An integrated approach to the design of digital signal processing algorithms for hearing instruments

A Data Management Plan created using DMPonline.be

Creator: Arnout Roebben

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Arnout Roebben

Grant number / URL: 11PDH24N

ID: 191521

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

A key component to the successful design of hearing instruments is the design of digital signal processing (DSP) algorithms. These algorithms realize specific functionalities such as noise reduction, dereverberation, dynamic range compression, adaptive feedback cancellation, and active noise control. While there has been a significant advance in the technology of such DSP functionalities, different DSP functionalities are mostly developed and analyzed in isolation. Whenever such DSP functionalities are cascaded in a complete system, their interaction -even counteraction- is rarely taken into consideration. This in particular compounds the general problem of robustness in typical dynamic acoustic environments, as observed by a hearing instrument. Hence, the objective of this project is to address this problem by following an integrated approach: integrating individual DSP functionalities in a common acoustic control problem statement and solution strategy. From this, suitable integrated algorithms will be designed and compared against the state-of-the-art based on off-line simulations, as well as based on real-time testing and perceptual evaluations.

Last modified: 23-02-2024

An integrated approach to the design of digital signal processing algorithms for hearing instruments Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

The entire required DMP for the FWO has already been developed as good practice. A summary of the used data types is as follows:

*) Manuscripts: .tex and .pdf (500 MB)
*) Administration: .ppt, .docx, .tex and .pdf (1GB)
*) Code (generated): .m, .slx and .c (5GB)
*) Simulation results and experiment output (generated): .mat and .txt (5 GB)
*) Recorded and numerically generated acoustic scenarios and experiment input (generated and reused): .mat and .txt (5 GB)
*) Metadata datasets, simulations, and equipment (generated and reused): .txt (1GB)
*) Subjective listening protocol (generated): .tex and .pdf (100 MB)
*) Subjective listening experiment - personal data (generated): .txt (10 MB)

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

Data will be preserved as open as possible and as closed as necessary. The main promotor (M. Moonen) is responsible for data preservation. The data will be stored on the ESAT-STADIUS servers (8 TB capacity). Access is regulated by an access control list granting read-write access to the project owner and read-only access to specific users, managed by the promotor and ESAT-STADIUS data manager. Also, GitLab repositories (10 GB/repository capacity) and KU Leuven RDR data repositories (50 GB capacity) are available, providing storage during and after the research.
Only one person is assigned the key info for subject identification. All anonymized data will be stored on encrypted hard drives.

What’s the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

We do not wish to deviate from this principle. The terms (and conditions) of data preservation are mentioned in the informed consent that subjects will sign prior to their participation in the subjective perceptual evaluation. Data preservation will be ensured through highly restricted overwrite and delete authorization. No data will be shared with external parties without the subject's consent.
Open science will be adhered to where possible, according to the 'as open as possible, as close as necessary' principle using Gitlab (code) and KU Leuven RDR (data) repositories. The subjective perceptual evaluation results are most relevant to store because of the difficult-to-recreate nature.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

Human subjects are involved whose personal data will be processed, falling under a declaration that will be submitted to the Ethical Committee. Subject data will be anonymized. Additionally, personal information will be stored in a separate database only accessible by the handling researcher.
The data protection policy of FWO and KU Leuven will be followed in a strict way. Access to data on the servers is regulated by an access control list (ACL), managed by the ESAT STADIUS data manager, granting read-write access to the project owner and read-only access to specific users. Client computers can access the data using: SMB2 from specific IP ranges NFSv4 from specific (ITmanaged) systems.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

New members in the research group get training on how to perform experiments and how to handle the collected data. Standard guidelines and instructions are available on the network drive which allows for the information, given during the training, to be refreshed. The 'open as

possible, closed as necessary' principle will be adhered to.
To gain more insights into good practices for data management, I have followed a 'research data management' seminar, as well as a seminar on 'open science'. By acting on the acquired insights of these seminars, the entire required DMP for the FWO has already been completed since I believe a well-defined DMP aids the organization of the project.

An integrated approach to the design of digital signal processing algorithms for hearing instruments
FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .cvs,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Manuscripts	Internal and external reports on the conducted research	Generate	Digital	Other	.tex, .pdf	< 1GB	
Administration and seminar paperwork	Administrative paperwork and research input	Generate and reuse	Digital	Other	.pdf, .docx, .ppt	< 100 GB	
Code	MATLAB, Simulink and speedgoat coding functions	Generate	Digital	Software	.m, .slx, .c	< 100 GB	
Simulation results and experiment output	Output files of the simulations and experiments	Generate	Digital	Simulation Data	.mat, .txt	< 100 GB	
Recorded and numerically generated acoustic scenarios and experiment input	Input files of the simulations and experiments	Generate and reuse	Digital	Simulation data	.mat, .txt	< 100 GB	
Metadata datasets, simulations, experiments and equipment	Metadata of the utilised input, output and tools	Generate and reuse	Digital	Simulation data	.txt	< 1 GB	
Subjective listening experiment protocol	Worked out protocol of the subjective listening experiments	Generate	Digital	Other	.pdf	< 100MB	
Subjective listening experiment output	Results of the subjective listening experiments	Generate	Digital	Experimental (Personal data)	.txt	< 100 MB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Impulse responses room: 10.5281/zenodo.7389996
Impulse response feedback path hearing aid: 10.1016/j.jfranklin.2006.08.002
HASPI, HASQI and HAAQI metrics: Received via an email to one of the developers, professor James M. Kates

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes, human subject data

Ethical approval will be obtained at the Ethics Committee Research UZ for the behavioural listening tests for normal-hearing young subjects in workpackage 4. For this, pseudonymised results of the listening tests, together with relevant information about diversity (gender and age).

