments, as it already comes with features related to configuration, audio device initialization, etc. The page, on which listeners responded with the number of estimated instruments, was developed from scratch and added as a new page type. In another experiment [24], the localization method was slightly modified and utilized to evaluate a novel re-panning method. An example of a MUSHRA-based experiment can be found in [25], in which the MUSHRA method was modified to investigate the perceived density of synthesized applause signals. Instead of having one reference signal, two reference signals were required by the evaluation. In order to investigate a novel approach for assessing audio systems, webMUSHRA was used in a listening test utilizing the technique of so-called “distributed pair evaluation” [26]. The idea of



**Figure 5:** A listening test in which a technique called “distributed pair evaluation” was used.

this technique originates from software development and is called (distributed) pair programming, in which two programmers work as a pair together on the same code. Applying this technique is known to improve code quality. In the experiment, participants were assigned into pairs and collaboratively evaluated the BAQ of the audio codecs by the MUSHRA method. When participants worked in pairs, they were spatially separated from each other, but able to communicate by video-, voice-, and text-chat. In order to enable this communication, a WebRTC client was integrated into the user interface of webMUSHRA (see **Figure 5**). As one can see, webMUSHRA has already been utilized in multiple scenarios. Therefore, it is expected that the frequent use of webMUSHRA will continue in the future.

#### Acknowledgements

The authors want to thank Alexander Adami for extensively using pre-release versions of webMUSHRA for his experiments. His feedback has been very valuable and significantly helped to stabilize and improve the initial release version of webMUSHRA.

#### Competing Interests

The authors have no competing interests to declare.

#### References


1. **Punch, J L, Rakerd, B and Amlani, A M** 2001 Paired-comparison hearing aid preferences: Evaluation of an unforced-choice paradigm. *Journal American Academy of Audiology*, 12: 190–201.
2. **International Telecommunication Union** 2015 Recommendation ITU-R BS.1534-3: Method for the subjective assessment of intermediate quality level of audio systems. <https://www.itu.int/rec/R-REC-BS.1534/en>.
3. **International Telecommunication Union** 2015 Recommendation ITU-R BS.1116-3: Methods for the subjective assessment of small impairments in audio systems. <https://www.itu.int/rec/R-REC-BS.1116/en>.
4. **Schoeffler, M, Stöter, F-R, Bayerlein, H, Edler, B and Herre, J** 2013 An Experiment about Estimating the Number of Instruments in Polyphonic Music: A Comparison Between Internet and Laboratory Results. *Proceedings of 14<sup>th</sup> International Society for Music*

