Scale

A Software Tool for Listening Experiments

Arnau Vazquez Giner[1]
[1] *Cologne University of Applied Sciences, Institute of Communication Systems*
arnau.vazquez@fh-koeln.de

Introduction

Scale is a software tool that covers the full chain of setup, conduction and analysis of psychoacoustic experiments. Several testing procedures are offered. In the first release, the software includes adaptive methods using forced choice paradigms, semantic differential, double-blind triple-stimulus with hidden reference and rating questions. The conduction part is self-explaining and fully operated by the subjects. The analysis module offers basic comparisons and averaging of results. The software is intended to be easy to understand and to use for both researcher and subjects. The interaction between researcher or subjects and the software is done via a Graphical User Interface (GUI) and does not require any programming skills. Created tests setups or results can be easily ported from one instance of the program to another. Thus everything is portable and exchangeable between different computers and researchers.

Running Environment

Scale is a MATLAB compiled application and requires the MATLAB Compiler Runtime (MCR) to be executed. The MCR is freely available and comes with the download package.
Distributing a compiled version of the software instead of sources has the following advantages:

- The software can be used without programming skills (besides possibly required skills that are needed for the preparation of the stiumuli)
- No MATLAB instance and license are required
- More simplicity due to a single executable only (The source code structure is quite complex and involves a big amount of interdependent functions and files)

The Interface

Scale is divided into three main modules, i.e. setup, conduction and analysis of listening experiments. The three modules enable to automate larger parts of the experiment and thus save the researcher time and effort. The options and the appearance of the user interface depend on the role of the user, i.e. subject or researcher.

Implemented Tests

The amount of available procedures for listening experiments is high and keeps constantly increasing. Therefore the goal of including all existing procedures is unattainable. The first release of *Scale* involves a selection of some popular test procedures like simple or transformed staircase adaptive procedures, semantic differential, double blind triple-stimulus with hidden reference, rating tests and combinations of these.

Adaptive Procedures

Adaptive procedures aim to find a threshold of detection in the psychometric function of a determined dimension of a sound. Stimuli are presented and varied in one dimension. The amount of variation is increased or reduced depending on the preceding subject's responses and on the respective adaptation method. In *Scale* the adaptation is made using different staircase methods like simple staircase (1up/1down) on the one hand, and some transformed staircase methods (1up/2down, 1up/3down, 2up/1down) as described in [5] on the other hand. The threshold estimation can either vary depending on the ending conditions of the trial (limited number of runs or reversals) or on how the average is calculated.

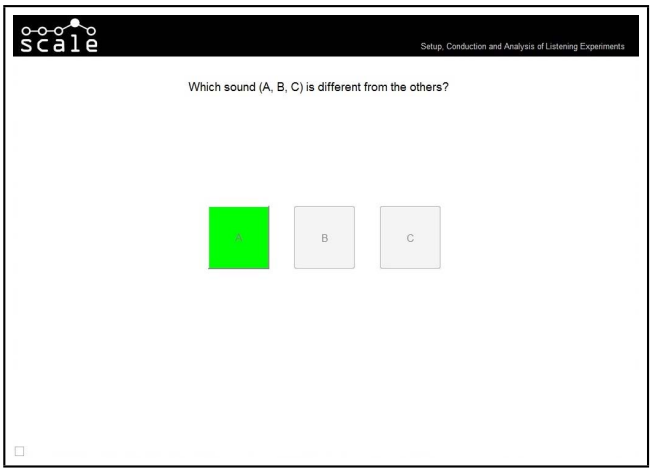

Figure 1: 3-AFC procedure trial window

The tests have to be combined with a paradigm. *Scale* provides different paradigms which can be divided into two groups: n-AFC (n alternative forced choice) paradigms and Yes/No paradigms. In the n-AFC paradigms n intervals (as used in [3]) are presented to the subject. When n is equal to 2 the subject has to decide in which of the two intervals a designated signal is present. When n is greater than 2 the subject has to decide in which of the n intervals the presented stimulus is different. The sample assignation to the intervals of the n-AFC paradigms is always automatically randomised. When using a Yes/No paradigm only one interval that includes one or more sounds is presented to the subject. The subject has to decide whether the signal occurs within the presented interval or not.

Semantic Differential

Semantic differential tests, described in [2], are aimed to obtain a multidimensional aesthetic judgement of sound samples using bipolar rating scales. The items comprise opposing adjective pairs that are assigned to the poles of a scale with a given number of steps, see Figure 2. The task of the subject is to judge and rate the items between the corresponding opposite poles.

