Speech data and the GDPR: First reflections from a legal perspective

Privacy: Speech Meets Legal Experts
Eurecom, 29 January 2020

Catherine Jasserand

STeP (Security, Technology and e-Privacy)

University of Groningen

Background

Foreword: distinction privacy v. data protection

➢ **Privacy:** no universally accepted definition, very broad scope (protection of family, home, correspondence, right to develop own identity, and also extends to the processing of personal data)

➢ **Data Protection:** relates to the protection of personal data.

At EU level: *2 distinct rights*

At level of Council of Europe**:** *right to privacy encompassing the protection of personal data*

Background – data protection

Focus on data protection at EU level

➢ **The right to data protection= fundamental right** (Art. 8 of Charter of Fundamental Rights)
   N.B. The right to privacy is a separate right
   (Art. 7 Charter Fundamental Rights)

➢ Right detailed in secondary legislation, such as in the General Data Protection Regulation (GDPR – Regulation 2016/679)

Background on the GDPR

Double Objective:

➢ (1) the protection of natural persons for the processing of their personal data (human rights objective)

➢ (2) facilitating the free flow of personal data within the EU (economic objective)

➢ Rules to protect data collection and processing: *What is 'personal data'? 'processing'? What are the conditions under which the data can be processed? etc.*

Speech Data

> Discussion based in part on two papers

__References:__

1. The GDPR & Speech Data: Reflections of Legal and Technology Communities, First steps towards a Common Understanding

2. Legal Nature of Biometric Data: From 'Generic' Personal Data to Sensitive Data

> Difficulties for non-technical experts: what does speech cover? How does it relate to voice? How is data speech processed? What are the formats resulting from the processing? For which purposes speech data is used? Etc…The easiest is to analyse it through case (e.g. voice assistants, voice data used in the context of healthcare)

Speech and the GDPR

What does the GDPR say about speech/voice?

- ➢ Nothing! There are references to fingerprint or facial images, but not to voice or speech.

- ➢ But **voice** or **speech data** can be considered as **personal data**…

  Personal data is any information relating to an identified or identifiable individual **(Article 4(1) GDPR)**

  ! identification from a data protection perspective has a specific meaning (= singling out someone)

- ➢ **Voice** or **speech data** can also fall in the category of **biometric data**, defined as

  Article 4 (14) GDPR:

  Biometric data' means personal data resulting from specific technical processing relating to the physical, physiological or behavioral characteristics of a natural person, **which allow or confirm the unique identification of that natural person**, *such as facial images or dactyloscopic data.*

Speech and the GDPR

➢ If **voice** or **speech data** is processed to 'uniquely identify' an individual, then it can be subject to the rules applicable to sensitive data

Article 9(1) GDPR:

Processing of…*biometric data for the purpose of uniquely identifying a natural person*…shall be prohibited…

➢ *!* The text of the GDPR is subject to (<u>various</u>) <u>interpretations</u>. This is one of the hurdles that we face when we try to understand and explain the notions. Example with the notion of biometric data and the enigmatic phrase 'uniquely identify'

Speech and the GDPR

➢ **Voice** or **speech data** is more than biometric data. If it reveals sensitive information, (such as the health condition of an individual [e.g. Parkinson], ethnicity [based on accent, intonation], it can still be **another type of sensitive data**.

Article 9(1) GDPR (exhaustive list of sensitive data)

Personal data, which falls in the category of sensitive data, cannot be processed unless an exception applies

Article 9(2) GDPR

One of the exceptions relates to the processing of 'manifestly made public by the data subjects themselves'

! Distinction between data publicly available (disclosure by third-parties) and data made publicly available *by the data subject*.

Issues to address

➢ **Identification of information in speech** that deserves protection: what is personal data and what is not. Which data can be derived/inferred from speech

➢ **Processing operations of speech data** (from the capture to their use and storage)- Which data qualifies as personal data? Which format?

➢ **Storage of speech data:** local, in a cloud? Decentralised?

➢ **Entities involved** in speech data to know each one's responsibility

Issues to address

> **Safeguards:** how to protect the content? De-identification of speech data? Different techniques need to be explored.

> **Case studies** (smart homes, social media, eLearning platforms, health care, etc.)

Some conclusions

➢ Necessary collaboration with scientists (understood broadly) to understand the implications of the GDPR (1st objective)

➢ The purpose is not that technical experts become lawyers, and lawyers technical experts. The purpose is to share knowledge across communities to have a basic understanding of the issues.

References

➢ A. Nautsch et al. *, Preserving Privacy in Speaker and Speech Characterisation* (2019) Computer Speech and Language, 58, pp. 441-480

➢ A. Nautsch, C. Jasserand, E. Kindt, M. Todisco, I. Trancoso, and N. Evans. *The GDPR & Speech Data: Reflections of Legal and Technology Communities, First steps towards a Common Understanding* (2019) Interspeech, 20th Annual Conference, conference paper

➢ C. Jasserand, *Legal Nature of Biometric Data: From 'Generic' Personal Data to Sensitive Data* (2016) European Data Protection Law Review, vol. 2, issue 3, pp.297-311

Thank you for your attention

Catherine Jasserand: c.a.jasserand@step-rug.nl