

Instructions for Web Audio Evaluation Tool

Nicholas Jillings, Brecht De Man and David Moffat

7 December 2015

These instructions are about use of the Web Audio Evaluation Tool on Windows and Mac OS X platforms.

We request that you acknowledge the authors and cite our work when using it [1], see also CITING.txt.

The tool is available in its entirety including source code on <https://code.soundsoftware.ac.uk/projects/webaudioevaluationtool/>, under the GNU General Public License v3.0 (<http://choosealicense.com/licenses/gpl-3.0/>), see also LICENSE.txt.

Contents

1	Installation	4
1.1	Contents	4
1.2	Compatibility	5
2	Test setup	7
2.1	Sample rate	7
2.1.1	Mac OS X	7
2.1.2	Windows	7
2.2	Local test	7
2.2.1	Mac OS X	8
2.2.2	Windows	9
2.3	Remote test	11
3	Interfaces	13
3.1	APE	13
3.2	MUSHRA	13

4	Features	14
4.1	Surveys	14
4.1.1	Pre- and post-page surveys	14
4.1.2	Pre- and post-test surveys	14
4.1.3	Survey elements	14
4.2	Randomisation	15
4.2.1	Randomisation of configuration XML files	15
4.2.2	Randomisation of page order	15
4.2.3	Randomisation of axis order	15
4.2.4	Randomisation of fragment order	15
4.2.5	Randomisation of initial slider position	15
4.3	Looping	15
4.4	Sample rate	15
4.5	Scrubber bar	15
4.6	Metrics	16
4.6.1	Time test duration	16
4.6.2	Time fragment playback	16
4.6.3	Initial positions	16
4.6.4	Track movements	16
4.6.5	Which fragments listened to	16
4.6.6	Which fragments moved	16
4.6.7	elementListenTracker	17
4.7	References and anchors	17
4.7.1	Reference	17
4.7.2	Hidden reference	17
4.7.3	Hidden anchor	17
4.8	Checks	17
4.8.1	Playback checks	17
4.8.2	Movement check	17
4.8.3	Comment check	17
4.8.4	Scale use check	18
4.8.5	Note on the use of multiple rating axes	18
4.9	Layout options	18
4.10	Multiple sliders	18
4.11	Platform information	18
4.12	Show progress	19
4.13	Gain	19
4.14	Loudness	19
5	Using the test create tool	20
5.1	Nodes to familiarise	20

5.2	Modifying <code>core.js</code>	20
5.3	Building the Interface	20
5.3.1	loadInterface	21
5.3.2	loadTest(audioHolderObject)	21
6	Analysis and diagnostics	22
6.1	In the browser	22
6.2	Python scripts	22
6.2.1	comment_parser.py	22
6.2.2	evaluation_stats.py	22
6.2.3	generate_report.py	22
6.2.4	score_parser.py	22
6.2.5	score_plot.py	23
6.2.6	timeline_view_movement.py	23
6.2.7	timeline_view.py	23
7	Troubleshooting	24
7.1	Reporting bugs and requesting features	24
7.2	First aid	24
7.3	Known issues and limitations	25
8	References	26
A	Legacy	27
B	Listening test instructions example	28
C	Terminology	29
	Contact details	30

1 Installation

Download the folder (<https://code.soundsoftware.ac.uk/hg/webaudioevaluationtool/archive/tip.zip>) and unzip in a location of your choice, or pull the source code from <https://code.soundsoftware.ac.uk/hg/webaudioevaluationtool> (Mercurial).

1.1 Contents

The folder should contain the following elements:

Main folder:

- `analyse.html`: analysis and diagnostics of a set of result XML files
- `ape.css`, `core.css`, `graphics.css`, `mushra.css`, `structure.css`: style files (edit to change appearance)
- `ape.js`: JavaScript file for APE-style interface [2]
- `CITING.txt`, `LICENSE.txt`, `README.txt`: text files with, respectively, the citation which we ask to include in any work where this tool or any portion thereof is used, modified or otherwise; the license under which the software is shared; and a general readme file referring to these instructions.
- `core.js`: JavaScript file with core functionality
- `index.html`: webpage where interface should appear (includes link to test configuration XML)
- `jquery-2.1.4.js`: jQuery JavaScript Library
- `loudness.js`: Allows for automatic calculation of loudness of Web Audio API Buffer objects, return gain values to correct for a target loudness or match loudness between multiple objects
- `mushra.js`: JavaScript file for MUSHRA-style interface [3]
- `pythonServer.py`: webserver for running tests locally
- `pythonServer-legacy.py`: webserver with limited functionality (no automatic storing of output XML files)
- `save.php`: PHP script to store result XML files to web server

Documentation (./docs/)

- DMRN+10: PDF and L^AT_EXsource of poster for 10th Digital Music Research Network One-Day workshop (“soft launch”)
- Instructions: PDF and L^AT_EXsource of these instructions
- Project Specification Document (L^AT_EX/PDF)
- Results Specification Document (L^AT_EX/PDF)
- SMC15: PDF and L^AT_EXsource of 12th Sound and Music Computing Conference paper [1]
- WAC2016: PDF and L^AT_EXsource of 2nd Web Audio Conference paper

Example project (./example_eval/)

- An example of what the set up XML should look like, with example audio files 0.wav-10.wav which are short recordings at 44.1kHz, 16bit of a woman saying the corresponding number (useful for testing randomisation and general familiarisation with the interface).