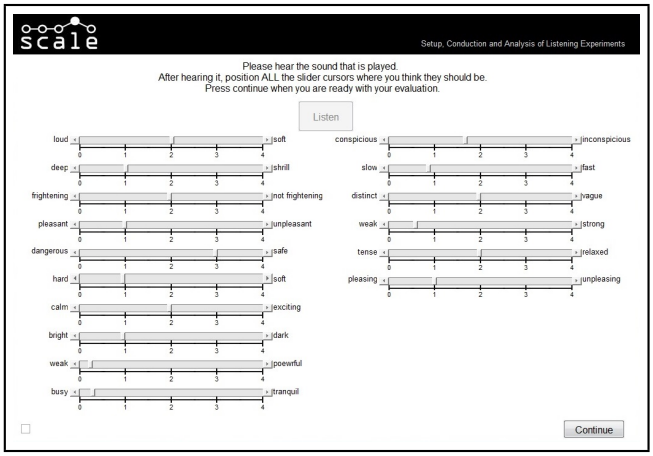

Figure 2: Semantic differential procedure trial window using the pairs of adjectives as described in [1]

All rating scales in *Scale* are continuous and the subjects give their assessments moving a slider cursor to the desired position. The researcher can set the number of steps for each item, the labels of the different marks and the width of the slider cursor. Pictures with different styles of anchors can be selected and placed between slider and labels. A maximum of 20 items and a minimum of three are allowed for every trial.

Double-blind Triple-stimulus with Hidden Reference

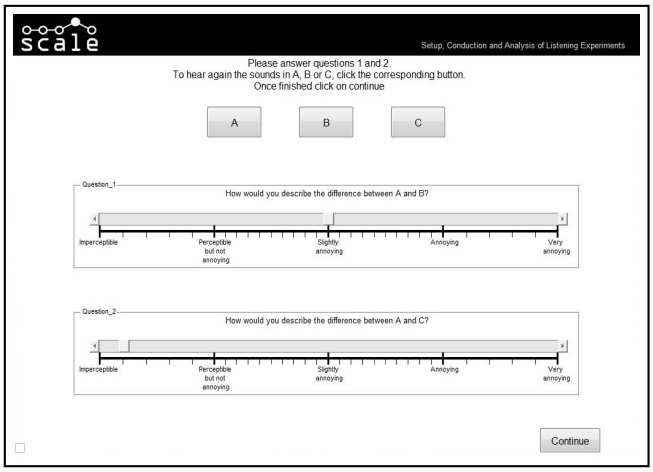

Figure 3: Double-blind triple-stimulus with hidden reference procedure trial window [9]

This procedure has become a standard in psychoacoustics and is used to assess small impairments between sound samples. In every trial, three stimuli are presented in three intervals (”A”, ”B” and ”C”). The stimulus in ”A” is presented as the known reference, the stimuli in ”B” or ”C” are randomly assigned, whereas one of them is a hidden reference and the other one is a sample which is varied in one determined dimension. After listening to the stimuli the subject is asked to assess the impairments between ”A” and ”B” and ”A” and ”C” using a rating scale. The rating is performed with a slider along a continuous scale with anchors. The number of grades in the scale as well as the text in the labels of every mark can be set. Thus all requirements to perform the test *”Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems”* described in the recommendations of the ITU [9] are met, see Figure 3.

Rating Questions

The rating method is useful in case certain opinion of the subject concerning one or more stimuli needs to be obtained. There are two main differences between this method and the semantic differential method. On the one hand the rating scales can have verbal descriptions instead of numerical values and on the other hand every presented item has its own instructions.

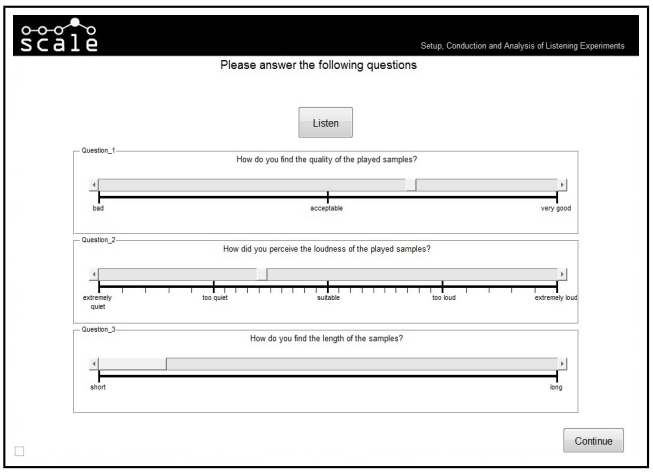

Figure 4: Rating questions procedure trial window

Combined

The combined tests are combinations of an adaptive test with one or more rating questions. The questions are depicted below the buttons in the run windows as shown in Figure 5.

