TÉCNICO LISBOA

Voice Biometrics: Privacy in Paralinguistic and Extralinguistic Tasks for Health Applications

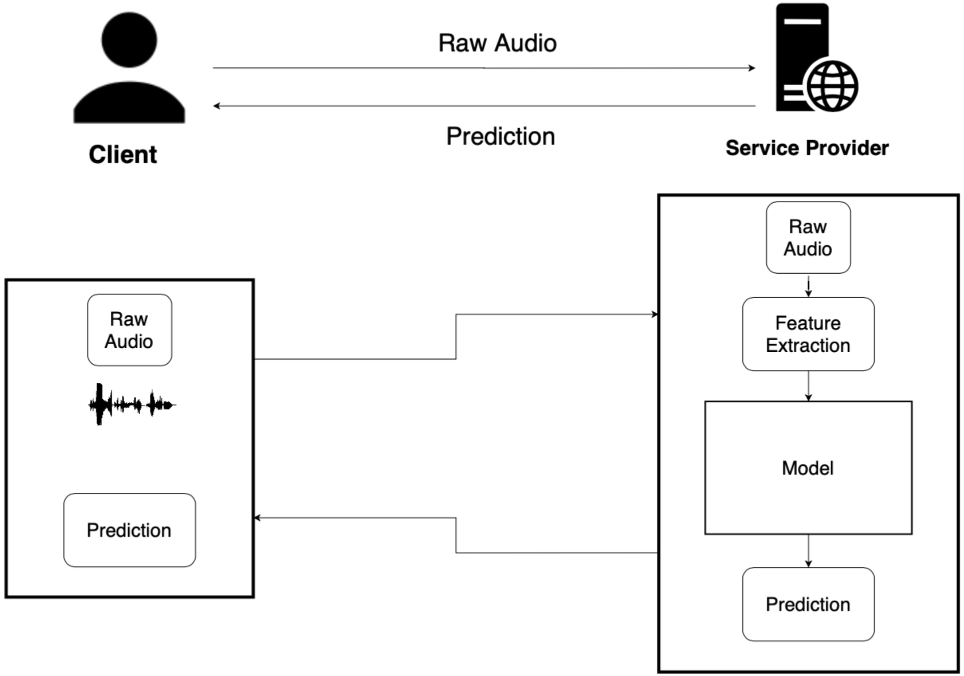
ISCA-SPSC Webinar
July 6th, 2020

Francisco Sepúlveda Teixeira,
francisco.s.teixeira@tecnico.ulisboa.pt

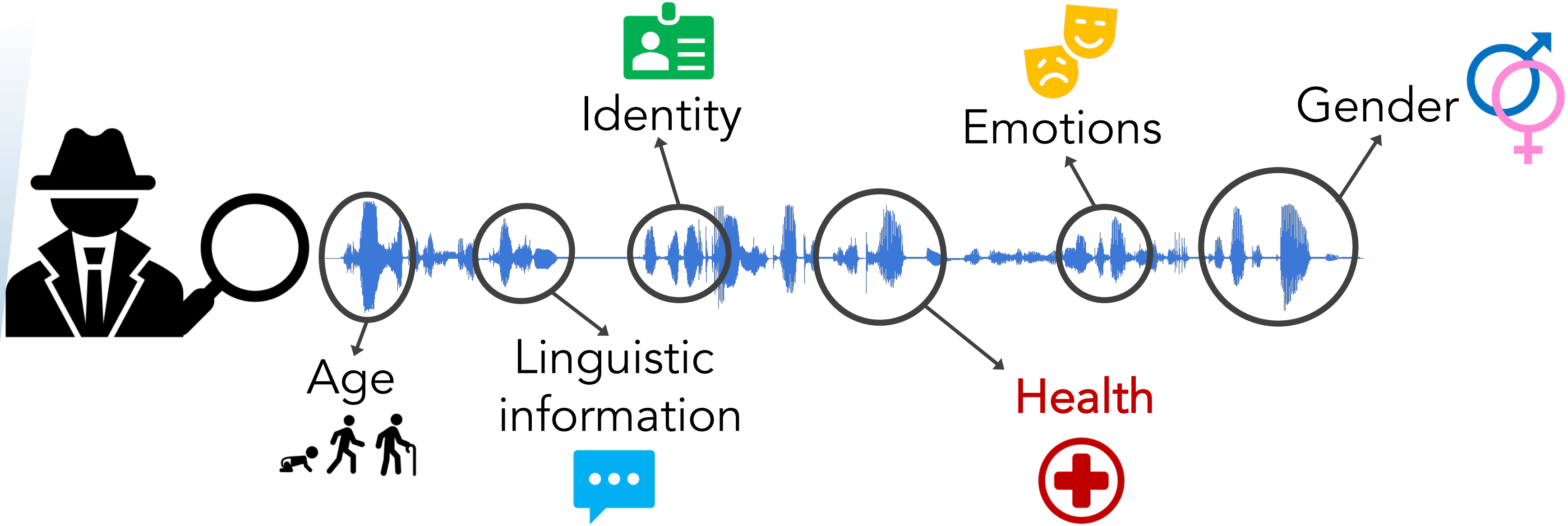
Isabel Trancoso, Alberto Abad and Bhiksha Raj