Output files (./saves/)

- The output XML files of tests will be stored here by default by the `pythonServer.py` script.

Auxiliary scripts (./scripts/)

- Helpful Python scripts for extraction and visualisation of data.

Test creation tool (./test_create/)

- Webpage for easily setting up your own test without having to delve into the XML.

1.2 Compatibility

As Microsoft Internet Explorer doesn’t support the Web Audio API¹, you will need another browser like Google Chrome, Safari or Firefox (all three are tested and confirmed to work).

¹<http://caniuse.com/#feat=audio-api>

Firefox does not currently support other bit depths than 8 or 16 bit for PCM wave files. In the future, this will throw a warning message to tell the user that their content is being quantised automatically.

The tool is platform-independent and works in any browser that supports the Web Audio API. It does not require any specific, proprietary software. However, in case the tool is hosted locally (i.e. you are not hosting it on an actual webserver) you will need Python (2.7), which is a free programming language - see the next paragraph.

2 Test setup

2.1 Sample rate

Depending on how the experiment is set up, audio is resampled automatically (the Web Audio default) or the sample rate is enforced. In the latter case, you will need to make sure that the sample rate of the system is equal to the sample rate of these audio files. For this reason, all audio files in the experiment will have to have the same sample rate.

Always make sure that all other digital equipment in the playback chain (clock, audio interface, digital-to-analog converter, ...) is set to this same sample rate.

Note that upon changing the sampling rate, the browser will have to be restarted for the change to take effect.

2.1.1 Mac OS X

To change the sample rate in Mac OS X, go to **Applications/Utilities/Audio MIDI Setup** or find this application with Spotlight (see Figure 1). Then select the output of the audio interface you are using and change the ‘Format’ to the appropriate number. Also make sure the bit depth and channel count are as desired. If you are using an external audio interface, you may have to go to the preference pane of that device to change the sample rate.

Also make sure left and right channel gains are equal, as some applications alter this without changing it back, leading to a predominantly louder left or right channel. See Figure 1 for an example where the channel gains are different.

2.1.2 Windows

To change the sample rate in Windows, right-click on the speaker icon in the lower-right corner of your desktop and choose ‘Playback devices’. Right-click the appropriate playback device and click ‘Properties’. Click the ‘Advanced’ tab and verify or change the sample rate under ‘Default Format’. If you are using an external audio interface, you may have to go to the preference pane of that device to change the sample rate.

2.2 Local test

If the test is hosted locally, you will need to run the local webserver provided with this tool.

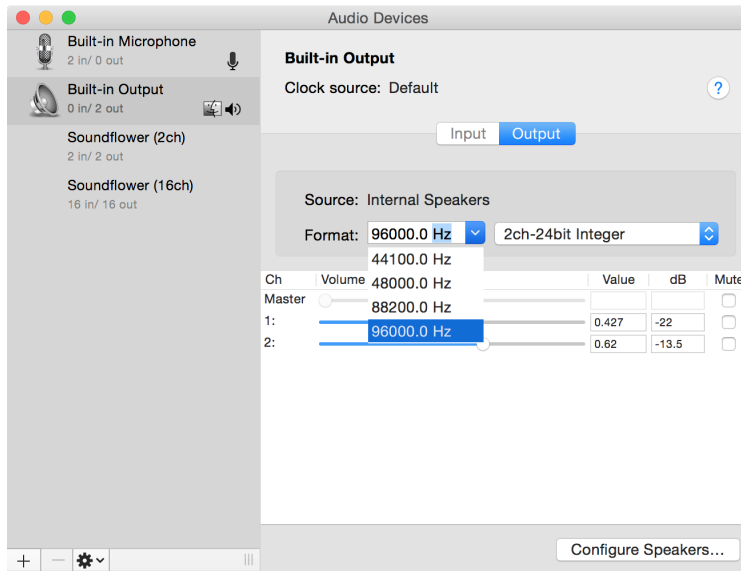


Figure 1: The Audio MIDI Setup window in Mac OS X

2.2.1 Mac OS X

On Mac OS X, Python comes preinstalled.

Open the Terminal (find it in **Applications/Terminal** or via Spotlight), and go to the folder you downloaded. To do this, type `cd [folder]`, where `[folder]` is the folder where to find the `pythonServer.py` script you downloaded. For instance, if the location is `/Users/John/Documents/test/`, then type

```
cd /Users/John/Documents/test/
```

Then hit enter and run the Python script by typing

```
python pythonServer.py
```

and hit enter again. See also Figure 2.

Alternatively, you can simply type `python` (followed by a space) and drag the file into the Terminal window from Finder.

You can leave this running throughout the different experiments (i.e. leave the Terminal open).

