



What is Speech/Voice from a data privacy perspective? Insights from the GDPR

Café Talk

Catherine Jasserand, PhD

STeP (Security, Technology and e-Privacy)
University of Groningen (RUG)



Background

- ✓ More and more applications use voice and speech data
- ✓ Need to understand the impact from a privacy/data protection perspective
- ✓ Need to assess the new data protection instrument at EU level (GDPR) in relation to these applications
- ✓ This talk is a first step on the notions of speech and voice



What is Speech? What is Voice?

- ✓ Linked (?) but distinct
- ✓ In texts written by non-experts, confusion between the two terms
[Special thanks to Isabel Trancoso to clarify the distinction and points out to the NIDCD's website
<https://www.nidcd.nih.gov/health/what-is-voice-speech-language>
- **Voice** understood as 'the sound produced by humans (and other vertebrates) using the lungs and the vocal folds in the larynx, or voice box. Voice is not always produced as speech.'
- **Speech** : 'Human express thoughts, feelings, and ideas *orally* to one another through a series of complex movements that alter and mold the basis tone created by voice into specific, decodable sounds.'



What is Speech? What is Voice?

- ✓ **Voice** linked to an individual

Voice recognition = to identify an individual

- ✓ **Speech** : not necessarily spoken, can be transcribed
Speech recognition = recognition of spoken words? The identification of the individual does not matter.

Applications:

- Interception of communications (whether authorized or not): both voice and speech can be valuable in that context, i.e. knowing who talks (voice) and what was said (speech).
- Smart assistants: reacting to voice command, but they can record conversations.



Privacy and data protection

- ✓ No single and universal definition of privacy; data protection is also known as 'data privacy' outside Europe
- ✓ Different approaches depending on legal order of reference
 - At the **level of the Council of Europe** (pan-European organization), the notion of privacy is interpreted as encompassing the protection of personal data (i.e. data protection) – only the right to privacy is a fundamental right
 - At the **EU level**: privacy and data protection are separate notions (and distinct fundamental rights)



Summary

Council of Europe

Privacy

(private and family life,
 home, and correspondence)
 Art. 8 European Convention
 on Human Rights

Data Protection

(judicial
 interpretation of
 Art. 8)

EU level

Privacy

(private and family
 life, home, and
 communications)
 Art. 7 Charter of
 Fundamental Rights

Data Protection

Art. 8 Charter of
 Fundamental Rights



Basics: GDPR

- ✓ **General Data Protection Regulation**
- ✓ Replaces an outdated instrument from 1995 (known as 'Data Protection Directive')
- ✓ Implementation of the fundamental right to data protection
- ✓ GDPR in a nutshell:
 - Principles to process personal data (transparency, purpose limitation, data minimisation, accuracy, data retention, data security, accountability)
 - Obligations for entities/individuals processing these data
 - Rights for individuals whose personal data are processed
 - Oversight by independent supervisory authorities
 - Remedies in case of non-compliance with GDPR's obligations (with possible hefty fines)



Voice, Speech, Personal data?

- ✓ **Voice or speech** ≠ **personal data**, only representations of voice and speech (e.g. voice recording, speech sample, written texts) can be considered as personal data if they relate to an identified/identifiable individual. Thus, only voice data or speech data can be personal data

See Art. 4(1) GDPR:

‘personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

- identifiability/identification has to be understood from a data protection perspective
- Meaning = to be singled out
- The GDPR does not refer to voice or speech data; the previous instrument did (reference to the processing of sound and image data falling within the scope of the Directive)



Voice, Biometric data?

- ✓ **Voice data could also qualify as biometric data**
- ✓ Legal notion of biometric data = type of personal data

Art. 4(14) GDPR

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

- Pre-requisite: need to be personal data,
- Resulting from technical processing of biometric characteristics
- Ex: fingerprints, facial images, but no reference to sound or voice data
- Problematic criterion: '*which confirm or allow the unique identification*'
!! Uncertainty concerning its meaning, confusion as well b/c of specific meaning of 'identification' in biometric recognition context!



Voice, Biometric Data ?

- ✓ Biometric data can also qualify as sensitive data (Art. 9 GDPR)
 - *N.B. **Sensitive data** is a **special category** of personal data benefitting from a higher level of protection because of their nature (or the purpose of their processing, see biometric data).*
 - Includes data revealing ethnicity, political opinions, religious or philosophical beliefs, but also genetic data or biometric data for the purpose of uniquely identifying an individual
 - General rule: processing prohibited, unless an exception applies (10 grounds)
- ✓ Issue: meaning of 'uniquely identifying' in relation to biometric data
Does it mean to process the data for biometric identification purposes? Or to process the data using characteristics that are deemed 'unique' to an individual?
 - ▶ No binding interpretation (yet).



Voice data, Speech data, type of data

- ✓ **Voice data** can reveal the identity of an individual but also sensitive information such as the demographic/ethnic origin (through accent), a possible disease (such as Parkinson?)
- ✓ **Speech data** can reveal the context, environment, with whom an individual speaks.
- ✓ They could both qualify as personal data, but also as sensitive data (even biometric data) due to the nature of the information they could reveal.

!! Important to identify the correct nature of the data b/c the legal grounds on which data can be processed are different

- 6 grounds to process personal data (Art. 6 GDPR)
- 10 grounds of exception to process sensitive data (including biometric data qualifying as sensitive data) (Art. 9(2) GDPR)



Food for Thought

- ✓ A **voice** can be linked to an individual, as such recordings, samples, and other forms could qualify as personal data.
 - If voice data are used for biometric recognition purposes, they could also qualify as biometric data
 - But a voice could also reveal other types of sensitive information (health related or ethnicity) and be considered sensitive data
- ✓ **Speech** is not limited to the scope of personal data. Speech can indeed provide information about the whereabouts of an individual, his/her connections, his/her habits, the identity of other individuals, etc. In short, the recording or use of speech can also infringe the right to privacy.

Questions from my side: what are the formats under which voice/speech data are processed? What is the linkability of these formats to an individual? Is it possible to 'de-identify' voice/speech data?



university of
 groningen

faculty of law

STeP
RESEARCH GROUP

27/08/2020 |

Thank you!

Contact details: c.a.jasserand@step-rug.nl