

**Data Usage Agreement (DUA)**

**SONIVA Dataset – Audio Files Access**  
Imperial College London – Clinical Language and Cognition Lab

**1. Purpose**

This Data Usage Agreement governs access to the audio recordings in the SONIVA dataset, collected by the Clinical Language and Cognition Lab (CLC Lab) at Imperial College London. These recordings feature speech from individuals’ post-stroke and matched controls, and are considered personally identifiable information under the UK General Data Protection Regulation (GDPR).

While transcriptions and extracted features are openly available via GitHub, the **audio files require controlled access** due to their sensitive biometric nature. This agreement defines the responsibilities of all users requesting access to the audio files and outlines the limitations placed on usage, redistribution, and data security.

**2. Data Description**

The audio recordings contain short spoken picture descriptions from stroke survivors and control participants. Each file is linked with anonymised metadata (e.g., age, sex, accent, clinical scores, lesion labels), but does **not** contain names, images, or directly identifying details.

Nonetheless, since human voice is a biometric identifier, these audio recordings constitute indirectly identifiable information and require appropriate handling.

**3. Requesting Access**

Access to the audio portion of the SONIVA dataset is granted only upon receipt and review of:

* A signed copy of this Data Usage Agreement (DUA).
* A certificate of Good Clinical Practice (GCP) (e.g., NIHR Learn certificate) or equivalent data protection training.
* A brief description of the intended research use, including affiliation and planned analysis.

Approved users will be granted secure access via a link maintained by Imperial College London.

**4. Conditions of Use**

By signing this agreement, I (the requesting researcher) and my affiliated institution agree to:

* Use the data solely for academic, non-commercial research as outlined in the access request
* Not attempt to re-identify participants, either directly or by linking with other datasets
* Not share, transfer, or redistribute the audio data to anyone not named in the DUA
* Host and store the data only on secure, encrypted systems under institutional control
* Ensure all team members accessing the data have appropriate training and institutional authorization
* Report any suspected data breach or misuse immediately to the CLC Lab team at Imperial College London
* Cite the dataset appropriately in any resulting publication or output, and include the required data availability statement

These obligations remain **in effect after termination of access** to the dataset.

**5. Publication & Citation**

If results derived from the SONIVA data are published or presented, the following citation must be included:

> Sanguedolce, G., Price, C. J., Brook, S., Gruia, D. C., Parkinson, N. V., Naylor, P. A., & Geranmayeh, F. (2025). \*SONIVA: Speech recOgNItion Validation in Aphasia\*. [Journal Name], [DOI to be assigned].

A statement must also be included in the Methods section, e.g.:

> “*Access to audio recordings from the SONIVA dataset was granted following approval from the data custodians and submission of a signed Data Usage Agreement. Audio data are available under controlled access conditions from Imperial College London*.”

**6. Termination and Limitations**

Imperial College London may revoke access at any time if this agreement is breached or if ethical/legal concerns arise. Continued use of the dataset after termination is prohibited.

This agreement does not transfer intellectual property, nor does it grant rights beyond those specified.

**7. Signature (To be completed by the requesting researcher)**

By signing below, I confirm that:

* I have read and understood this Data Usage Agreement
* I agree to comply with all of its terms and conditions
* I understand that failure to do so may result in access being revoked and potential institutional reporting
* I agree that these obligations remain in force even after access to the data is withdrawn

**Full Name:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
**Institutional Affiliation:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
**Institutional Email Address:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
**Position/Title:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Signature:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
**Date:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_