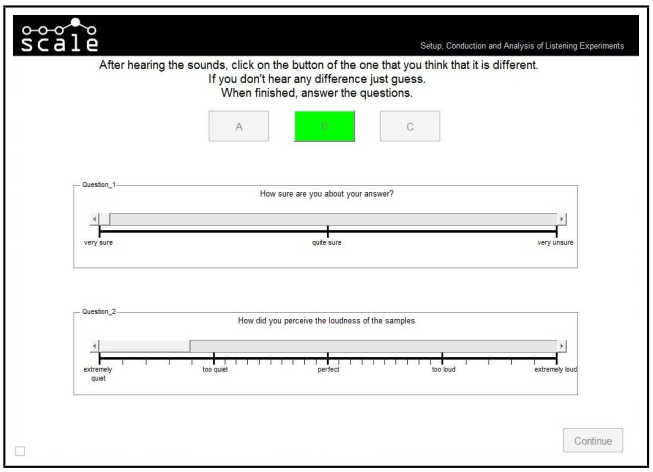

Figure 5: Combined test trial window

Creating Tests

Results from two similar tests can be affected by several factors. On the one hand there are the external factors like the devices used along the playback chain and their settings or the characteristics of the experiment's location. On the other hand there are factors like e.g. the sequence of the presented stimuli or the information given to subjects in advance of, or during the test. Taking the latter into account, different settings are adjustable in *Scale*. The settings are accessed in the creation module where different configuration options are available depending on which type of procedure is selected. Among other features, *Scale* offers the possibility to set example scenarios or to write the instructions appearing in the different phases of the test. The integrated interface for importing the audio samples enables the efficient and comfortable creation of the involved stimuli pool. Furthermore, the stimuli assignment can be automated using batch processing.

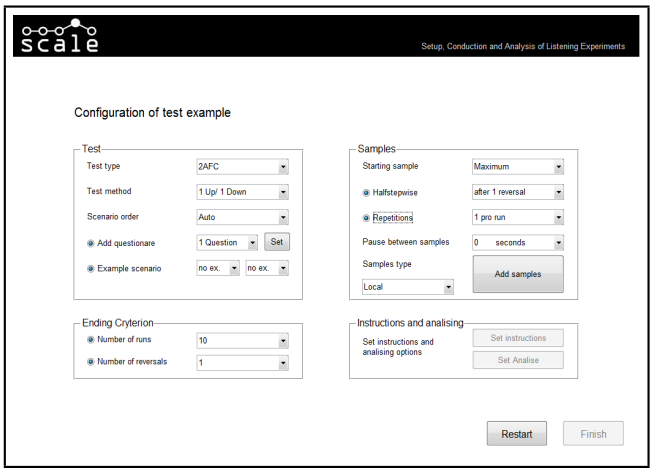

Figure 6: Main create window

Analysis of Results

Results obtained in tests are analysed at the researcher's discretion. Hence, the data can be represented in many different ways. In *Scale* some of the most common methods of representation have been implemented in order to facilitate this process. Those include different averaging options and the calculation of confidence intervals. The analysis module basically offers a graphical representation of the collected data and the researcher can change several properties like scales, labels and grid or line styles. Data and results can be exported to different formats i.e. Raw Text (*.txt*), MATLAB Structs (*.mat*), Excel Tables (*.xls*) and MATLAB Figures (*.fig*). In case a more exhaustive analysis is desired, the full result files can also be exported directly.

Visualisation of Averages

The visualisation of averages is useful to obtain an overview of the results or to give some feedback to the researcher about how a currently running experiment is going. The plots show one averaged value per scenario or trial. It is possible to visualise all available trials or just a dedicated subset of these.

Figure 7 shows an example of an ABX procedure with 12 scenarios containing results from 28 subjects. The Y axis represents the different values obtained in every trial and the X axis represents the trial number. The green line shows an average of all results and the (+) marks denote the values obtained separately for each subject. The blue line is the representation of the confidence interval.

