

The Newcastle Audio Ranking test

User manual

Michael Drinnan
6th January 2012

With great thanks due (in alphabetical order) to:

Meike Brockmann
Paul Carding
Jim Gould
Jessie Waugh

We hope you find NeAR useful.

For the latest version of software, advice, improvements, bug reports or anything else, please
contact me on one of these addresses:

michael.drinnan@ncl.ac.uk
michael.drinnan@nhs.net
michael@drinnan.net

Contents

Installation and preparing the audio files	2
Installation.....	2
Getting the file format right	2
Put together the rating folder.....	2
Optionally, put together a reference folder.....	2
Preparing the rating session.....	3
Select the files for rating	3
If necessary, select your reference files.....	3
Choose a name to identify the rating session	3
Tick the rating options you want to use.....	3
Starting the rating session	4
Performing the rating	5
End the rating session.....	5
Look at the results	5
Finer details	6
The file format.....	6
Using reference files.....	6
How does the program deal with the results file NeAR.csv?	6
Analysing the data.....	7
Measures of agreement.....	7
Unrated samples	7
Using reference files.....	7
References	7

Installation and preparing the audio files

The Newcastle audio ranking test is a mini-application for sorting and rating audio clips by repeated comparison. It is very simple to use.

Installation

You should only need to install the program once, using the enclosed self-installation program. Click on the file *setup.exe* and just follow the instructions. NOTE: The program uses version 3.5 of Microsoft's *.NET* framework. This comes bundled with *Vista*, but might not be installed if you've got an earlier version of Windows.

If the program doesn't run and doesn't offer to help you, the easiest way is to simply type: ".NET framework 3.5" into Google, follow the links and download the .NET framework from Microsoft's website. It is perfectly safe to install, and there is no cost.

You also need Windows Media Player, version 9 or more recent. The chances are that your computer will already have it, but if not you'll need to get it.

Getting the file format right

At the moment, NeAR needs to have your audio files in .WAV format. This is pretty much the standard for MS-Windows audio; if your software doesn't create WAV files already, then it will almost certainly have an option to do so. If not, then there are many software applications that will convert between a bewildering array of different formats. I'd recommend the free audio editor *Audacity* (<http://audacity.sourceforge.net>), which will do the job, and will also let you edit down the audio files to the right size for your rating tasks.

NOTE: If you've got an over-riding reason to need to use wma or mp3 and you really can't convert them to WAVs, then it would be fairly straightforward to adapt the software – I just didn't do it yet. In principle, it ought to be possible to rate any type of file that Media Player can show; video or jpg stills, for example.

Put together the rating folder

Create a folder on your hard disk. Put there a copy of all the WAV files you want to be rated. Any WAV file in this folder will be included in the rating task.

Optionally, put together a reference folder

You only need to do this if you want to rate the samples against some reference files you have already created. Create a different folder on your hard disk. Put there a copy of all the WAV files you want to use as references. They will always be displayed in ascending alphanumeric order: 0...9, then A...Z and a...z.

When you come to do the rating, you won't be able to move the reference files around. Therefore they need to be named correctly from the start. For example, if you've got 12 reference files you could name them: 01.wav ... 12.wav, or A.wav ... L.wav. The lowest number or letter will appear to the left on the screen, and should correspond to the 'best' sample – however you choose to interpret 'best'.

Preparing the rating session

Select the files for rating

Start the NeAR program, and use the *Choose Folder* button at the top to pick out the folder you just created. In the box underneath should be indicated the folder and the number of files found there to rate.

If necessary, select your reference files

If you want to use some reference files, tick the box and pick the relevant folder in the next box down.

Choose a name to identify the rating session

To make the analysis easier, all the results go into the same file. The name you type here will help you identify this rating session. You can use the same name twice, but you'll get a warning. The date and time of the session will also be saved, which might let you figure out which rater was which if you've forgotten.

Tick the rating options you want to use

Most of these are self-explanatory. It's probably easier to have a trial run and just watch what happens when you pick the various options, but here is a short explanation of each (on the next page is a picture that might help guide you).

