

# SoundSoftware.ac.uk

## Survey Report

# Index

1 Introduction	
1.1 Survey design and execution.....	2
1.2 Acknowledgements.....	2
2 Survey results	
2.1 General information.....	3
2.1.1 Academic position.....	3
2.1.2 Institution affiliation.....	3
2.1.3 Areas of work/research.....	3
2.2 Software use and development.....	4
2.2.1 Operating systems and programming languages.....	4
2.2.2 Software development.....	5
2.2.3 Comments from respondents about software use and/development.....	6
2.3 Reproducible Research.....	8
2.3.1 Comments on reproducible research .....	8
2.4 Comments from the Final Remarks section .....	10

# 1 Introduction

This online survey aimed to help the SoundSoftware project understand the UK audio and music research community. The survey focused on software use and development and use of datasets in research, and was intended to provide a quick but broad overview of the reality of this community.

## 1.1 Survey design and execution

The survey opened in October 2010 and was advertised to a number of senior researchers in research groups around the UK. We asked for detailed information about software use, authorship and publication practices of researchers, with the aim of obtaining a number of individual case points for further examination as well as some broad numerical results.

The survey closed in April 2011, with 54 complete and 23 partially complete responses. There were responses from at least 16 different institutions, with a number of common issues reported. It was conducted using an online tool (LimeSurvey).

The survey was organized by sections as follows:

1. General—related to researcher's academic background;
2. Software Usage—in this section we tried to find out what kind of software the different researchers were using. We hoped to use these results to identify old or no longer maintained programs, as well as the most common applications for the different kinds of tasks;
3. Software Development—this section of the survey aimed at trying to find out what kind of software development skills the researchers have (version control, operating systems, programming languages, etc);
4. Datasets—this section aimed to identify dataset usage and limitations;
5. Reproducible Research—in this brief section we tried to get some idea of how widely reproducible research principles are adopted within the community;
6. Final Remarks—this section was intended to provide general feedback.

Note that, because the main focus of the survey was on collecting observations and comments, most of the questions were optional.

## 1.2 Acknowledgements

The authors would like to express their gratitude to Aleksandra Pawlik for many suggestions about this document.

## 2 Survey results

The results are organized in four main sections: general information about respondents' affiliations; software and data usage; software and dataset development; and reproducible research.

### 2.1 General information

#### 2.1.1 Academic position

Present position	
PhD student in your first 2 years (first 3 if part-time)	13
PhD student in year 3 or beyond (4 or beyond if part-time)	10
Postdoc student or Research Assistant	16
Academic	31

#### 2.1.2 Institution affiliation

We received answers from 16 different universities, namely:

University of East Anglia, University of Sheffield, Queen's University Belfast (Sonic Arts Research Centre), Queen Mary University of London (Centre for Digital Music, MAT), University of Essex, University of Stirling, De Montfort University, University of Sussex, University of Surrey, University of Southampton, University of Edinburgh (Institute for Digital Communications), Lancaster University, Cambridge University, Medical Research Council–Institute of Hearing Research, University of York, University of Nottingham (School of Psychology).

#### 2.1.3 Areas of work/research

Research Areas	
Machine Listening	14
Music Informatics	8
Audio Engineering	6
Auditory analysis	6
Environmental Acoustics	3
Sound synthesis or design	3

Other areas include: statistical signal processing for audio; physical modeling for sound synthesis; sound synthesis, interactive live animation and improvisational music; games; models of the auditory periphery; speech synthesis, recognition and perception; composition; dialogue systems; radar signal processing; psychoacoustics; musical education; music information retrieval.

## 2.2 Software use and development

### 2.2.1 Operating systems and programming languages

Operating System	
Windows	54
MacOS	44
Linux	33

Most respondents use more than one operating system (55). 15 respondents reported they use all three operating systems. No user reported using only Linux.

Programming Language Used	
Matlab	54
C/C++	39
Max/MSP or Pd	18
Python	15
Java	15
Other	15

Some other languages/environments mentioned were: Perl, Objective C, bash, SuperCollider, Android SDK/NDK, .NET, Praat, CUDA-C, Clojure, Presentation and R.

45 respondents said they use more than one programming language. Two respondents did not mention using any programming language.

### 2.2.2 Software development

In this section the main interest was to find about software development practice, especially the use of code version control software and source code hosting services. This section is significant in this exploratory survey, since version control is a key feature of sustainable software development.

Use of audio-related or acoustics software packages, toolboxes, or libraries in research	
Yes	59
No	13

Do you write your own software, even if it is only intended for your own personal research use?	
Yes	59
No	13

Although the figures are the same for these two questions, they do not represent the same groups—ten respondents that write their own software said they did not use any audio-related packages, toolboxes or libraries.

Version control software	
Subversion	14
Git	7
CVS	5
Mercurial	4
Other	4
None	26

From the results we can see that Subversion (14) and git (7) were the most popular systems. CVS (5) and Mercurial (4) were the other available options. Other systems used include DARCS, RCS and Perforce.

- 6 respondents use more than one system.
- 24 users that develop software said explicitly that they do not use version control, while
- 26 users that develop software did not answer this question.

Use of source code hosting services	
Yes	18
No	30

Of the respondents that said they develop software, 18 said they use code hosting services. 30 respondents said explicitly that their code stays on their computer.

The most used third-party source code hosting services is SourceForge (6 users). 10 users are using university source code version control tools.

Production/maintenance of publicly available software, toolboxes, or libraries (related to the field)	
Yes	18
No	31

Software packages, toolboxes, libraries, or other code planned to become public	
Yes	12
No	36

Software packages, toolboxes, libraries, or other code not planned to become public	
Yes	16
No	32

### 2.2.3 Comments from respondents about software use and/development

- Most [of what I develop] is interesting work, little is good enough for the public
- I'm not intending to publish my code but I would like to make it available privately to other researchers
- I often give people my software when they ask but I do not publish or publicize it
- I'm not sure if something I develop during my PhD will be put into a software tool

- I have written a large number of programs for teaching speech and hearing science [and they're publicly available]
- [My application] is written in Python and works very well for most things. There are lots of sound libraries out there but most aim to provide lots of sound formats or lots of features (3d sound, filtering etc) whereas most of my users are more interested in knowing that when they ask a sound to play it happens instantly.
- While this software is for music, none of it is strictly audio-related, except that it is capable of making musical sound
- At the moment [the software I've developed] is hosted on an SVN server in my office that I have set up, since my institution does not provide such a facility. It would be better if it was hosted externally
- I believe that the majority of audio research software should be open source and developed in the public domain, ideally on an open distribution platform such as GitHub
- Am currently looking for suitable version control. Have toyed with SVN in the past, but with little success



## 2.3 Reproducible Research

Steps taken to ensure the reproducibility of research results	
Yes	23
No	21

23 respondents say they take steps to ensure the reproducibility of their research results, while 21 explicitly say they don't take those steps.

Asked to describe the steps they take, many respondents identify actions which might be considered preliminary to reproducible research, such as preparing to make code available on request.

Some researchers also say that they publish their code in their own pages. There are indications that this procedure can lead to problems with sustainability itself, as their pages sometimes disappear or become (in the case of institutional home pages) inaccessible to them.

Some respondents provided answers explaining why they do not embrace reproducibility. Several complain about the amount of time involved and/or complexity of making research reproducible. Some observe that their "input" data is itself not reproducible or that their results are not deterministic and do not lend themselves to being reproduced. And a few respondents were too early in their work to have published anything yet.

### 2.3.1 Comments on reproducible research

- The software and datasets have been previously described and will be published shortly. I appreciate that it should have been published before the paper, but this was not possible due to practical limitations, especially time
- Published results of experiment on my own webpage
- Documentation of [data and] scripts used in order that other (internal) people (including myself) can reproduce the results
- Still within the first month of research and have yet to put any procedures in place
- I work with physics people and they I don't think they operate this way because I think the complexity of their work is in the equations they use, not their programs. I think its a great idea when complex code is involved

- Provided my code for research purposes on request
- Published software and data used to evaluate [my] algorithm
- Data results published but not full source yet so not entirely reproducible
- Use the McGill dataset, use known cochlea models like the gammatone filterbank
- I'd tend to publish the research code if it isn't horribly messy. I've open sourced quite a few things over the years [...] and also released MATLAB code on the side to other researchers
- Just put the data online and give the webpage address in paper
- I have published Matlab code for my thesis and other papers on my personal website
- Published Matlab toolbox containing algorithms and codes to reproduce figures of some of my papers
- Publishing some datasets, making some software available publicly
- I have planned to, but not yet got round to it!
- It's always made available to interested researchers
- Published measurements are generally calibrated by standard procedures
- All the software I write for my research is open source and available in the public domain - including all development history and commit logs
- Designed a website for additional information to a book I wrote, but the website is time consuming to keep up-to-date, especailly since the host university has kept changing its policy, IP addresses etc. Now I have no direct access to the site any more
- Yes, methods are published; code, data, results given out on request
- Publish software and source code freely on internet and give out datasets upon request once results have been published
- The generic methods used are discussed and relatively easy to produce. In its current state, the amount of refinement and training needed to use the software I have developed would require a significant investment of additional time, which cannot be justified against my other research commitments

- I haven't published any results yet
- So far I use a publicly available corpus (the AMI-meeting corpus) and will make the one we recorded also publicly available. The steps used to come to the results are explained in the papers
- Making code available that has been used to generate results, although I have not done this as much as I would like as it is time consuming
- Code is all available and papers describing how the code works
- Participation in MIREX tasks (<http://www.music-ir.org/mirex/>)
- Max/MSP results aren't 100% the same each time
- I'm too protective of my work to share
- Ongoing journal article referring to software available on [site], need to add the Matlab code to produce the figures with the data

## 2.4 Comments from the Final Remarks section

- Many of the answers to these questions will change within the next few months as I am still within the first month of my PhD
- Did not have time to put as many details as wanted, but hopefully most important is there. Survey worked well
- This survey was a bit too literal
- I am at a very early stage in my research so my needs are probably going to change. I hope this is still of some use
- The best thing that can be done is to identify and support standards for data which would allow interoperability of tools