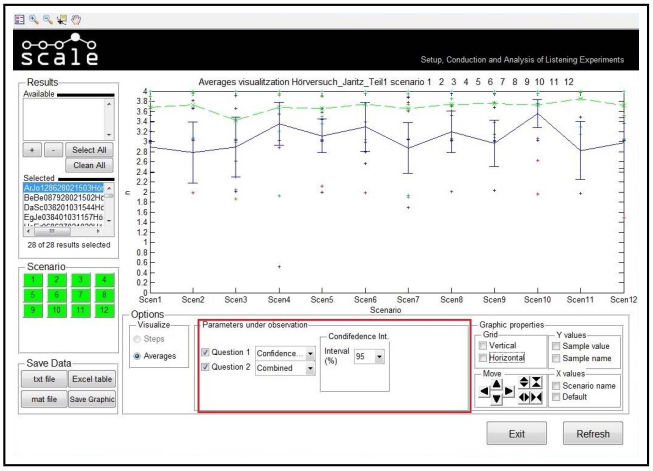

Figure 7: Visualisation of averages from an ABX test

Visualisation of Runs

The visualisation of runs is only available for the adaptive and combined methods since they are the only methods in *Scale* that have more than one step per scenario or trial.

Figure 8 shows an example of one scenario from a 3-AFC test combined with one rating question. The Y axis represents the values obtained in every run and the X axis the run number. The continuous lines represent the response to the AFC paradigm (thick) and the response to the rating question (thin). The dotted line represents the average of the rating question for the whole trial.

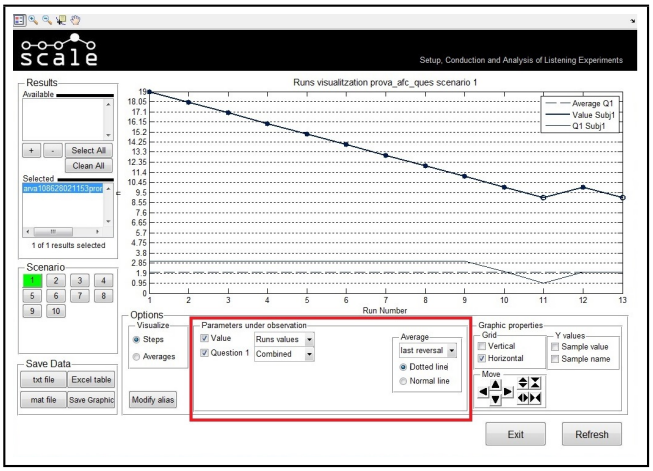

Figure 8: Visualisation of runs from a 3AFC test combined with a rating question

History and Versions

Scale was developed and implemented during 2012-2013 at Cologne University of Applied Sciences as part of the project MARA (Microphone Arrays for Room Acoustics

and Auralization) funded by the Federal Ministry of Education and Research in Germany. Support Code: 17009X11 MARA. Besides for the MARA project, the software is used for several other activities involving research on psychoacoustics at Cologne University of Applied Sciences and is now made available to the acoustic community. The current release includes several basic procedures as presented in this paper. Next versions will include communication interfaces to external reproduction systems and binaural renders, making *Scale* useful in the field of spatial audio research.

Download and Release

The software and a detailed documentation will be released and available for download from April 2013 at: `http://www.audiogroup.web.fh-koeln.de/`.

References

[1] Kuwano S., Namba S., Schick A., Höge H., Fastl H., Filippou T., Florentine M., "Subjective impression of auditory danger signals in different countries", In: Journal Acoustical Society of Japan, Volume 28(5), pp 360-362, 2007.

[2] Osgood C.E., Suci G.J., Tannenbaum P.H., "The measurement of meaning", University of Illinois Press, Urbana, 1967.

[3] Zwicker E., Fastl H., "Psychoacoustics. Facts and Models", 2nd ed. Springer, Heidelberg, New York, 1999, pp. 8-15.

[4] You Y., "Audio Coding: Theory and Applications", Springer, Heidelberg, New York, 2010, pp. 253-254.

[5] Levitt H., "Transformed up Down Methods in Psychoacoustics", In: Doctoral Program in Speech, The City University of New York, 1970.

[6] Fastl H., "Psychoacoustic Basis of Sound Quality Evaluation and Sound Engineering", ICSV-13, Vienna.

[7] Boley J., Lester M., "Statistical Analysis of ABX Results Using Signal Detection Theory", In: Proceedings of the 127th Convention of the AES.

[8] Ciba S., Wlodarski A., Maempel H., "WhisPER - A new tool for performing listening tests", In: Proceedings of the 126th Convention of the AES, Convention Paper 7749.

[9] I.T.U., "Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems", In: Rec. ITU-R BS.1116-1, p. 4.