



Voice anonymisation and the GDPR

Brij Mohan Lal Srivastava, CEO

LDS Technology Workshop
29/01/2024



www.nijta.com



Context

Sharing data creates value !

1M hours of calls recorded **per month**
for large Contact Centres

Call recording is essential for:

- Listening (Quality assurance), and
- Extracting key strategic information for their business at scale.



Voice recordings

May contain, besides the linguistic content of speech

voice characteristics

(e.g., loudness, pitch, nasality, roughness, hoarseness)

speech characteristics

(e.g., speech, tempo, rhythm, pronunciation, accent, articulation, intonation, clarity of expression)

non-speech human sounds

(e.g., sneezes, coughs, laughter, cries, breaths, sighs)

background sounds

(e.g., vehicle sounds, natural soundscapes, media sounds, eating sounds)

May allow inferences about user's...

Physical health (e.g., Parkinson's disease, asthma, common cold, flu)	Gender	Mental health (e.g., depression, schizophrenia, post-traumatic stress disorder)	Age	Personality traits (e.g., openness, extroversion, neuroticism)
Socioeconomic status	Communication disorders (e.g., language disorders, voice disorders, speech disorders)	Geographical origin	Impression on other people (e.g., perceived as charismatic, competent, emotionally stable)	
Moods and emotions (e.g., happy, angry, sad, bored, frustrated, anxious)	Body measures (e.g., body, weight, body height)	Biometric Identity	Medium-term states (e.g., sleepiness, intoxication)	

Overview of some sensitive attributes **discernable** from speech data

Context

Call recordings contain sensitive information about the customers.

This raises challenges around privacy, data security and protection, regulatory compliance, and customer trust.



Key problem

Valuable voice data is simply **erased** to comply with the **GDPR** and other privacy similar regulations

-€€€€,€€€€
/cc/month.



Contact Center Challenges

- Dependence on **third-party vendors** for voice analytics
- No data ownership! **No in-house AI!**
- 30-90 Days to **delete** the data
- Fear of non-compliance

AI powered Anonymization,
rather than data **deletion**.

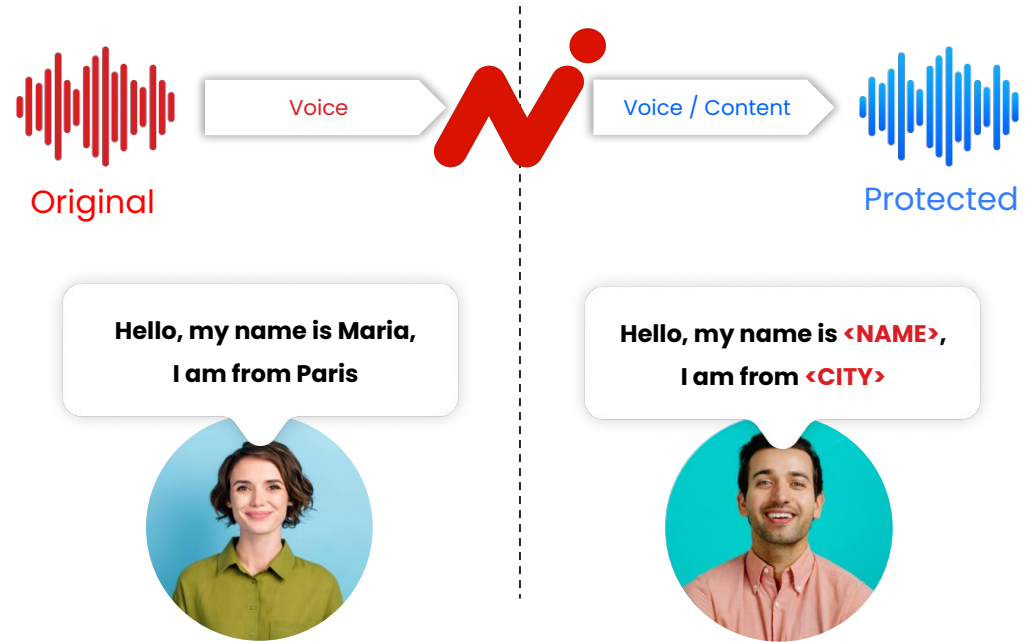


Our proposition

Nijta removes **biometric and sensitive information** from **speech data** while preserving the utility* of voice for AI analytics