inesc id lisboa

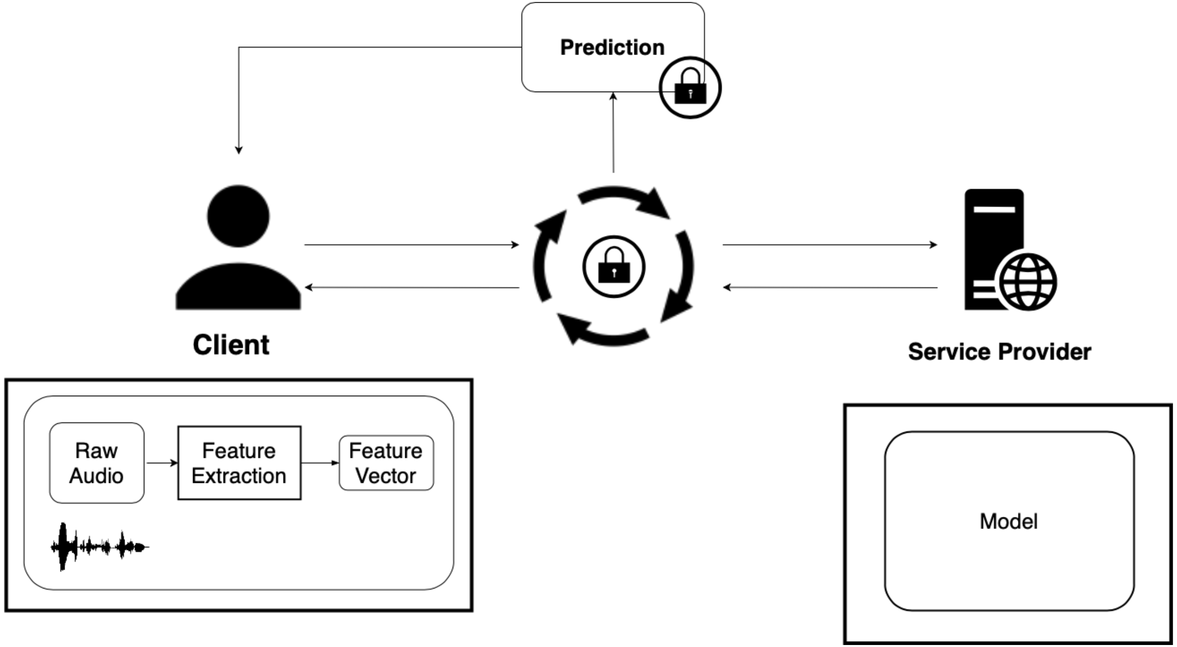
Machine Learning as a Service (MLaaS)



Personally Identifiable Information (PII) in speech



Privacy-preserving Machine Learning as a Service



Outline

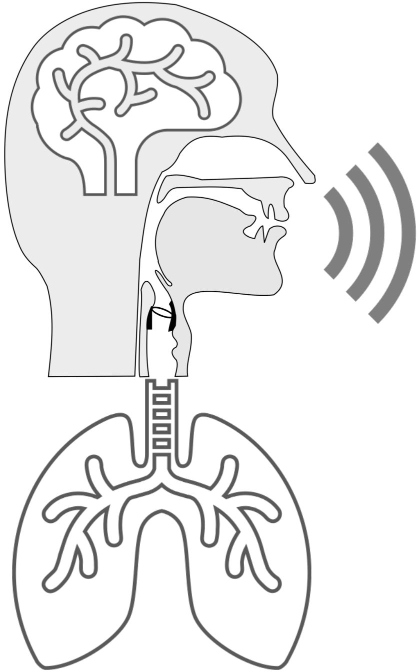
- Speech-affecting diseases
- Paralinguistic and Extralinguistic Tasks for Health: Methods and Features
- Cryptographic primitives and Secure Multiparty Computation protocols
- Privacy-preserving Paralinguistic and Extralinguistic tasks
- Conclusions

Speech-affecting diseases

Speech-affecting diseases

- Speech production is a very complex process.
- It can be affected by numerous factors.

- Outside *sigmatism* and *stuttering* (speech and language disorders), many other diseases affect speech:
    - Diseases that concern respiratory organs;
    - Mood disorders;
    - Neurodegenerative diseases;

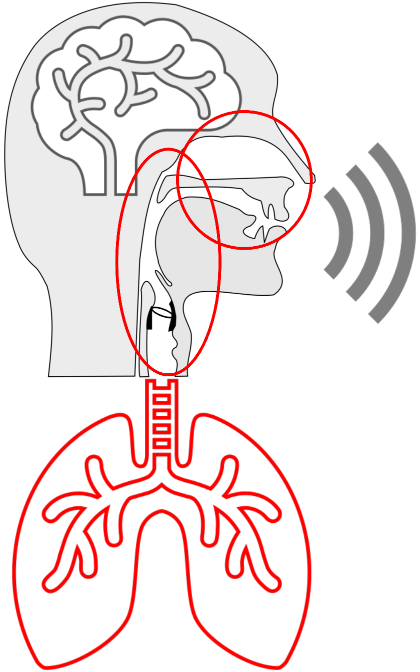


Speech-affecting diseases

- Diseases that concern respiratory organs, e.g.:
  - Obstructive Sleep Apnea (OSA);
  - Common Cold.

Obstructive Sleep Apnea (OSA):

- **Articulatory disorders**, especially dysarthria, causing slurred speech;
- **Phonation anomalies**, due to larynx inflammation caused by snoring;
- **Resonance anomalies**, due to abnormal coupling of vocal and nasal tracts.



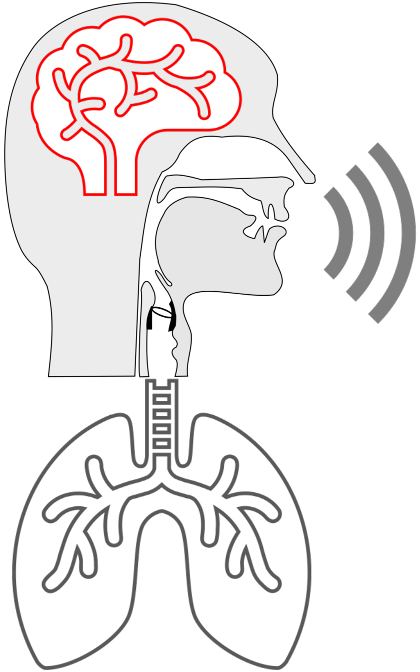
Speech-affecting diseases

- Diseases that concern respiratory organs;

- **Mood disorders, e.g.:**
  - Anxiety;
  - Depression;
  - Bipolar Disorder;
  - Post-traumatic stress disorder (PTSD).

Depressed speech is described as:

- Dull;
- Monotone;
- Mono-loud;
- Lifeless;
- Metallic.



Speech-affecting diseases

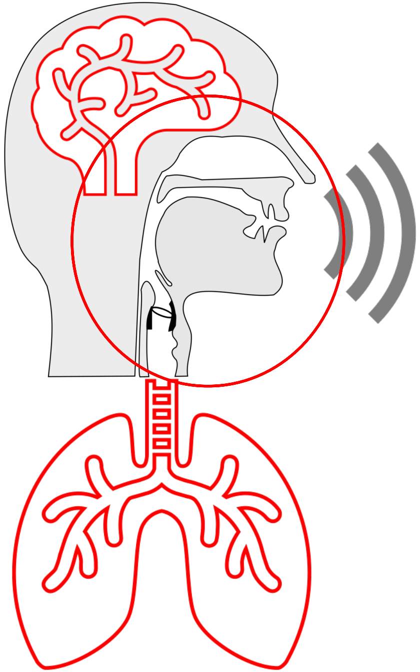
- Diseases that concern respiratory organs;

- Mood disorders;

- **Neurodegenerative diseases, e.g.:**
  - Alzheimer's disease (AD);
  - Parkinson's disease (PD);
  - Huntington's disease (HD);
  - Amyotrophic lateral sclerosis (ALS).

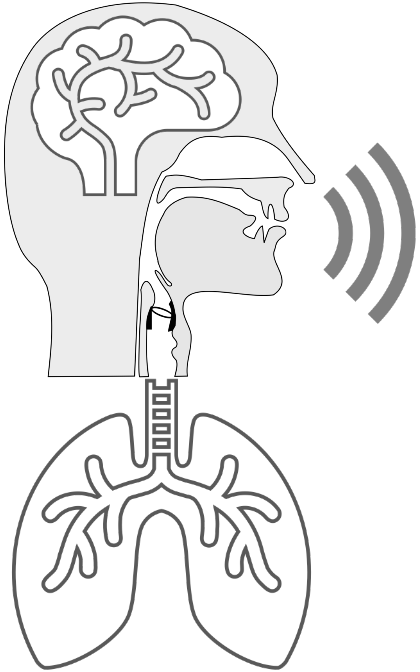
Parkinson's Disease (PD):

- Dysarthria (difficulty in articulation);
- Hypophonia (reduced loudness);
- Hurried speech;



Speech-affecting diseases

- **Speech has the potential to become an inexpensive and non-intrusive biomarker for health.**
- Speech can be applied to the diagnosis and monitoring of many disorders.
- Pathological speech data is hard to obtain due to legal and ethical constraints.
- Pathological speech databases are usually very small.
- The sensitive nature of speech for medical applications **demands privacy-preserving solutions.**



Paralinguistic and Extralinguistic Tasks for Health

Para/Extralinguistic tasks for Health: Knowledge-based (KB) features

- Mood disorders (e.g. Depression):
  - Lower mean F0.
  - Smaller range of formant frequencies.
  - Reduced variation in loudness.
  - Higher jitter, shimmer and HNR.

- Respiratory and neurodegenerative disorders (e.g. OSA, PD):
  - **Phonation:** jitter, shimmer, amplitude perturbation quotient, pitch perturbation quotient, harmonics-to-noise ratio (HNR), etc.
  - **Articulation:** vowel space area, vowel articulation index, formant centralization ratio, diadochokinetic analysis (DDK), onset energy, MFCCs, etc.
  - **Prosody:** pitch & energy contours, durations, etc.

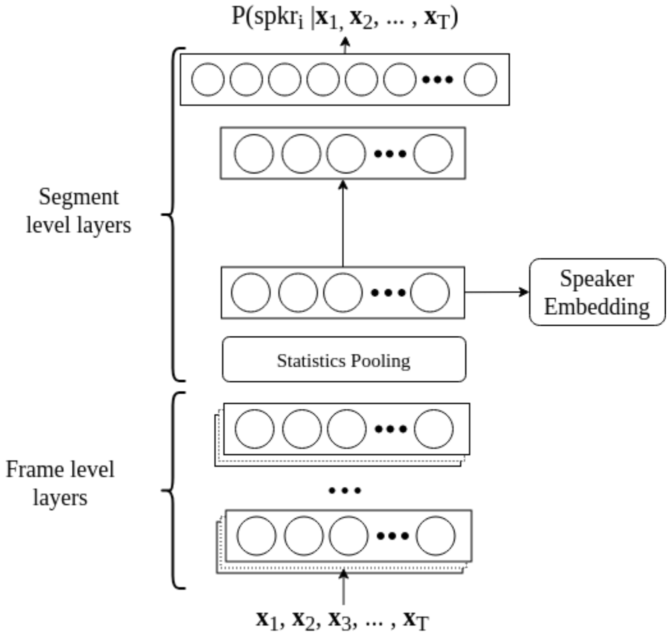
- KB features are more easily **explainable**, facilitating the creation of bridges between data scientists and healthcare professionals.

Para/Extralinguistic tasks for Health: Generic features

- GeMAPS and eGeMAPS (OpenSMILE Toolkit):
  - (Extended) Geneva Minimalistic Acoustic Parameter Set – Originally proposed for generic paralinguistic tasks.
  - Includes functionals of Low-Level Descriptors (LLDs) that model: spectral, cepstral, prosodic, and voice quality parameters.
- Bags-of-Audio-Words:
  - Represent audio frames as distributions of LLDs from a codebook (e.g. MFCCs, PLPs, …).
- Deep-spectrum features:
  - Obtained by forwarding spectrograms through pre-trained CNNs.
- Bottleneck features:
  - Obtained from intermediate layers of auto-encoder networks, trained with out-of-domain data.

Para/Extralinguistic tasks for Health: Speaker representations

- Generic **speaker** representations inherently model disease symptoms:
  - GMM-UBM supervectors;
  - *i-vectors;*
  - *x-vectors.*
- Speaker representations have been shown to work as well as KB features for the detection and assessment of PD [1], OSA [2] and Alzheimer's Disease [3].


$P(\text{spkr}_i \mid \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$
Segment level layers
Speaker Embedding
Statistics Pooling
Frame level layers
$\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_T$

Para/Extralinguistic tasks for Health: Classifiers

- Support Vector Machines (SVMs) – Linear, Polynomial and Radial Basis Function (RBF) kernels;
- Deep Neural Networks (DNNs);
- Convolutional Neural Networks (CNNs);
- Recurrent Neural Networks (RNNs);
- Probabilistic Linear Discriminant Analysis (PLDA) (together with speaker representations);
- End-to-end models: CNNs+ Long-Short-Term Memory (LSTM) recurrent layers.
- …..

Cryptographic primitives and Secure Multiparty Computation protocols

Homomorphic Encryption (HE)

- Homomorphic Encryption is a family of cryptographic methods that respect the following:

$$E(x) \otimes E(y) = E(x \times y)$$
$$E(x) \oplus E(y) = E(x + y)$$

- **Advantages** of HE:
  - + Usually requires only two communication rounds, Client -> Server and Server -> Client.
  - + Very efficient at performing linear operations (e.g. inner products).
  - + Some **(Leveled) HE** schemes **allow** several values to be encrypted into a single *ciphertext*, allowing for **SIMD operations** (e.g. BGV [4], BFV [5] and CKKS [6]).
  - + Many of these cryptosystems are based on the Learning with Errors (LWE) and Ring LWE problems which are assumed to be post-quantum secure [7].

Homomorphic Encryption (HE)

- **Disadvantages** of HE:
  - Very high computational cost.
  - Most schemes are limited in the number or type of operations allowed.
  - Increasing the number of operations allowed implies increasing the computational cost of the scheme.

e.g. Computational cost of HE: Take as an example the BFV cryptosystem, which is based on the Ring Learning with Errors (RLWE) problem:

- HE *ciphertexts* based on RLWE are defined as polynomials.
- Consider a polynomial of degree 4096;
- Assuming each coefficient of this polynomial to be at most 110 bits-long (value recommended for 128-bit security [8]):

The ciphertext occupies **55KB long**.

Secure Multiparty Computation (MPC)

- MPC is a family of cryptographic protocols that allow two or more parties to interactively (and privately) compute a certain function. The most important types of MPC protocols are:
  - Arithmetic and Boolean Secret Sharing [9,10].
  - Garbled Circuits [11].

$$\langle x \rangle = \langle x \rangle_0 \oplus \langle x \rangle_1$$

- **Advantages** of MPC:
  - + Additions are computed for free in Arithmetic Secret Sharing:
  - + XORs are computed for free in Boolean Secret Sharing and Garbled Circuits.
  - + Garbled Circuits are very efficient to compute complex functions (e.g. comparisons).

- **Disadvantages** of MPC:
  - - MPC requires the constant presence of all parties involved in the computation.
  - - The high number of communication rounds entails a large bandwidth usage.

Secure Multiparty Computation – Security Models

- *Honest-but-curious* model:
  - Both parties are assumed to follow the established protocol, but to try to get as much information as possible from the data that is visible to them.
  - Can be used in applications where both parties are trustworthy (e.g. interaction between hospitals or clinics and companies).

- *Malicious* model:
  - Parties are assumed to thwart the protocol.
  - Require Zero Knowledge (ZK) proofs and cut-and-choose methods to ensure all parties are "behaving" correctly.
  - Can be used in settings where parties do not trust each other (e.g. two competing companies that need to perform a computation over their private data).
  - Although much more secure, this model significantly increases the computational cost of MPC protocols.

Secure Modular Hashing (SMH)

- Information-theoretically secure distance-preserving hash function proposed by Jímenez et al., 2015 [12]:

$$\boldsymbol{Q_k}(\boldsymbol{x}) = \lfloor A\boldsymbol{x} + \boldsymbol{w} \rfloor (\mathrm{mod\ k})$$

$$\boldsymbol{A} \in \mathbb{R}^{N \times M} \sim N\left(0, \tfrac{1}{\delta^2} I_N\right)$$

$$\boldsymbol{w} \in \mathbb{R}^{N} \sim \mathrm{unif}[0, k]$$

  - Irreversible transformation.
  - It is not possible to gain information from a hash vector without having any prior information about it, or having access to the transformation parameters (or *hash key*) - (A, w).

Secure Modular Hashing (SMH)

$$Q_k(x) = \lfloor Ax + w \rfloor (\text{mod k})$$

$$\left\lfloor \overbrace{\begin{pmatrix} 0.1 & 0.7 & 0.5 & 0.2 \\ 0.6 & 0.6 & 0.3 & 0.1 \\ 0.3 & 0.4 & 0.5 & 0.7 \\ 0.2 & 0.8 & 0.5 & 0.3 \\ 0.4 & 0.1 & 0.8 & 0.9 \\ 0.7 & 0.9 & 0.4 & 0.7 \\ 0.8 & 0.6 & 0.2 & 0.1 \\ 0.9 & 0.1 & 0.3 & 0.7 \end{pmatrix}}^{A} \times \overbrace{\begin{pmatrix} 0.5 \\ 0.8 \\ 0.1 \\ 0.3 \end{pmatrix}}^{x} + \overbrace{\begin{pmatrix} 0.7 \\ 0.4 \\ 0.1 \\ 0.5 \\ 0.6 \\ 0.8 \\ 0.1 \\ 0.2 \end{pmatrix}}^{w} \right\rfloor \overbrace{(\text{mod } 2)}^{k} = \overbrace{\begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}}^{Q(x)}$$

Secure Modular Hashing (SMH)

- In SMH the **Hamming Distance (HD)** between two **hash vectors** (created using the same key) is:
  - **Proportional** (i.e. provides information) to the **Euclidean Distance (ED)** between the **original vectors** if the **ED** is **smaller** than a **threshold**.
  - Provides **no information** about either vector **otherwise**.

$$d_E(x,y) = \sqrt{\sum_{i=0}^{M}(x_i - y_i)^2}$$

$$d_H(x,y) = \frac{1}{M}\sum_{i=0}^{M} x_i \oplus y_i$$

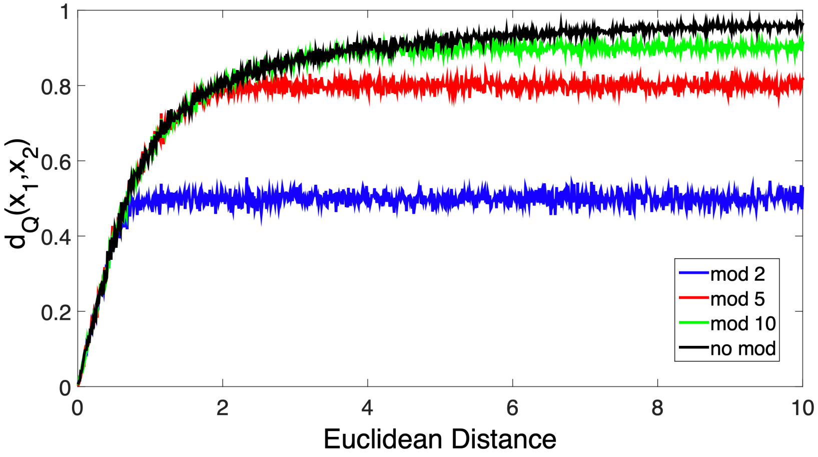


Figure 1: SMH - ED & HD proportionality.

Other cryptographic/secure methods

Other methods include:

- Dedicated protocols (e.g. GSHADE [13] for secure hamming distance computation);
- Functional Encryption;
- Secure enclaves (e.g. Intel's SGX).

Privacy-preserving training:

- Differential Privacy;
- Federated Learning;

Privacy-preserving Machine Learning (PPML) for Paralinguistic and Extralinguistic Tasks

Privacy-preserving (Encrypted) Neural Networks (ENNs)

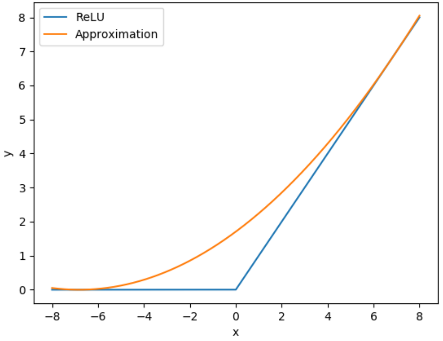
- Originally proposed as **Cryptonets** by Gilad-Bachrach et al., 2016 [14] (later improved by Chabanne et al., 2017 [15], and Hesamifard et al., 2017 [16], among many others).

- All operations in the NN are replaced by their HE counterparts.
- Non-linear Activation layers are replaced by polynomials.

- When using batching allows for multiple predictions to be computed at the same time (useful for para/extralinguistic tasks).

- Network architecture is limited by noise growth, weight scaling and computational complexity.

- Requires discretized weights and inputs, which may lead to accuracy degradation.



Polynomial ReLU:

$$y = 0.037x^2 + 0.5x + 1.71$$

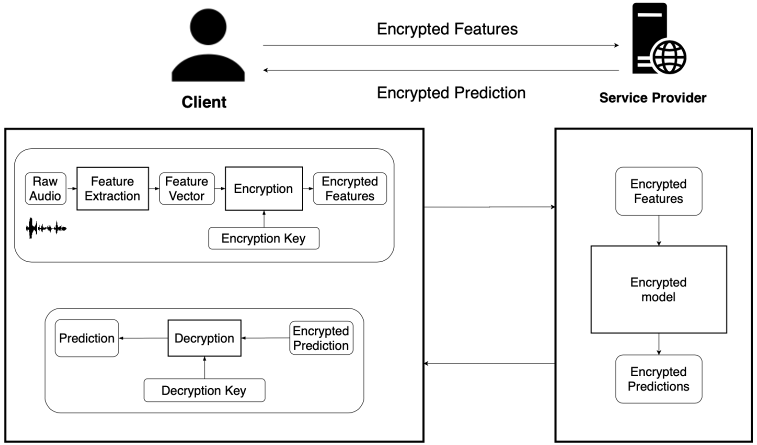
Privacy-preserving ENNs for Para/Extralinguistic Tasks

- First applied to a paralinguistic task (**emotion recognition**) by *Dias et al.*, 2018 [17].

- Later applied to the detection and assessment of **Parkinson's Disease**, the common **Cold** and **Depression** [18, 19]:
  - Shallow neural network (1 hidden layer).
  - Knowledge based features (eGeMAPS and variations).

  - **Negligible accuracy degradation** when compared to results obtained *in-the-clear.*

  - ~4.5 s for a single prediction without the use of batching;
  - ~23 s for 16,384 simultaneous predictions, yielding an amortized cost of ~1.4 ms per prediction.



Privacy-preserving SVM+RBF kernel

- SVMs are widely used for tasks where data is scarce.
- This is the case of many paralinguistic and extralinguistic tasks, in particular for tasks related with health, where privacy is of the utmost importance.
- Due to its computational complexity, few works have proposed solutions to the problem of privacy-preserving inference using the RBF kernel.
- **Our solution:** using SMH we can replace the Euclidean Distance (ED) with the Hamming Distance (HD), making the kernel much cheaper to compute privately.
- We take advantage of the fact that the HD is a sum of XORs, and that both of these operations are essentially “free” to compute with Arithmetic and Boolean Secret Sharing, respectively [20].

Privacy-preserving RBF kernel

$$d_E(x,y) = \sqrt{\sum_{i=0}^{M}(x_i - y_i)^2}$$

$$\mathrm{k}(\boldsymbol{x}, \boldsymbol{x}'_i) = \exp(-\gamma d_E^2(\boldsymbol{x}, \boldsymbol{x}'_i))$$

$$\mathrm{k}(d'_H) = \exp(-\frac{\gamma}{M^2}{d'}^2_H)$$

$$d_H(x,y) = \cancel{\frac{1}{M}}\sum_{i=0}^{M} x_i \oplus y_i$$

$$\mathrm{k}(d'_H) = x^5 - 1.54 \times 10^{-14}x^4 - 1.20 \times 10^{-08}x^3 - 1.76 \times 10^{-16}x^2 + 7.77 \times 10^{-17}x - 6.05 \times 10^{-20}.$$

Computed using Arithmetic and Boolean Secret Sharing

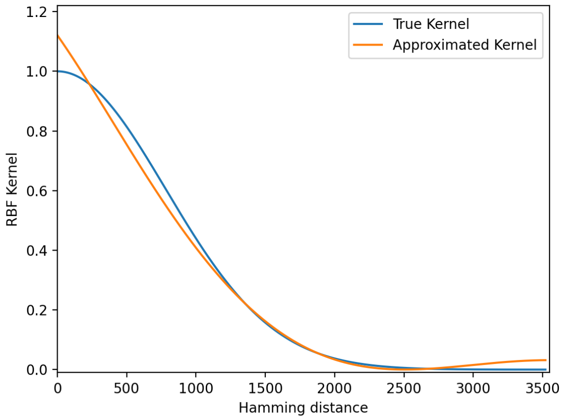


Figure 2: Polynomial Aproximation

Privacy-preserving Support Vector Machine (eSVM)

Support Vectors
transformed using SMH

$$\hat{y}(\boldsymbol{x}) = \operatorname{sign}\Big(w_0 + \sum_{i=0}^{n} \alpha_i y_i \underbrace{k_H(\boldsymbol{x}, \boldsymbol{x}_i')}\Big)$$

Garbled Circuits

Secret Sharing
+
Homomorphic Encryption

Input
transformed using SMH

Homomorphic Encryption

PPSVM - Experiments

Our method was tested for the detection of two speech affecting diseases:

- Obstructive Sleep Apnea (OSA):
  - Portuguese Sleep Disorders Corpus [21].
  - Modelled using the 109-long feature set proposed by Botelho et al. [21].
- Parkinson's Disease (PD):
  - Subset of the New Spanish Parkinson's Disease Corpus [22].
  - Modelled using the 114-long feature set proposed by Pompili et al. [23].

- Models were trained and evaluated using Leave-One-Speaker-Out (LOSO) cross-validation.
- Implemented in C++ using SEAL [24] for HE and ABY [25] for MPC.
- Results were obtained using 4-second long audio segments.

PPSVM – Results

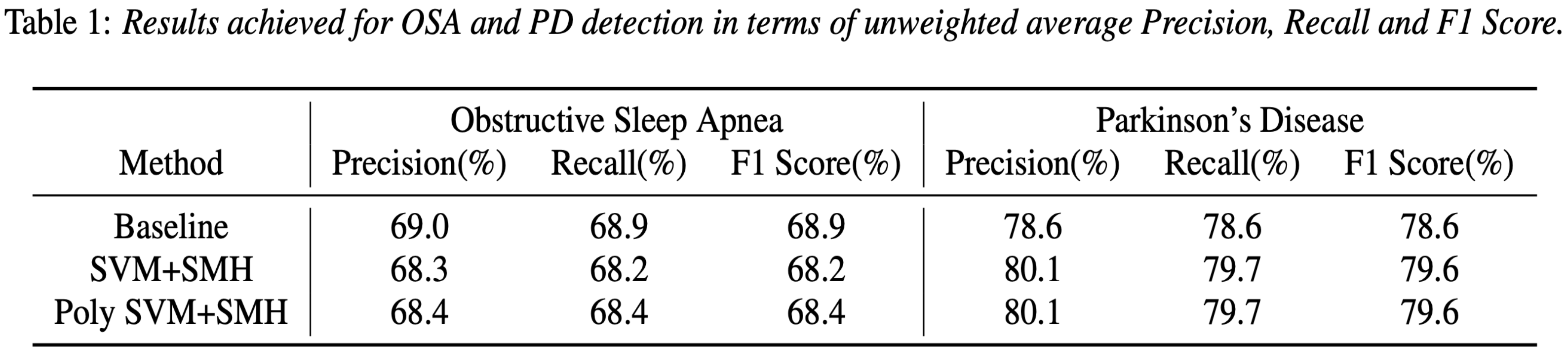


Table 1: *Results achieved for OSA and PD detection in terms of unweighted average Precision, Recall and F1 Score.*

Method	Obstructive Sleep Apnea Precision(%)	Recall(%)	F1 Score(%)	Parkinson's Disease Precision(%)	Recall(%)	F1 Score(%)
Baseline	69.0	68.9	68.9	78.6	78.6	78.6
SVM+SMH	68.3	68.2	68.2	80.1	79.7	79.6
Poly SVM+SMH	68.4	68.4	68.4	80.1	79.7	79.6

PPSVM - Security and Computational Cost

The security of this method comes from the security of the underlying protocols:

- Components of the method that use HE are secure under the RLWE problem.
- MPC-based components that use Secret Sharing or Garbled Circuits assume the *semi-honest* model.

For an SVM with:

- 1432 support vectors;
- 109 features-long vectors, hashed with *k = 4 & mpc = 32,* yielding SMH vectors of size *6976.*

Our method takes **~600m**s and **3MB** of bandwidth to compute **a single prediction** on a laptop with an Intel Core Quad-Core i5 CPU @ 1.4GHz and 16GB of RAM.

Other works: Privacy-preserving BFCC Extraction

- Work done by Thaine et al., 2019 [26].
- The authors show that while it is hard to extract Mel Frequency Cepstral Coefficients (MFCCs) privately, it is possible to replace them with Bark Frequency Cepstral Coefficients (BFCCs) with little accuracy degradation for an ASR system.
- The BFCC extraction is implemented using Homomorphic Encryption (BFV scheme implemented in PALISADE).
- Although not directly related, this is an important first step towards the private extraction of features for paralinguistic and extralinguistic tasks.

Other works: Privacy-aware Gender recognition

1. In a very different approach, Nelus et al., 2018 & 2019 [27], have conducted a series of works with the goal of determining a speaker's gender while trying to hide the speaker's identity.

- This has been done using:
  - Generative Adversarial Networks;
  - Siamese Neural Networks.

- It is shown that speaker identification accuracy significantly drops while it is still possible to determine gender.

- Although not perfectly secure, this work is an interesting take on the problem of privacy in speech processing.

Conclusions

- The problem of privacy in speech processing goes well beyond identity due to the amount of personal information that speech contains.
- Speech has the potential to become an inexpensive and non-intrusive biomarker for speech affecting diseases.
- Current societal concerns on privacy (e.g. EU's GDPR, California's CCPA, …), demand private solutions for this type of applications.
- Very few and limited contributions exist in relation to this topic.
- It is essential to increase the efforts to develop privacy-preserving techniques adapted to paralinguistic and extralinguistic tasks.

HE and MPC resources

- HE libraries:
  - SEAL - https://www.microsoft.com/en-us/research/project/microsoft-seal/
  - HElib - https://github.com/homenc/HElib
  - TFHE - https://github.com/tfhe/tfhe
  - PALISADE - https://palisade-crypto.org

- MPC libraries:
  - ABY - https://github.com/encryptogroup/ABY
  - MP-SPDZ - https://github.com/data61/MP-SPDZ
  - Awesome MPC - https://github.com/rdragos/awesome-mpc – Comprehensive list of courses, tutorials, software and workshops related with MPC.

- Tensor libraries working over HE and/or MPC backends:
  - nGraphHE - https://github.com/NervanaSystems/he-transformer
  - tf-encrypted - https://github.com/tf-encrypted/tf-encrypted
  - PySyft - https://github.com/OpenMined/PySyft
  - CrypTen - https://github.com/facebookresearch/crypten

References

[1] L. Moro-Velazquez, J. Villalba and N. Dehak - Using X-Vectors to Automatically Detect Parkinson's Disease from Speech. In ICASSP, 2020.

[2] Perero-Codosero, J. M., et al. - Modeling Obstructive Sleep Apnea Voices Using Deep Neural Network Embeddings and Domain-Adversarial Training. In STSP, 2019.

[3] Pompili, A., Rolland, T. and Abad, A. - The INESC-ID Multi-Modal System for the ADReSS 2020 Challenge. In arXiv preprint arXiv:2005.14646, 2020.

[4] Brakerski, Z., Gentry, C., and Vaikuntanathan, V. - (Leveled) fully homomorphic encryption without bootstrapping. In TOCT, 6 (3), 1-36, 2014.

[5] Fan, J., and Vercauteren, F. - Somewhat Practical Fully Homomorphic Encryption. In IACR Cryptol. ePrint Arch., 2012, 144.

[6] Cheon, J. H., Kim, A., Kim, M., and Song, Y. - Homomorphic encryption for arithmetic of approximate numbers. In Asiacrypt , 2017.

[7] Lyubashevsky, V., Peikert, C., and Regev, O.. On ideal lattices and learning with errors over rings. In Eurocrypt, 2010.

[8] Chase, M., Chen, H., Ding, J., Goldwasser, S., Gorbunov, S., Hoffstein, J.,et al. - Security of homomorphic encryption. *HomomorphicEncryption.org, Tech. Rep,* 2017.

[9] Goldreich, O., Micali, S., & Wigderson, A. - How to play any mental game. 1987.

[10] Ben-Or, M., Goldwasser, S., & Wigderson, A. - Completeness theorems for non-cryptographic fault-tolerant distributed computation. 1988.

[11] Yao, A. C. C. - How to generate and exchange secrets. 1986.

[12]] Jiménez, A. and Raj, B. and Portêlo, J. and Trancoso, I. - Secure Modular Hashing. In WIFS, 2015.

[13] Bringer, J., Chabanne, H., Favre, M. Patey, A., Schneider, T. and Zohner, M. - GSHADE: Faster Privacy-Preserving Distance Computation and Biometric Identification. In IH&MMSEC, 2014.

[14] Gilad-Bachrach, R., Dowlin, N., Laine, K., Lauter, K., Naehrig, M., and Wernsing, J. - Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. In ICML, 2016.

[15] Chabanne, H., de Wargny, A., Milgram, J., Morel, C., and Prouff, E. - Privacy-Preserving Classification on Deep Neural Network. In IACR Cryptol. ePrint Arch., 35, 2017.

[16] Hesamifard, E., Takabi, H., and Ghasemi, M. - CryptoDL: Deep neural networks over encrypted data. arXiv preprint arXiv:1711.05189, 2017.

[17] Dias, M, Abad, A., Trancoso, I. – Exploring Hashing and Cryptonet based approaches for Privacy-preserving Speech Emotion Recognition. In ICASSP, 2018.

[18] Teixeira, F., Abad, A., Trancoso, I. – Patient Privacy in Paralinguistic Tasks. In Interspeech, 2018.

[19] Teixeira, F., Abad, A., Trancoso, I. – Privacy-preserving Paralinguistic Tasks. In ICASSP, 2019.

[20] Jiménez, A. and Raj, B. – Privacy Preserving Distance Computation Using Somewhat-Trusted Parties. In ICASSP, 2017.

[21] Botelho, M. C. and Trancoso, I. and Abad, A. and Paiva, T. - Speech as a biomarker for obstructive sleep apnea detection. ICASSP, 2019.

[22] Orozco, J. R., Arias-Londoño, J. D., Vargas-Bonilla, J., González-Rátiva, M. and Nöth, E. - New Spanish speech corpus database for the analysis of people suffering from Parkinsons disease. In LREC, 2014.

[23] Pompili, A,, et al, - Automatic detection of parkinson's disease: An experimental analysis of common speech production tasks used for diagnosis. In TSD, 2017.

[24] Microsoft SEAL (release 3.5). URL: https://github.com/Microsoft/SEAL.

[25] Demmler, Daniel, Thomas Schneider, and Michael Zohner - ABY- A framework for efficient mixed-protocol secure two-party computation. In NDSS, 2015.

[26] Thaine, P. and Penn, G. - Extracting Mel-Frequency and Bark-Frequency Cepstral Coefficients from Encrypted Signals. In Interspeech, 2019.

[27] Nelus, A. and Rainer, M. - Privacy-aware Feature Extraction for Gender Discrimination versus Speaker Identification. In ICASSP, 2019.

inesc id lisboa

DEFINING TECHNOLOGY

Thank you!

francisco.s.teixeira@tecnico.ulisboa.pt