

**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 contained in the SONIVA dataset, collected by the Clinical Language and Cognition Lab (CLC Lab) at Imperial College London. The dataset includes speech recordings of individuals post-stroke and healthy controls, which are considered personally identifiable information under UK GDPR regulations.

While transcriptions and extracted features are openly available via GitHub, the **audio files** require controlled access due to their sensitive nature. This agreement outlines the responsibilities of researchers requesting access to ensure ethical and legal data handling.

**2. Data Description**

The audio recordings include spoken language from participants involved in post-stroke recovery research. Each file is associated with metadata (age, sex, accent, clinical scores, lesion labels) but contains no direct identifiers such as names or images. However, because human voice is considered biometric data, it constitutes **indirect identifiable information**.

**3. Requesting Access**

Access to the audio portion of the SONIVA dataset is granted only upon submission of:

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

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

**4. Conditions of Use**

By signing this agreement, the researcher and their institution agree to:

* **Use the data solely for non-commercial, academic research** as outlined in the request.
* **Refrain from any attempt to re-identify participants**.
* **Not redistribute the audio data** in any form or upload it to public repositories.
* **Store the data securely** on institutional or encrypted devices (e.g., within Imperial OneDrive or equivalent).
* **Report any data breaches or misuse immediately** to the CLC Lab team.

**5. Publication & Citation**

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

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].

Additionally, include a statement in the Methods section noting that audio data were shared under controlled access following approval from the authors and a signed DUA.

**6. Termination of Access**

Imperial College London reserves the right to revoke access at any time if terms of this agreement are breached or if new ethical or legal considerations arise.

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

By signing below, I confirm that I have read, understood, and agree to comply with all terms outlined in this Data Usage Agreement.

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

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