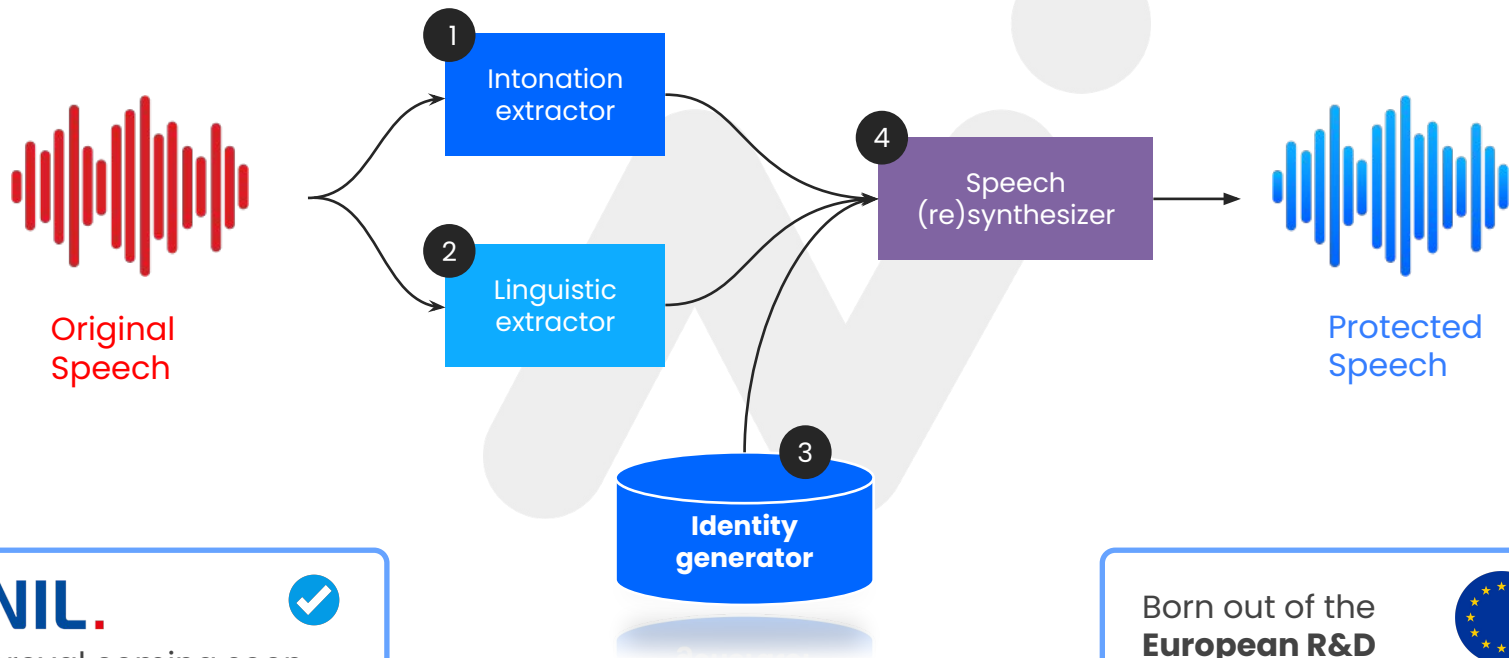
**prosody, emotions, age, health, ...*

Speech to Speech Anonymization



Our key feature

STS – Voice anonymization



CNIL.



Approval coming soon


Born out of the
European R&D



GDPR Compliance and privacy guarantees*

Accountability
Accuracy
Confidentiality

Data minimisation

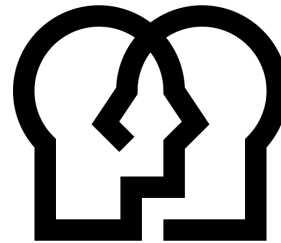


Lawfulness
Purpose limitation
Storage limitation

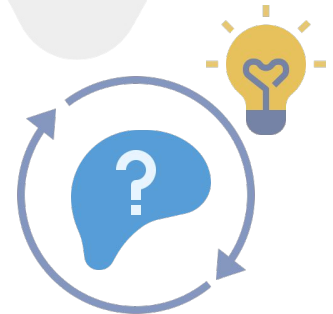
Singling out



Linkability



Inference



*Opinion 05/2014 : Article 29 Data Protection Working Party



Partnerships



Get in touch

For follow-up questions and demo, reach out to:

Address:

Campus Cyber Hauts-de-France
Euratechnologies
165 avenue de Bretagne
59000 Lille



Brij SRIVASTAVA

Co-founder and CEO

+33 (0)6 09 39 66 85

brij@nijta.com

<https://www.linkedin.com/in/brijsri/>



Demo link

www.nijta.com