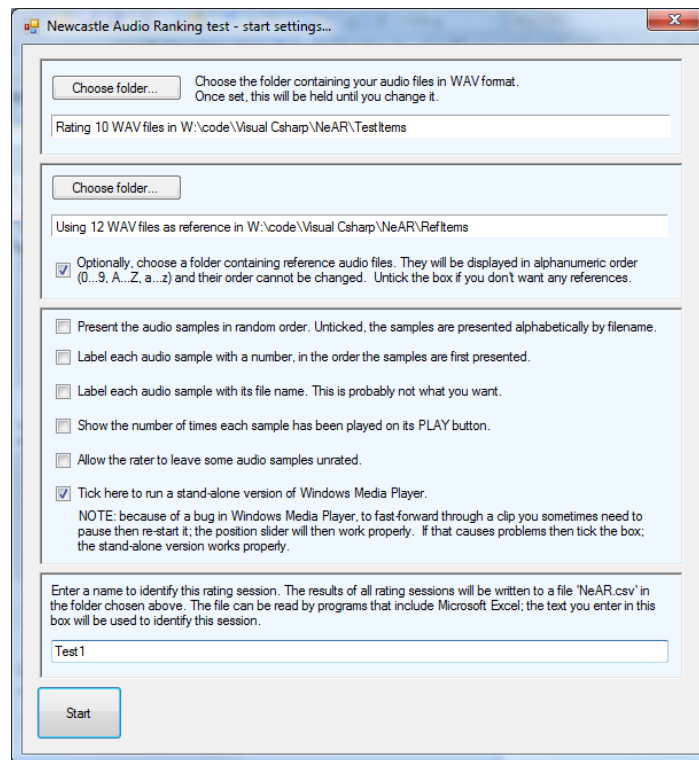
- If you tick the first box, then the audio samples are presented in a different random order each time you run the rating test. This would be useful if you are giving the same test to the same person on multiple occasions to assess test-retest agreement, for example. Otherwise, the audio samples are presented in alphabetical order by file name. This means that every rater gets the samples in the same order.
- If you tick the second box, then the samples will be labelled from 1 upwards. This isn't very helpful.
- If you tick the third box, then the samples will be labelled with their file name. For example, you could deliberately label the samples with the patient's age to see whether that affects the ranking process. Be careful using this option; it's quite likely you used meaningful file names for the original audio samples. For example, they might give away the fact that two samples came from the same patient.
- The fourth tick box indicates how many times each sample has been played. Initially, it was in mind to restrict the users to playing the samples only a fixed number of times but this hasn't proven necessary. If it would be useful then it would be a reasonably easy change.

NOTE: NeAR can't tell if the user has played the sample using the Windows Media Player controls, so this count may in any case not be accurate.

- The fifth tick box allows the user to leave some samples unrated. This can make statistical analysis of your data more difficult. Before you do this, read the discussion of statistics at the end of the manual.
- The final tick box allows you to run a stand-alone version of Windows Media player. Normally, you shouldn't need to use it but on some occasions the fast-forward and rewind buttons don't work properly. If you need them to work and they don't, then use the stand-alone media player.

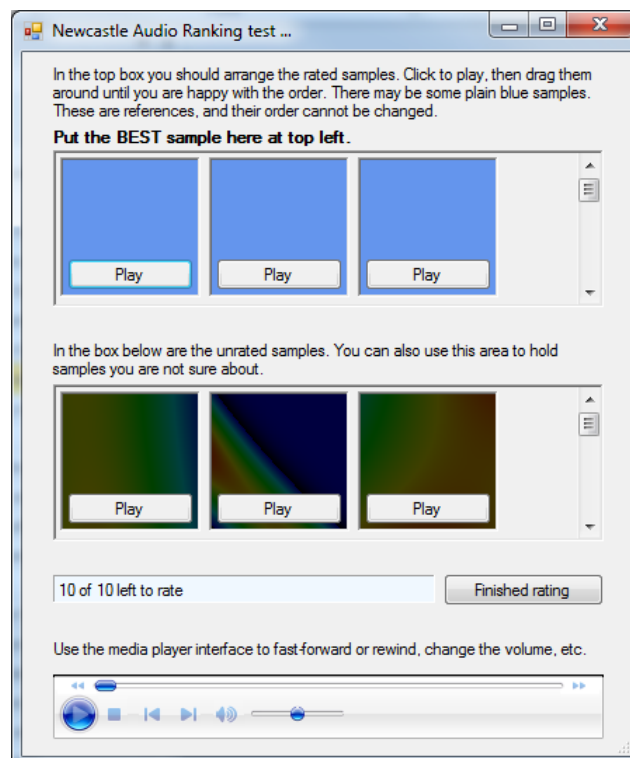
Starting the rating session

If you've done everything correctly, it should look a bit like this:



Click on the start button, and the rating window appears like in the picture below. As normal, you can resize the window to adjust your screen.

- The *unrated* audio samples appear at the bottom of the screen. Each one is coloured differently, so that you can recognise which sample is which.
- If you've specified any reference files, they will appear at the top of the screen in blue as in the example below.

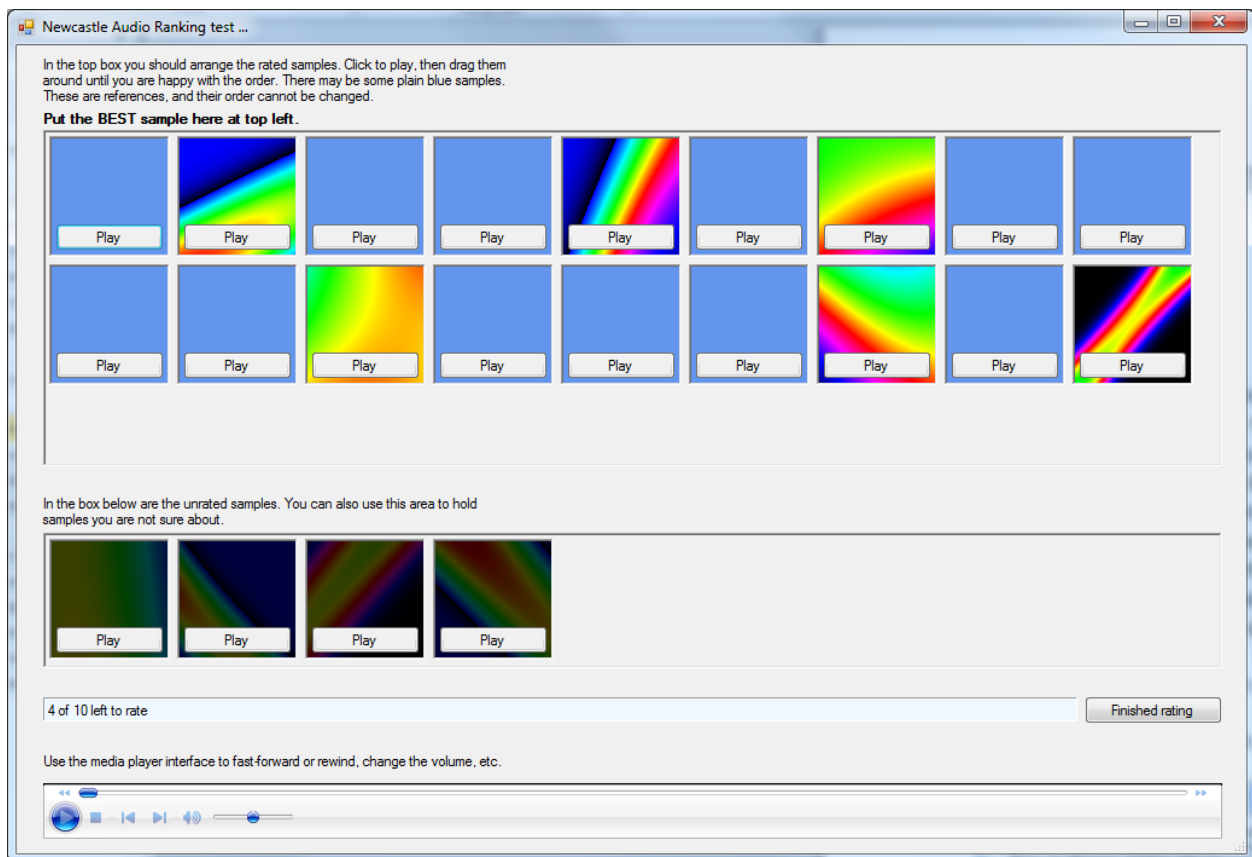


Performing the rating

The rating process is easy, but here are the steps:

- Resize the window to make it easier to see all the samples.
- Click the *play* button on any audio sample to hear it.
- Then, drag it up to the top of the screen using the left mouse button. It will turn bright-coloured when you drop to indicate that it has been rated.
- Repeat for the second audio sample. Drop it to the left or right of the first sample, according to whether you think it was better or worse. And so on.
- You can play and move the audio samples as often as you like, until you are happy with the order.
- If you are not sure about a sample, drag it back to the bottom of the screen where it joins the unrated samples.
- Note that you can't change the order of the blue reference samples.

Here is a rating session after six of the ten samples have been rated:



End the rating session

When you are happy, click the *Finished rating* button. That's it – the results are saved automatically, and you can start another rating session.

NOTE: If you want to abandon a rating session without rating all the samples, but you forgot to tick the box on the front page, then hold down the CTRL key as you click the *Finished rating* button. This feature isn't advertised to the user in the software.

Look at the results

The results are saved in a file called *NeAR.csv* in the same folder as your WAV files. The file can be opened straight into Microsoft Excel, or just about any text editor or statistics package. There is some more detail about the intricacies of the file format on the next page.

Finer details

There might be odd occasions when the program does something unexpected. The error messages are quite informative, but the detail here might help explain what is going on.

The file format

Your results are stored in a file *NeAR.csv*. The CSV (comma separated value) file is in a text format and can be read by any text editor, word processor, spreadsheet or statistics package. However, it is probably best suited to Microsoft Excel for initial viewing. Here's how it might look in Excel:

RATER	DATE	TIME	SOURCE	REFERENCE	NREFS	zz10.wav	zz24.wav	zz89.wav	zz90.wav
Anne	25-Jul-08	18:59:12	C:\src		0	4	3	2	1
Brian	25-Jul-08	18:59:37	C:\src		0	1	4	3	2
Claire	25-Jul-08	18:59:57	C:\src		0	0	3	2	1

- The folder C:\src contained 4 WAV files, named *zz10*, *zz24*, *zz89* and *zz90.wav* respectively.
- Three rating sessions had been conducted to date, with the dates and times as shown.
- Anne thought *zz90* was the best audio sample, and *zz10* was the worst.
- Brian thought *zz10* was best, but the others were in the same order as before.
- Claire was unable to rate *zz10*, but had the other three samples in the same order as before.

Notice that Claire was only able to leave a sample unrated because the corresponding option was ticked on the front page. If it wasn't ticked, then she would have been obliged to rate all the audio samples to complete the task. See the pitfall discussed below.

Using reference files

If you've used 10 reference samples stored in a folder c:\ref, the file might look like this:

RATER	DATE	TIME	SOURCE	REFERENCE	NREFS	zz10.wav	zz24.wav	zz89.wav	zz90.wav
Anne	25-Jul-08	18:59:12	C:\src	C:\ref	10	14	7	2	1
Brian	25-Jul-08	18:59:37	C:\src	C:\ref	10	1	11	10	9
Claire	25-Jul-08	18:59:57	C:\src	C:\ref	10	0	10	7	3

The results for the reference samples aren't included in the table because they always stay in the same order. Therefore there are still only 4 rankings. However, the rankings will now be in the range 1... 14. So for example, Ann thought that sample *zz90* was better than *all* the reference samples, and sample *zz10* was worse than all of them.

How does the program deal with the results file *NeAR.csv*?

- First, the program checks to see if the results file *NeAR.csv* already exists. If not, it is created new.
- If *NeAR.csv* DOES exist, the program reads the file names from the first line of the file. It then compares these file names with the WAV files that are currently in the folder. If the files have been changed, then the results file must be cleared. The program will give you the opportunity to save the existing results.
- Then you perform the rating task.
- After the rating task, the program will add the new results to the end of the file *NeAR.csv* as a new row.
- Normally, that's the end of the story but occasionally the program fails to write the new results. This is probably because you've got the file open in another program; Excel will lock the file and stop any other program from accessing it. In this case, you are given the opportunity to save the results as a new file. Be aware this is a one-off; the next set of results will go back to *NeAR.csv*. So - if you have a problem, you need to sort it out immediately. Otherwise, you will have some results in one file, and some in another.

Analysing the data

Measures of agreement

The results of the ranking test have some special properties that make them amenable to some simple statistical measures of agreement. Consider a group of 10 audio samples in a rating task. Each rank from 1 to 10 will appear just once. Thus each set of rankings will have the same mean (5.5) and the same standard deviation (3.0). Therefore there can be no systematic bias between raters. Under these circumstances many coefficients of agreement, such as Pearson's and Spearman's correlation, intra-class correlation and quadratic weighted kappa become equivalent.

Unrated samples

The biggest potential stumbling block to analysis is where you allow the raters to leave some audio samples unrated. Think carefully whether this is really what you want. Here are issues to think about:

- First, the unrated samples will all be given the value zero. For most analysis, this is NOT the same as missing data. The zeros might be useful in some circumstances, for example to create a histogram showing missing data. HOWEVER in most cases you will need to replace the zeros with a blank, or whatever your own software uses to indicate missing data.
- Second, ranks always start at 1 so it is the higher ranks that will be missing. Some statistical tests (Spearman's rank correlation, for example) are relatively insensitive to this problem, but others are not. Even a simple mean score for each audio sample is probably wrong if you have missing data.

Using reference files

For similar reasons, if you're using reference audio samples these special properties don't hold true because the ranks of the reference files are not included in the results. Therefore one rater might rank all the samples as better than the reference samples, while the next rater ranks them all as worse.

References

Clopper CG. Auditory free classification: Methods and analysis. *Behavior Research Methods* 2008; 40: 575-581.

Granqvist S. The visual sort and rate method of perceptual evaluation in listening tests. *Logopedics Phoniatrics Vocology* 2003; 28: 109-116.

Imai S, Garner WR. Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology* 1965; 69: 596-608.

Rae G. The equivalence of multiple rater kappa statistics and intraclass correlation coefficients. *Educational and psychological measurement* 1988; 48: 367-374.