When sound processing meets data protection
François Pellegrini, Félicien Vallet

To cite this version:
François Pellegrini, Félicien Vallet. When sound processing meets data protection. Linux Audio Conference 2020, scrime; LaBRI, Nov 2020, Bordeaux / Virtual, France. hal-03096429

HAL Id: hal-03096429
https://inria.hal.science/hal-03096429
Submitted on 5 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
When sound processing meets data protection

François Pellegrini & Félicien Vallet
WHO IS THE CNIL?
The CNIL (1)

* “Commission nationale de l'informatique et des libertés”
  * French Data Protection Authority
* Created by the Law n° 78-17 on “Informatics & Freedom” of 6 January 1978
  * Independent Administrative Authority
    * Belongs to the Executive branch but independent from the Government and empowered to oversee it
* 18 Commissionners and 230+ staff
  * Plenary sessions + Restricted Committee
The CNIL (2)

* Article 1 of the French “Informatics & Freedow” Law:
  * “Informatics must be at the service of every citizen. Its development must take place within the framework of international cooperation. It must not interfere with human identity, human rights, privacy, or individual or public liberties”
What is "data with personal character"

* Any information related to an identified or identifiable person, directly or indirectly,
* in particular relating to an identification number
  * E.g.: social security number, phone number, etc.
* or to elements related to the concerned person (data subject)
  * E.g.: biometrics, psychological character, etc.
* Even more so: "data with inter-personal character"
  * Data is produced and exchanged for interaction
Rights of data subjects

* Right to object
  * May be restricted, depending on the file
* Right of access
  * Indirect access, for some police files
* Right to request rectification
  * Data must always be accurate
* Right to be de-listed
* Right to data portability
  * Revolutionary right brought by the GDPR
The GDPR

* More an evolution than a revolution
* Formalization of some key principles
  * Accountability instead of prior authorization
  * Data protection by design, data minimization, etc.
* Creation of the Data Protection Officer
* Strengthened European collaboration
  * Through the European Data Protection Board
* Fully applicable to non-EU data controllers
The CNIL in action(s)...

* Advising and regulating
  * Advice on draft laws, Authorizations, ...
* Accompanying compliance
  * Appointment of DPOs, ...
* Protecting data subjects
  * Verification, indirect right of access, ...
* Investigating
  * On-site or on-line
* Rendering orders and issuing sanctions
* Informing
  * +8M visits to the CNIL website in 2019
WHAT ABOUT THE VOICE?
A private piece of data

* Something *we* use daily
* Something *that* defines us
* Something *that* “says” a lot
A volatile piece of data

* Something **intangible**
* Yet that can be **recorded**
* As for image, **voice right exists**
A reproducible piece of data

* Usurpation, an old issue

* Likely to evolve in the coming years
An intellectual property protected piece of data

* Vocal performance is a creation of form exhibiting originality, thus eligible as a work in author's right
  * Voice acting, a specialized profession

* But with new technologies, new questions are raised
A variable geometry data

* Personal data, a broad scope

* Many information can be extracted from the voice signal

* In some cases, deemed “sensitive data”
  * As of art. 9 GDPR
  * When used for biometry
WHERE DOES CNIL STAND?
Biometrics (1)

* Biometric data is a highly sensitive kind of data
  * Not revocable
* Two mains uses of biometry
  * Authentication
  * Identification
Biometrics (2)

* Identification does require a centralized database, while authentication does not
  * E.g.: biometric passport
* Architecture does matter!
  * Data protection by design
* CNIL promotes biometry “at the hand of the persons”
  * People should always retain control over their biometric data
Voice Biometrics

* Many attempts to use biometry as a “quick” authentication means
  * Yet only one of the three factors
    * Inherence, possession, knowledge
    * Not always the most secure one
* CNIL authorized several pre-GDPR experiments by banks, etc.
  * Yet without the creation of huge stocks of biometric data
  * Biometry “at the hand of the persons”
    * Biometric data are encrypted with user-specific keys
Voice Assistants (1)

* Study started in 2017

* Tests at LINC, CNIL’s lab
Voice Assistants (2)

* September 2020: white paper on voice assistants [EN soon]

* Use-cases for professionals

* Good practices for:
  * Voice assistant designers
  * Application developers
  * Voice assistant integrators
  * Voice assistants “deployers”
  * Users
Voice Assistants (3)

* 4 major pieces of advice for privacy-friendly voice assistants:

* “Keep positive frictions”
* Prefer the local to the remote
* Ensure the means of control
* Adapt to the voice medium
AND TOMORROW?
Speech processing + Privacy by design = ?

* A **new scientific field**

* Research projects financed in France and EU

* A **strategic positioning** for some economic actors

**COMPRISE**
Cost-effective, Multilingual, Privacy-driven voice-enabled Services
Open questions...

* What about the use of voice samples to infer the health condition?

Carnegie Mellon University

COVID Voice Detector

Record your voice to help AI beat Covid!
Carnegie Mellon University is collaborating with researchers around the world to develop an automated AI system that can detect signatures of Covid-19 infection in the human voice.

In order to build the system we require a large number of voice samples -- from Covid patients, from healthy people, and from people affected by other ailments.

Please sign in and record your voice, to contribute to this effort.
Open questions...

* What about the automated evaluation of emotions?

* The quantification of “social skills” in a recruitment process?
Open questions...

* What rights for a post-mortem use of voice characteristics?
THANK YOU

{fpellegrini, fvallet}@cnil.fr