To start the test, open the browser and type

```
localhost:8000
```

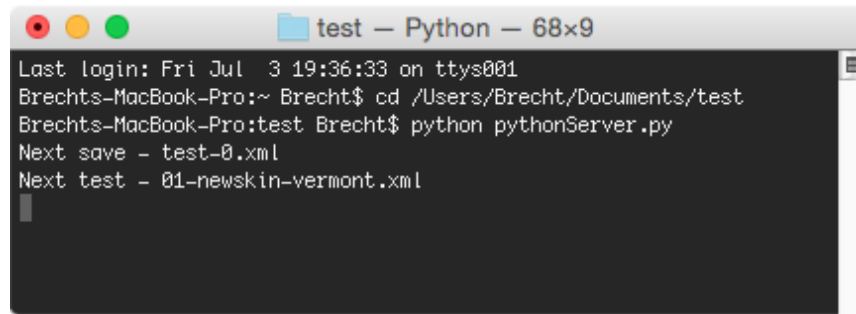



Figure 2: Mac OS X: The Terminal window after going to the right folder (`cd [folder_path]`) and running `pythonServer.py`.

and hit enter. The test should start (see Figure 5).

To quit the server, either close the terminal window or press `Ctrl+C` on your keyboard to forcibly shut the server.

2.2.2 Windows

On Windows, Python 2.7 is not generally preinstalled and therefore has to be downloaded² and installed to be able to run scripts such as the local webserver, necessary if the tool is hosted locally.

Simply double click the Python script `pythonServer.py` in the folder you downloaded.

You may see a warning like the one in Figure 3. Click ‘Allow access’.

The process should now start, in the Command prompt that opens - see Figure 4.

You can leave this running throughout the different experiments (i.e. leave the Command Prompt open).

To start the test, open the browser and type

`localhost:8000`

and hit enter. The test should start (see Figure 5).

If at any point in the test the participant reports weird behaviour or an error of some kind, or the test needs to be interrupted, please notify the experimenter and/or refer to Section 7.

²<https://www.python.org/downloads/windows/>



Figure 3: Windows: Potential warning message when executing `pythonServer.py`.

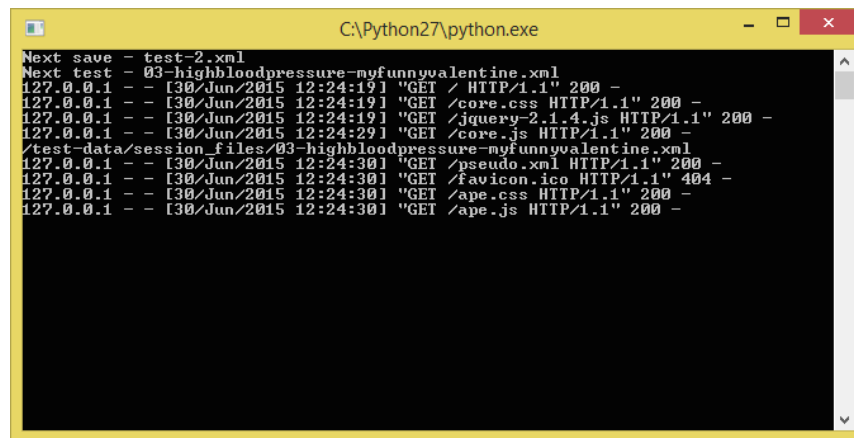


Figure 4: Windows: The Command Prompt after running `pythonServer.py` and opening the corresponding website.

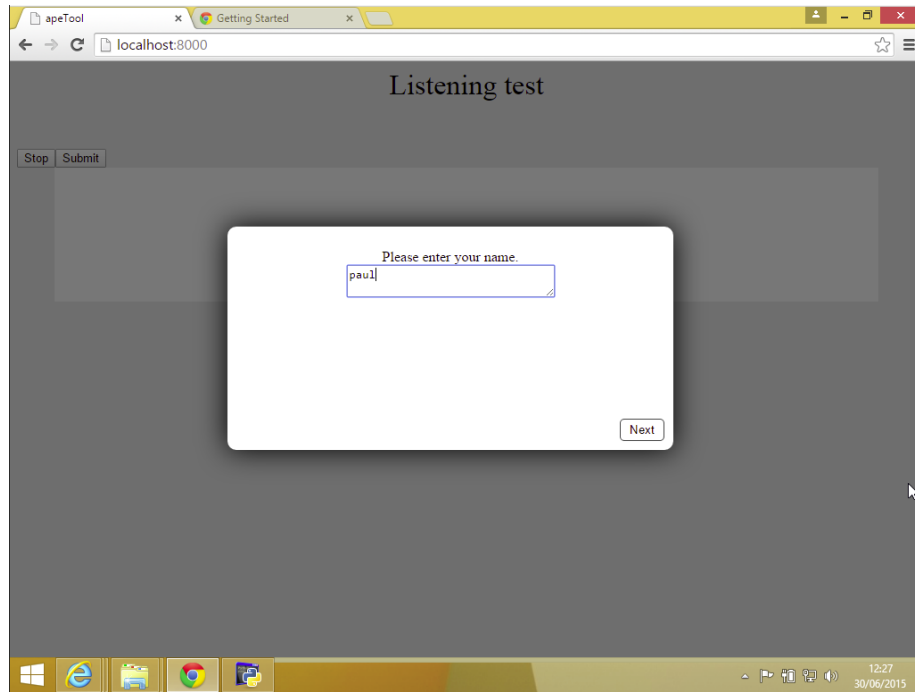


Figure 5: The start of the test in Google Chrome on Windows 7.

When the test is over (the subject should see a message to that effect, and click ‘Submit’ one last time), the output XML file containing all collected data should have appeared in ‘saves/'. The names of these files are ‘test-0.xml’, ‘test-1.xml’, etc., in ascending order. The Terminal or Command prompt running the local web server will display the following file name. If such a file did not appear, please again refer to Section 7.

It is advised that you back up these results as often as possible, as a loss of this data means that the time and effort spent by the subject(s) has been in vain. Save the results to an external or network drive, and/or send them to the experimenter regularly.

To start the test again for a new participant, you do not need to close the browser or shut down the Terminal or Command Prompt. Simply refresh the page or go to `localhost:8000` again.

2.3 Remote test

Put all files on a web server which supports PHP. This allows the ‘save.php’ script to store the XML result files in the ‘saves/’ folder. If the web server is not able to store the XML

file there at the end of the test, it will present the XML file locally to the user, as a ‘Save file’ link.

Make sure the `projectReturn` attribute of the `setup` node is set to the `save.php` script.

Then, just go to the URL of the corresponding HTML file, e.g. `http://server.com/path/to/WAET/index.htm`. If storing on the server doesn’t work at submission (e.g. if the `projectReturn` attribute isn’t properly set), the result XML file will be presented to the subject on the client side, as a ‘Save file’ link.

3 Interfaces

The Web Audio Evaluation Tool comes with a number of interface styles, each of which can be customised extensively, either by configuring them differently using the many optional features, or by modifying the JavaScript files.

To set the interface style for the whole test, add `interface="APE"` to the `setup` node, where "APE" is one of the interface names below.

3.1 APE

The APE interface is based on [2], and consists of one or more axes, each corresponding with an attribute to be rated, on which markers are placed. As such, it is a multiple stimulus interface where (for each dimension or attribute) all elements are on one axis so that they can be maximally compared against each other, as opposed to rated individually or with regards to a single reference. It also contains an optional text box for each element, to allow for clarification by the subject, tagging, and so on.

3.2 MUSHRA

This is a straightforward implementation of [3], especially common for the rating of audio quality, for instance for the evaluation of audio codecs.

4 Features

This section goes over the different features implemented in the Web Audio Evaluation Tool, how to use them, and what to know about them.

Unless otherwise specified, *each* feature described here is optional, i.e. it can be enabled or disabled and adjusted to some extent.

As the example project showcases (nearly) all of these features, please refer to its configuration XML document for a demonstration of how to enable and adjust them.

4.1 Surveys

4.1.1 Pre- and post-page surveys

4.1.2 Pre- and post-test surveys

4.1.3 Survey elements

All survey elements (which ‘pop up’ in the centre of the browser) have an `id` attribute, for retrieval of the responses in post-processing of the results, and a `mandatory` attribute, which if set to “true” requires the subjects to respond before they can continue.

statement Simply shows text to the subject until ‘Next’ or ‘Start’ is clicked.

question Expects a text answer (in a text box). Has the `boxsize` argument: set to “large” or “huge” for a bigger box size.

number Expects a numerical value. Attribute `min="0"` specifies the minimum value - in this case the answer must be strictly positive before the subject can continue.

radio Radio buttons.

checkbox Checkboxes. Note that when making a checkbox question “mandatory”, the subject is forced to select at least one option (which could be e.g. ‘Other’ or ‘None’).

Example usage:

```
<PostTest>
  <question id="location" mandatory="true" boxsize="large">Please enter your location. (example
    mandatory text question)</question>
  <number id="age" min="0">Please enter your age (example non-mandatory number question)</
    number>
  <radio id="rating">
```

```

        <statement>Please rate this interface (example radio button question)</statement>
        <option name="bad">Bad</option>
        <option name="poor">Poor</option>
        <option name="good">Good</option>
        <option name="great">Great</option>
    </radio>
    <checkbox id="checkboxtest" mandatory="true">
        <statement>Please select with which activities you have any experience (
            example checkbox question)</statement>
        <option name="musician">Playing a musical instrument</option>
        <option name="soundengineer">Recording or mixing audio</option>
    </checkbox>
    <statement>Thank you for taking this listening test. Please click 'Submit' and your results
        will appear in the 'saves/' folder.</statement>
</PostTest>

```

4.2 Randomisation

4.2.1 Randomisation of configuration XML files

4.2.2 Randomisation of page order

4.2.3 Randomisation of axis order

4.2.4 Randomisation of fragment order

4.2.5 Randomisation of initial slider position

4.3 Looping

Loops the fragments If the fragments are not of equal length initially, they are padded with zeros so that they are equal length, to enable looping without the fragments going out of sync relative to each other.

Note that fragments cannot be played until all fragments are loaded when in looped mode, as the engine needs to assess whether all

4.4 Sample rate

If you require the test to be conducted at a certain sample rate (i.e. you do not tolerate resampling of the elements to correspond with the system's sample rate), add `sampleRate="96000"` - where "96000" can be any support sample rate - so that a warning message is shown alerting the subject the system's sample rate is different from this enforced sample rate. This of course means that in one test, all sample rates must be equal as it is impossible to change the system's sample rates during the test (even if you were to manually change it, then the browser must be restarted for it to take effect).

4.5 Scrubber bar

The scrubber bar, or transport bar (that is the name of the visualisation of the playhead thing with an indication of time and showing the portion of the file played so far) is at this point just a visual, and not a controller to adjust

the playhead position.

Make visible by adding `<option name='playhead' />` to the `interface` node (see Section 4.8: Checks).

4.6 Metrics

Enable the collection of metrics by adding `collectMetrics='true'` in the `setup` node.

The **Metric** node, which contains the metrics to be tracked during the complete test, is a child of the `setup` node, and it could look as follows.

```
<Metric>
  <metricEnable>testTimer</metricEnable>
  <metricEnable>elementTimer</metricEnable>
  <metricEnable>elementInitialPosition</metricEnable>
  <metricEnable>elementTracker</metricEnable>
  <metricEnable>elementFlagListenedTo</metricEnable>
  <metricEnable>elementFlagMoved</metricEnable>
  <metricEnable>elementListenTracker</metricEnable>
</Metric>
```

When in doubt, err on the inclusive side, as one never knows which information is needed in the future. Most of these metrics are necessary for post-processing scripts such as `timeline_view_movement.py`.

4.6.1 Time test duration

`testTimer`

4.6.2 Time fragment playback

`elementTimer`

This keeps track of when each fragment is played back and stopped again, *and* which part of the fragment has been played back at that time.

4.6.3 Initial positions

`elementInitialPosition`

Tracks the initial position of the sliders, especially relevant when these are randomised, so their influence

4.6.4 Track movements

4.6.5 Which fragments listened to

4.6.6 Which fragments moved

Binary check whether or not a the marker corresponding with a particular fragment was moved at all throughout the experiment.

4.6.7 elementListenTracker

4.7 References and anchors

4.7.1 Reference

4.7.2 Hidden reference

4.7.3 Hidden anchor

4.8 Checks

These checks are enabled in the `interface` node, which is a child of the `setup` node.

4.8.1 Playback checks

Enforce playing each sample at least once, for at least a little bit (e.g. this test is satisfied even if you only play a tiny portion of the file), by alerting the user to which samples have not been played upon clicking ‘Submit’. When enabled, one cannot proceed to the next page, answer a survey question, or finish the test, before clicking each sample at least once.

Alternatively, one can check whether the *entire* fragment was listened to at least once.

Add `<check name="fragmentPlayed"/>` to the `interface` node.

4.8.2 Movement check

Enforce moving each sample at least once, for at least a little bit (e.g. this test is satisfied even if you only play a tiny portion of the file), by alerting the user to which samples have not been played upon clicking ‘Submit’. When enabled, one cannot proceed to the next page, answer a survey question, or finish the test, before clicking each sample at least once. If there are several axes, the warning will specify which samples have to be moved on which axis.

Add `<check name="fragmentMoved"/>` to the `interface` node.

4.8.3 Comment check

Enforce commenting, by alerting the user to which samples have not been commented on upon clicking ‘Submit’. When enabled, one cannot proceed to the next page, answer a survey question, or finish the test, before putting at least one character in each comment box.

Note that this does not apply to any extra (text, radio button, checkbox) elements, unless these have the ‘mandatory’ option enabled.

Add `<check name="fragmentComments"/>` to the `interface` node.

4.8.4 Scale use check

It is possible to enforce a certain usage of the scale, meaning that at least one slider needs to be below and/or above a certain percentage of the slider.

Add `<check name="scalerrange" min="25" max="75"/>` to the `interface` node.

4.8.5 Note on the use of multiple rating axes

I.e. what if more than one axis? How to specify which axis the checks relate to?

4.9 Layout options

`title`, `scale`, `position`, `commentBoxPrefix`

4.10 Multiple sliders

(APE example)

```
<interface name="preference">
  <title>Preference</title>
  <scale position="0">Min</scale>
  <scale position="100">Max</scale>
  <scale position="50">Middle</scale>
  <commentBoxPrefix>Comment on fragment</commentBoxPrefix>
</interface>
<interface name="depth">
  <title>Depth</title>
  <scale position="0">Low</scale>
  <scale position="100">High</scale>
  <scale position="50">Middle</scale>
  <commentBoxPrefix>Comment on fragment</commentBoxPrefix>
</interface>
```

where the `interface` nodes are children of the `audioholder` node.

4.11 Platform information

For troubleshooting and usage statistics purposes, information about the browser and the operating system is logged in the results XML file. This is especially useful in the case of remote tests, when it is not certain which operating system, browser and/or browser were used. Note that this information is not always available and/or accurate, e.g. when the subject has taken steps to be more anonymous, so it should be treated as a guide only.

Example:

```
<navigator>
  <platform>MacIntel</platform>
  <vendor>Google Inc.</vendor>
  <uagent>Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_5) AppleWebKit/537.36 (KHTML, like Gecko)
    Chrome/43.0.2357.134 Safari/537.36</uagent>
</navigator>
```

4.12 Show progress

Add `<option name="page-count"/>` to the `interface` node (see Section 4.8: Checks) to add the current page number and the total number of pages to the interface.

4.13 Gain

It is possible to set the gain (in decibel) applied to the different audioelements, as an attribute of the `audioelement` nodes in the configuration XML file:

```
<audioElements url="sample-01.wav" gain="-6" id="sample01quieter"/>
```

4.14 Loudness

Set the loudness for a complete test by adding `loudness="-23"` to the `setup` node in the configuration XML file, where -23 is an example loudness in LUFS.

5 Using the test create tool

We provide a test creation tool, available in the directory `test.create`. This tool is a self-contained web page, so doubling clicking will launch the page in your system default browser.

The test creation tool can help you build a simple test very quickly. By simply selecting your interface and clicking check-boxes you can build a test in minutes.

Include audio by dragging and dropping the stimuli you wish to include.

The tool examines your XML before exporting to ensure you do not export an invalid XML structure which would crash the test.

This guide will help you to construct your own interface on top of the WAET (Web Audio Evaluation Tool) engine. The WAET engine resides in the `core.js` file, this contains prototype objects to handle most of the test creation, operation and data collection. The interface simply has to link into this at the correct points.

5.1 Nodes to familiarise

`Core.js` handles several very important nodes which you should become familiar with. The first is the Audio Engine, initialised and stored in variable `'AudioEngineContext'`. This handles the playback of the web audio nodes as well as storing the `'AudioObjects'`. The `'AudioObjects'` are custom nodes which hold the audio fragments for playback. These nodes also have a link to two interface objects, the comment box if enabled and the interface providing the ranking. On creation of an `'AudioObject'` the interface link will be nulled, it is up to the interface to link these correctly.

The specification document will be decoded and parsed into an object called `'specification'`. This will hold all of the specifications various nodes. The test pages and any pre/post test objects are processed by a test state which will proceed through the test when called to by the interface. Any checks (such as playback or movement checks) are to be completed by the interface before instructing the test state to proceed. The test state will call the interface on each page load with the page specification node.

5.2 Modifying `core.js`

Whilst there is very little code actually needed, you do need to instruct `core.js` to load your interface file when called for from a specification node. There is a function called `'loadProjectSpecCallback'` which handles the decoding of the specification and setting any external items (such as metric collection). At the very end of this function there is an if statement, add to this list with your interface string to link to the source. There is an example in there for both the APE and MUSHRA tests already included. Note: Any updates to `core.js` in future work will most likely overwrite your changes to this file, so remember to check your interface is still here after any update that interferes with `core.js`. Any further files can be loaded here as well, such as css styling files. `jQuery` is already included.

5.3 Building the Interface

Your interface file will get loaded automatically when the `'interface'` attribute of the setup node matches the string in the `'loadProjectSpecCallback'` function. The following functions must be defined in your interface file.

- **loadInterface** - Called once when the document is parsed. This creates any necessary bindings, such as to the metric collection classes and any check commands. Here you can also start the structure for your test such as placing in any common nodes (such as the title and empty divs to drop content into later).
- **loadTest(audioHolderObject)** - Called for each page load. The `audioHolderObject` contains a specification node holding effectively one of the `audioHolder` nodes.

- **resizeWindow(event)** - Handle for any window resizing. Simply scale your interface accordingly. This function must be here, but can be an empty function call.

5.3.1 loadInterface

This function is called by the interface once the document has been parsed since some browsers may parse files asynchronously. The best method is simply to put 'loadInterface()' at the top of your interface file, therefore when the JavaScript engine is ready the function is called.

By default the HTML file has an element with id "topLevelBody" where you can build your interface. Make sure you blank the contents of that object. This function is the perfect time to build any fixed items, such as the page title, session titles, interface buttons (Start, Stop, Submit) and any holding and structure elements for later on.

At the end of the function, insert these two function calls: `testState.initialise()` and `testState.advanceState()`; This will actually begin the test sequence, including the pre-test options (if any are included in the specification document).

5.3.2 loadTest(audioHolderObject)

This function is called on each new test page. It is this functions job to clear out the previous test and set up the new page. Use the function `audioEngineContext.newTestPage()`; to instruct the audio engine to prepare for a new page. "audioEngineContext.audioObjects = [];" will delete any audioObjects, `interfaceContext.deleteCommentBoxes()`; will delete any comment boxes and `interfaceContext.deleteCommentQuestions()`; will delete any extra comment boxes specified by commentQuestion nodes.

This function will need to instruct the audio engine to build each fragment. Just passing the constructor each element from the audioHolderObject will build the track, `audioEngineContext.newTrack(element)` (where element is the audioHolderObject audio element). This will return a reference to the constructed audioObject. Decoding of the audio will happen asynchronously.

You also need to link `audioObject.interfaceDOM` with your interface object for that audioObject. The interfaceDOM object has a few default methods. Firstly it must start disabled and become enabled once the audioObject has decoded the audio (function call: `enable()`). Next it must have a function `exportXMLDOM()`, this will return the xml node for your interface, however the default is for it to return a value node, with `textContent` equal to the normalised value. You can perform other functions, but our scripts may not work if something different is specified (as it will breach our results specifications). Finally it must also have a method `getValue`, which returns the normalised value.

It is also the job the interfaceDOM to call any metric collection functions necessary, however some functions may be better placed outside (for example, the APE interface uses drag and drop, therefore the best way was to call the metric functions from the `dragEnd` function, which is called when the interface object is dropped). Metrics based upon listening are handled by the audioObject. The interfaceDOM object must manage any movement metrics. For a list of valid metrics and their behaviours, look at the project specification document included in the repository/docs location. The same goes for any checks required when pressing the submit button, or any other method to proceed the test state.

6 Analysis and diagnostics

6.1 In the browser

See ‘analysis.html’ in the main folder: immediate visualisation of (by default) all results in the ‘saves/’ folder.

6.2 Python scripts

The package includes Python (2.7) scripts (in ‘scripts/’) to extract ratings and comments, generate visualisations of ratings and timelines, and produce a fully fledged report.

Visualisation requires the free matplotlib toolbox (<http://matplotlib.org>), numpy and scipy. By default, the scripts can be run from the ‘scripts’ folder, with the result files in the ‘saves’ folder (the default location where result XMLs are stored). Each script takes the XML file folder as an argument, along with other arguments in some cases. Note: to avoid all kinds of problems, please avoid using spaces in file and folder names (this may work on some systems, but others don’t like it).

6.2.1 comment_parser.py

Extracts comments from the output XML files corresponding with the different subjects found in ‘saves/’. It creates a folder per ‘audioholder’/page it finds, and stores a CSV file with comments for every ‘audioelement’/fragment within these respective ‘audioholders’/pages. In this CSV file, every line corresponds with a subject/output XML file. Depending on the settings, the first column containing the name of the corresponding XML file can be omitted (for anonymisation). Beware of Excel: sometimes the UTF-8 is not properly imported, leading to problems with special characters in the comments (particularly cumbersome for foreign languages).

6.2.2 evaluation_stats.py

Shows a few statistics of tests in the ‘saves/’ folder so far, mainly for checking for errors. Shows the number of files that are there, the audioholder IDs that were tested (and how many of each separate ID), the duration of each page, the duration of each complete test, the average duration per page, and the average duration in function of the page number.

6.2.3 generate_report.py

Similar to ‘evaluation_stats.py’, but generates a PDF report based on the output files in the ‘saves/’ folder - or any folder specified as command line argument. Uses pdflatex to write a LaTeX document, then convert to a PDF.

6.2.4 score_parser.py

Extracts rating values from the XML to CSV - necessary for running visualisation of ratings. Creates the folder ‘saves/ratings/’ if not yet created, to which it writes a separate file for every ‘audioholder’/page in any of the output XMLs it finds in ‘saves/’. Within each file, rows represent different subjects (output XML file names) and columns represent different ‘audioelements’/fragments.

6.2.5 score_plot.py

Plots the ratings as stored in the CSVs created by score_parser.py. Depending on the settings, it displays and/or saves (in 'saves/ratings/') a boxplot, confidence interval plot, scatter plot, or a combination of the aforementioned. Requires the free matplotlib library. At this point, more than one subjects are needed for this script to work.

6.2.6 timeline_view_movement.py

Creates a timeline for every subject, for every 'audioholder'/page, corresponding with any of the output XML files found in 'saves/'. It shows the marker movements of the different fragments, along with when each fragment was played (red regions). Automatically takes fragment names, rating axis title, rating axis labels, and audioholder name from the XML file (if available).

6.2.7 timeline_view.py

Creates a timeline for every subject, for every 'audioholder'/page, corresponding with any of the output XML files found in 'saves/'. It shows when and for how long the subject listened to each of the fragments.

7 Troubleshooting

7.1 Reporting bugs and requesting features

Thanks to feedback from using the interface in experiments by the authors and others, many bugs have been caught and fatal crashes due to the interface seem to be a thing of the past entirely.

We continually develop this tool to fix issues and implement features useful to us or our user base. See <https://code.soundsoftware.ac.uk/projects/webaudioevaluationtool/issues> for a list of feature requests and bug reports, and their status.

Please contact the authors if you experience any bugs, if you would like additional functionality, if you spot any errors or gaps in the documentation, if you have questions about using the interface, or if you would like to give any feedback (even positive!) about the interface. We look forward to learning how the tool has (not) been useful to you.

7.2 First aid

Meanwhile, if things do go wrong or the test needs to be interrupted for whatever reason, all data is not lost. In a normal scenario, the test needs to be completed until the end (the final ‘Submit’), at which point the output XML is stored in the `saves/`. If this stage is not reached, open the JavaScript Console (see below for how to find it) and type

```
createProjectSave()
```

to present the result XML file on the client side, or

```
createProjectSave(specification.projectReturn)
```

to try to store it to the specified location, e.g. the ‘saves/’ folder on the web server or the local machine (on failure the result XML should be presented directly in the web browser instead)

and hit enter. This will open a pop-up window with a hyperlink that reads ‘Save File’; click it and an XML file with results until that point should be stored in your download folder.

Alternatively, a lot of data can be read from the same console, in which the tool prints a lot of debug information. Specifically:

- the randomisation of pages and fragments are logged;
- any time a slider is played, its ID and the time stamp (in seconds since the start of the test) are displayed;
- any time a slider is dragged and dropped, the location where it is dropped including the time stamp are shown;
- any comments and pre- or post-test questions and their answers are logged as well.

You can select all this and save into a text file, so that none of this data is lost. You may choose to do this even when a test was successful as an extra precaution.

If you encounter any issue which you believe to be caused by any aspect of the tool, and/or which the documentation does not mention, please do let us know!

Opening the JavaScript Console

- In Google Chrome, the JavaScript Console can be found in **View>Developer>JavaScript Console**, or via the keyboard shortcut `Cmd + Alt + J` (Mac OS X).

- In Safari, the JavaScript Console can be found in **Develop>Show Error Console**, or via the keyboard shortcut **Cmd + Alt + C** (Mac OS X). Note that for the Developer menu to be visible, you have to go to Preferences (**Cmd + ,**) and enable ‘Show Develop menu in menu bar’ in the ‘Advanced’ tab. **Note that as long as the Developer menu is not visible, nothing is logged to the console, i.e. you will only be able to see diagnostic information from when you switched on the Developer tools onwards.**
- In Firefox, go to **Tools>Web Developer>Web Console**, or hit **Cmd + Alt + K**.

7.3 Known issues and limitations

The following is a non-exhaustive list of problems and limitations you may experience using this tool, due to not being supported yet by us, or by the Web Audio API and/or (some) browsers.

- Issue **#1463**: **Firefox** only supports 8 bit and 16 bit WAV files. Pending automatic requantisation (which deteriorates the audio signal’s dynamic range to some extent), WAV format stimuli need to adhere to these limitations in order for the test to be compatible with Firefox.
- Issues **#1474** and **#1462**: On occasions, audio is not working - or only a continuous ‘beep’ can be heard - notably in **Safari**. Refreshing, quitting the browser and even enabling Developer tools in Safari’s Preferences pane (‘Advanced’ tab: “Show ‘Develop’ menu in menu bar”) has helped resolve this. If no (high quality) audio can be heard, make sure your entire playback system’s settings are all correct.

8 References

- [1] N. Jillings, D. Moffat, B. De Man, and J. D. Reiss, “Web Audio Evaluation Tool: A browser-based listening test environment,” in *12th Sound and Music Computing Conference*, July 2015.
- [2] B. De Man and J. D. Reiss, “APE: Audio Perceptual Evaluation toolbox for MATLAB,” in *136th Convention of the Audio Engineering Society*, April 2014.
- [3] *Method for the subjective assessment of intermediate quality level of coding systems*. Recommendation ITU-R BS.1534-1, 2003.

A Legacy

The APE interface and most of the functionality of the first WAET editions are inspired by the APE toolbox for MATLAB [2]. See <https://code.soundsoftware.ac.uk/projects/ape> for the source code and <http://brechtdeman.com/publications/aes136.pdf> for the corresponding paper.

B Listening test instructions example

Before each test, show the instructions below or similar and make sure it is available to the subject throughout the test. Make sure to ask whether the participant has any questions upon seeing and/or reading the instructions.

- You will be asked for your name (“John Smith”) and location (room identifier).
- An interface will appear, where you are asked to
 - click green markers to play the different mixes;
 - drag the markers on a scale to reflect your preference for the mixes;
 - comment on these mixes, using text boxes with corresponding numbers (in your **native language**);
 - optionally comment on all mixes together, or on the song, in ‘General comments’.
- You are asked for your personal, honest opinion. Feel free to use the full range of the scale to convey your opinion of the various mixes. Don’t be afraid to be harsh and direct.
- The markers appear at random positions at first (which means some markers may hide behind others).
- The interface can take a few seconds to start playback, but switching between mixes should be instantaneous.
- This is a research experiment, so please forgive us if things go wrong. Let us know immediately and we will fix it or restart the test.
- When the test is finished (after all songs have been evaluated), just call the experimenter, do NOT close the window.
- After the test, please fill out our survey about your background, experience and feedback on the test.
- By participating, you consent to us using all collected data for research. Unless asked explicitly, all data will be anonymised when shared.

C Terminology

As a guide to better understand the Instructions, and to expand them later, here is a list of terms that may be unclear or ambiguous unless properly defined.

Subject The word we use for a participant, user, ... of the test, i.e. not the experimenter who designs the test but the person who evaluates the audio under test as part of an experiment (or the preparation of one).

User The person who uses the tool to configure, run and analyse the test - i.e. the experimenter, most likely a researcher - or at least

Page A screen in a test; corresponds with an **audioholder**

Fragment An element or sample in a test; corresponds with an **audioelement**

Test A complete test which can consist of several pages; corresponds with an entire configuration XML file

Configuration XML file The XML file containing the necessary information on interface, samples, survey questions, configurations, ... which the JavaScript modules read to produce the desired test.

Results XML file The output of a successful test, including ratings, comments, survey responses, timing information, and the complete configuration XML file with which the test was generated in the first place.

Contact details

- Nicholas Jillings: `nicholas.jillings@mail.bcu.ac.uk`
- Brecht De Man: `b.deman@qmul.ac.uk`
- David Moffat: `d.j.moffat@qmul.ac.uk`