Web Audio Evaluation Tool something something

Nicholas Jillings
n.g.r.jillings@se14.qmul.ac.uk

Brecht De Man
b.deman@qmul.ac.uk

David Moffat
d.j.moffat@qmul.ac.uk

Joshua D. Reiss
joshua.reiss@qmul.ac.uk

Centre for Digital Music
School of Electronic Engineering and Computer Science
Queen Mary University of London
Mile End Road, London E1 4NS
United Kingdom

ABSTRACT

Here comes the abstract.

1. INTRODUCTION

Introducing the paper. Referring to [5]. Talking about what we do in the various sections of this paper. Pointing out that the header of the paper kind of looks like the Batsign.

2. ARCHITECTURE

A slightly technical overview of the system. Talk about XML, JavaScript, Web Audio API, HTML5. Describe and/or visualise audioholder-audioelement-... structure.

Streaming audio?

3. REMOTE TESTS

The following features allow easy and effective remote testing:

- PHP script to collect result XML files
- Randomly pick specified number of audioholders
- Functionality to participate multiple times
  - Possible to log in with unique ID (no password)
  - Pick 'new user' (need new, unique ID) or 'already participated' (need already available ID)
  - Store XML on server with IDs plus which audioholders have already been listened to
  - Don't show 'post-test' survey after first time
  - Pick 'new' audioholders if available
  - Copy survey information first time to new XMLs
- Intermediate saves
- Collect IP address information (privacy issues?) –> geo-related API?

4. INTERFACES

We could add more interfaces, such as:

- Multi attribute ratings
- MUSHRA (ITU-R BS. 1534) [13]
- Interval Scale [14]
- Rank Scale [8]
- 2D Plane rating - e.g. Valence vs. Arousal [1]
- Likert scale [6]
- **All the following are the interfaces available in HULTI-GEN [4]**
- ABC/HR (ITU-R BS. 1116) [12]
  - Continuous Scale (5-1) Imperceptible, Perceptible but not annoying, slightly annoying, annoying, very annoying. (default Inaudible?)
- -50 to 50 Bipolar with Ref
  - Scale -50 to 50 on Mushra with default values as 0 in middle and a comparison "Reference" to compare to 0 value
- Absolute Category Rating (ACR) Scale [10]
  - 5 point Scale - Bad, Poor, Fair, Good, Excellent (Default fair?)
- Degredation Category Rating (DCR) Scale [10]
  - 5 point Scale - Inaudible, Audible but not annoying, slightly annoying, annoying, very annoying. (default Inaudible?) - *Basically just quantised ABC/HR?*
- Comparison Category Rating (CCR) Scale [10]

Web Audio Conference WAC-2016, April 4–6, 2016, Atlanta, USA

  - 7 point scale: Much Better, Better, Slightly Better, About the same, slightly worse, worse, much worse - Default about the same with reference to compare to

- 9 Point Hedonic Category Rating Scale [9]
  - 9 point scale: Like Extremely, Like Very Much, Like Moderate, Like Slightly, Neither Like nor Dislike, dislike Extremely, dislike Very Much, dislike Moderate, dislike Slightly - Default Neither Like nor Dislike with reference to compare to

- ITU-R 5 Point Continuous Impairment Scale [11]
  - 5 point Scale (5-1) Imperceptible, Perceptible but not annoying, slightly annoying, annoying, very annoying. (default Inaudible?)- *Basically just quantised ABC/HR, or Different named DCR*

- Pairwise Comparison (Better/Worse) [3]
  - 2 point Scale - Better or Worse - (not sure how to default this - they default everything to better, which is an interesting choice)

There are also the following interfaces, which would require a slightly different 'engine' underneath, e.g. loading a different page for every possible pair.

- AB Test [7]
- ABX Test [2]
- JND

A screenshot would be nice.

5. ANALYSIS AND DIAGNOSTICS

It would be great to have easy-to-use analysis tools to visualise the collected data and even do science with it. Even better would be to have all this in the browser. Complete perfection would be achieved if and when only limited setup, installation time, and expertise are required for the average non-CS researcher to use this.

The following could be nice:

- Web page showing all audioholder IDs, file names, subject IDs, audio element IDs, ... in the collected XMLs so far (`saves/*.xml`)
- Check/uncheck each of the above for analysis (e.g. zoom in on a certain song, or exclude a subset of subjects)
- Click a mix to hear it (follow path in XML setup file, which is also embedded in the XML result file)
- Box plot, confidence plot, scatter plot of values (for a given audioholder)
- Timeline for a specific subject (see Python scripts), perhaps re-playing the experiment in X times realtime. (If actual realtime, you could replay the audio...)
- Distribution plots of any radio button and number questions (drop-down menu with 'pretest', 'posttest', ...; then drop-down menu with question 'IDs' like 'gender', 'age', ...; make pie chart/histogram of these values over selected range of XMLs)
- All 'comments' on a specific audioelement
- A 'download' button for a nice CSV of various things (values, survey responses, comments) people might want to use for analysis, e.g. when XML scares them
- Validation of setup XMLs (easily spot 'errors', like duplicate IDs or URLs, missing/dangling tags, ...)

A subset of the above would already be nice for this paper. Some pictures here please.

6. CONCLUDING REMARKS

Perhaps an 'engineering brief' such as this one doesn't really have a lot of conclusion, except 'We made this'.

You can check it out at code.soundsoftware.ac.uk/projects/webaudioevaluationtool.

7. FUTURE WORK

Perhaps here, perhaps not. Talking a little bit about what else might happen. Unless we really want to wrap this up.

8. REFERENCES

[1] J. D. Carroll. *Individual differences and multidimensional scaling.* Bell Telephone Labs., 1969.

[2] D. Clark. High-resolution subjective testing using a double-blind comparator. *Journal of the Audio Engineering Society*, 30(5):330–338, 1982.

[3] H. A. David. *The method of paired comparisons*, volume 12. DTIC Document, 1963.

[4] C. Gribben and H. Lee. Toward the development of a universal listening test interface generator in max. In *Audio Engineering Society Convention 138*. Audio Engineering Society, 2015.

[5] 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.

[6] R. Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.

[7] S. P. Lipshitz and J. Vanderkooy. The great debate: Subjective evaluation. *Journal of the Audio Engineering Society*, 29(7/8):482–491, 1981.

[8] G. C. Pascoe and C. C. Attkisson. The evaluation ranking scale: a new methodology for assessing satisfaction. *Evaluation and program planning*, 6(3):335–347, 1983.

[9] D. R. Peryam and N. F. Girardot. Advanced taste-test method. *Food Engineering*, 24(7):58–61, 1952.

[10] I. Rec. P. 800: Methods for subjective determination of transmission quality. *International Telecommunication Union, Geneva*, 1996.

[11] I. Rec. Bs. 562-3,'subjective assessment of sound quality'. *International Telecommunications Union*, 1997.

[12] I. Recommendation. 1116-1: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. *International Telecommunication Union, Geneva*, 1997.

[13] I. Recommendation. Bs. 1534-1: Method for the subjective assessment of intermediate quality levels of coding systems. *International Telecommunication Union*, 2003.

[14] N. Zacharov, J. Huopaniemi, and M. Hämäläinen. Round robin subjective evaluation of virtual home theatre sound systems at the aes 16th international conference. In *Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction*. Audio Engineering Society, 1999.