- Information Retrieval Conference*. Curitiba, Brazil, International Society for Music Information Retrieval. [http://www.ppgia.pucpr.br/ismir2013/wp-content/uploads/2013/09/59\\_Paper.pdf](http://www.ppgia.pucpr.br/ismir2013/wp-content/uploads/2013/09/59_Paper.pdf).
5. **Cartwright, M, Pardo, B, Mysore, G J and Hoffman, M** 2016 Fast and easy crowdsourced perceptual audio evaluation. *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, 619–623. IEEE.
  6. **Schoeffler, M, Stöter, F-R, Edler, B and Herre, J** 2015 Towards the Next Generation of Web-based Experiments: A Case Study Assessing Basic Audio Quality Following the ITU-R Recommendation BS.1534 (MUSHRA). *1st Web Audio Conference*. Paris, France. [http://wac.ircam.fr/pdf/wac15\\_submission\\_8.pdf](http://wac.ircam.fr/pdf/wac15_submission_8.pdf).
  7. **Kraft, S and Zölzer, U** 2014 BeaqlJS: HTML5 and JavaScript based Framework for the Subjective Evaluation of Audio Quality. *Linux Audio Conference*. Karlsruhe, Germany.
  8. **Jillings, N, Moffat, D, De Man, B and Reiss, J D** 2015 Web Audio Evaluation Tool: A browser-based listening test environment. *12th Sound and Music Computing Conference*. July.
  9. **Novo, P** 2005 Auditory virtual environments. *Communication Acoustics*, 277–297. Springer Berlin Heidelberg.
  10. **International Telecommunication Union** 2003 Recommendation ITU-R BS.1284-1: General methods for the subjective assessment of sound quality. <https://www.itu.int/rec/R-REC-BS.1284/en>.
  11. **Zielinski, S K, Rumsey, F and Bech, S** 2008 On some biases encountered in modern audio quality listening tests—a review. *J. Audio Eng. Soc.*, 56: 427–451. <http://www.aes.org/e-lib/browse.cfm?elib=14393>.
  12. **Liebetrau, J, Nagel, F, Zacharov, N, Watanabe, K, Colomes, C, Crum, P, Sporer, T and Mason, A** 2014 Revision of Rec. ITU-R BS.1534. *Audio Engineering Society Convention 137*, 1–8. Los Angeles, United States, Oct. <http://www.aes.org/e-lib/browse.cfm?elib=17495>.
  13. **Schoeffler, M** 2017 *Overall Listening Experience – a new Approach to Subjective Evaluation of Audio*. Ph.D. thesis. [https://opus4.kobv.de/opus4-fau/files/8290/dissertation\\_schoeffler.pdf](https://opus4.kobv.de/opus4-fau/files/8290/dissertation_schoeffler.pdf).
  14. **Schoeffler, M, Conrad, S and Herre, J** 2014 The Influence of the Single/Multi-Channel-System on the Overall Listening Experience. *Proc. of the AES 55<sup>th</sup> Conference on Spatial Audio*. Helsinki, Finland. <http://www.aes.org/e-lib/browse.cfm?elib=17350>.
  15. **Schoeffler, M, Gernert, J L, Neumayer, M, Westphal, S and Herre, J** 2015 On the validity of virtual reality-based auditory experiments: A case study about ratings of the overall listening experience. *Springer Virtual Reality*, 1–20. [http://www.mschoeffler.de/wp-content/uploads/2017/01/manuscript\\_springer\\_vr\\_\\_Validity\\_Virtual\\_Reality\\_Auditory\\_Experiments.pdf](http://www.mschoeffler.de/wp-content/uploads/2017/01/manuscript_springer_vr__Validity_Virtual_Reality_Auditory_Experiments.pdf).
  16. **Schoeffler, M, Silzle, A and Herre, J** 2017 Evaluation of spatial/3d audio: Basic audio quality versus quality of experience. *IEEE Journal of Selected Topics in Signal Processing*, 11: 75–88. <http://ieeexplore.ieee.org/document/7782315/>. DOI: <https://doi.org/10.1109/JSTSP.2016.2639325>
  17. **Westphal, S, Schoeffler, M and Herre, J** 2015 A framework for reporting spatial attributes of sound sources. *Workshop für Innovative Computerbasierte Musikinterfaces (ICMI)*. <http://www.icmi-workshop.org/papers/2015/spacialattributes.pdf>.
  18. **Morimoto, M** 1997 The Role of Rear Loudspeakers in Spatial Impression. *Audio Engineering Society Convention 103*. New York, United States. <http://www.aes.org/e-lib/browse.cfm?elib=7225>.
  19. **Norcross, S G, Souloudre, G A and Lavoie, M C** 2003 Temporal Aspects of Listener Envelopment in Multichannel Surround Systems. *Audio Engineering Society Convention 114*. Amsterdam, Netherlands.
  20. **Paine, G, Sazdov, R and Stevens, K** 2007 Perceptual investigation into envelopment, spatial clarity, and engulfment in reproduced multi-channel audio. *Audio Engineering Society Conference: 31st International Conference: New Directions in High Resolution Audio*. Jun. <http://www.aes.org/e-lib/browse.cfm?elib=13961>.
  21. **Haber, L, Haber, R N, Penningroth, S, Novak, K and Radowski, H** 1993 Comparison of nine methods of indicating the direction to objects: data from blind adults. *Perception*, 22: 35–47. <http://www.ncbi.nlm.nih.gov/pubmed/8474833>. DOI: <https://doi.org/10.1068/p220035>
  22. **Schoeffler, M, Westphal, S, Adami, A, Bayerlein, H and Herre, J** 2014 Comparison of a 2D- and 3D-Based Graphical User Interface for Localization Listening Tests. *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, 107–112. Berlin, Germany. <https://depositonce.tu-berlin.de/handle/11303/175>.
  23. **Stöter, F-R, Schoeffler, M, Edler, B and Herre, J** 2013 Human ability of counting the number of instruments in polyphonic music. *Proceedings of Meetings on Acoustics*, 19. Montreal, Canada, Acoustical Society of America. <http://asa.scitation.org/doi/10.1121/1.4805760>.
  24. **Adami, A, Schoeffler, M and Herre, J** 2015 Re-Panning of Directional Signals and its Impact on Localization. *23rd European Signal Processing Conference (EUSIPCO)*, 549–553. <http://ieeexplore.ieee.org/document/7362443/>.
  25. **Adami, A, Disch, S, Steba, G and Herre, J** 2016 Assessing Applause Density Perception Using Synthesized Layered Applause Signals. *Proceedings of the 19th International Conference on Digital Audio Effects (DAFx-16)*, 183–189. Brno, Czech Republic. [http://dafx16.vutbr.cz/dafxpapers/26-DAFx-16\\_paper\\_15-PN.pdf](http://dafx16.vutbr.cz/dafxpapers/26-DAFx-16_paper_15-PN.pdf).
  26. **Neumayer, M, Schoeffler, M and Herre, J** 2015 Distributed pair evaluation in context of mushra listening tests. *Workshop für Innovative Computerbasierte Musikinterfaces (ICMI)*. <http://www.icmi-workshop.org/papers/2015/mushra.pdf>.

**How to cite this article:** Schoeffler, M, Bartoschek, S, Stöter, F-R, Roess, M, Westphal, S, Edler, B and Herre, J 2018 webMUSHRA — A Comprehensive Framework for Web-based Listening Tests. *Journal of Open Research Software*, 6: 8. DOI: <https://doi.org/10.5334/jors.187>

**Submitted:** 31 July 2017    **Accepted:** 11 January 2018    **Published:** 05 February 2018

**Copyright:** © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Journal of Open Research Software* is a peer-reviewed open access journal published by Ubiquity Press

**OPEN ACCESS** 