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes

Relevant information about diversity (gender and age) will be noted down in the behavioural listening tests of workpackage 4.

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

The integrated system holds promise for valorisation when outperforming the state-of-the-art cascaded systems. The potential for valorisation will be assessed before sharing any data, and if necessary KU Leuven Research & Development (LRD) will be contacted.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

Code will be documented by means of comments in the code and a header.
GitLab repositories will be accompanied by a README file.
Behavourial listening outcome results will be accompanied by a README file.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

Metadata with specific information on experiments carried out with the developed algorithms and software will be stored in associated text files.
Data generated as output from measurements and simulation codes will be stored in standard formats to be imported into software packages.

3. Data storage & back-up during the research project

Where will the data be stored?

The data will be stored on ESAT servers, of which the access is regulated by an access control list (ACL) that grants:
- read-write access to the project owner
- read-only access to specific users

The ACL is managed by the promotor and STADIUS Data manager. Client computers can access the data using:
- SMB2 (or higher) from specific IP ranges
- NFSv4 from specific (IT managed) systems

Code and manuscripts will additionally be saved to GitLab repositories. Data of the behavioural listening tests will be made available, depending on the informed consent, using KU Leuven RDR repositories.

How will the data be backed up?

All data on the ESAT servers are backed up daily and replicated to an off-site storage system housed in the ICTS data center.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

The GitLab repositories (10 GB/repository) and ESAT servers possess over more than enough storage capacity for the data as described in section 1.1 . The same applies to the KU Leuven RDR repositories offering 50 GB/researcher each year.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The data will be stored on ESAT servers, of which the access is regulated by an access control list (ACL) that grants:
- read-write access to the project owner
- read-only access to specific users

The ACL is managed by the promotor and STADIUS Data manager. Client computers can access the data using:
- SMB2 (or higher) from specific IP ranges
- NFSv4 from specific (IT managed) systems

The KU Leuven servers protected by the KU Leuven IT department.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The servers, GitLab repositories and KU Leuven RDR are available without costs to the project.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

All data will be kept according to KU Leuven's RDM policy.

Where will these data be archived (stored and curated for the long-term)?

The code and manuscripts will be managed on the ESAT servers and GitLab repositories. The behavioural listening experiment results will be stored safely on the ESAT servers and can be made available on KU Leuven RDR depending on the outcomes of the informed consent of the experiment. Five years after the project ends the data on the ESAT servers will be moved to the offline archive storage.
Any data generated in the course of the project that is not longer in use will be removed from the ESAT servers.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

No costs, as the code and manuscripts will not exceed the free 10 GB space and the KU Leuven RDR is freely accessible by the project members. Furthermore, the research division provides data storage facilities for their researchers for no cost.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in a restricted access repository (after approval, institutional access only, …)

The project adheres to the 'as open as possible, as closed as necessary' data sharing principle.
Code and manuscripts will be stored in GitLab repositories that can be made publically available.
The behavioural listening experiment results can be made available on KU Leuven RDR depending on the outcome of the informed consent of the participants.

If access is restricted, please specify who will be able to access the data and under what conditions.

By the informed consent, the behavioural test subjects can indicate whether and to whom their data can be shared with. This will be adhered to strictly.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- Yes, Ethical aspects

By the informed consent, the behavioural test subjects can indicate whether and to whom their data can be shared with. This will be strictly adhered to.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

Code and manuscripts: GitLab
Data: When possible by informed consent KU Leuven RDR

When will the data be made available?

Upon publication

Which data usage licenses are you going to provide? If none, please explain why.

Depending on the informed consent results.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

When possible by the informed consent

What are the expected costs for data sharing? How will these costs be covered?

No costs, as the code and manuscripts will not exceed the free 10 GB space and the KU Leuven RDR is freely accessible by the project members. Furthermore, the research division provides data storage facilities for their researchers for no cost.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Arnout Roebben. After the contract ends, the responsibility is transferred to PI Marc Moonen.

Who will manage data storage and backup during the research project?

The IT division of the department (ESAT) will be responsible for the data storage and back-up during the project.

Who will manage data preservation and sharing?

PI Marc Moonen.

Who will update and implement this DMP?

PI Marc Moonen.

An integrated approach to the design of digital signal processing algorithms for hearing instruments
GDPR

GDPR

Have you registered personal data processing activities for this project?

- Yes

An integrated approach to the design of digital signal processing algorithms for hearing instruments
DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable