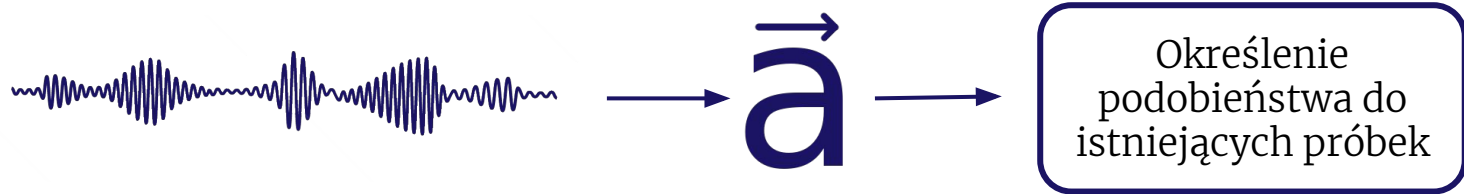


Głos jako hasło, czyli weryfikacja za pomocą głosu

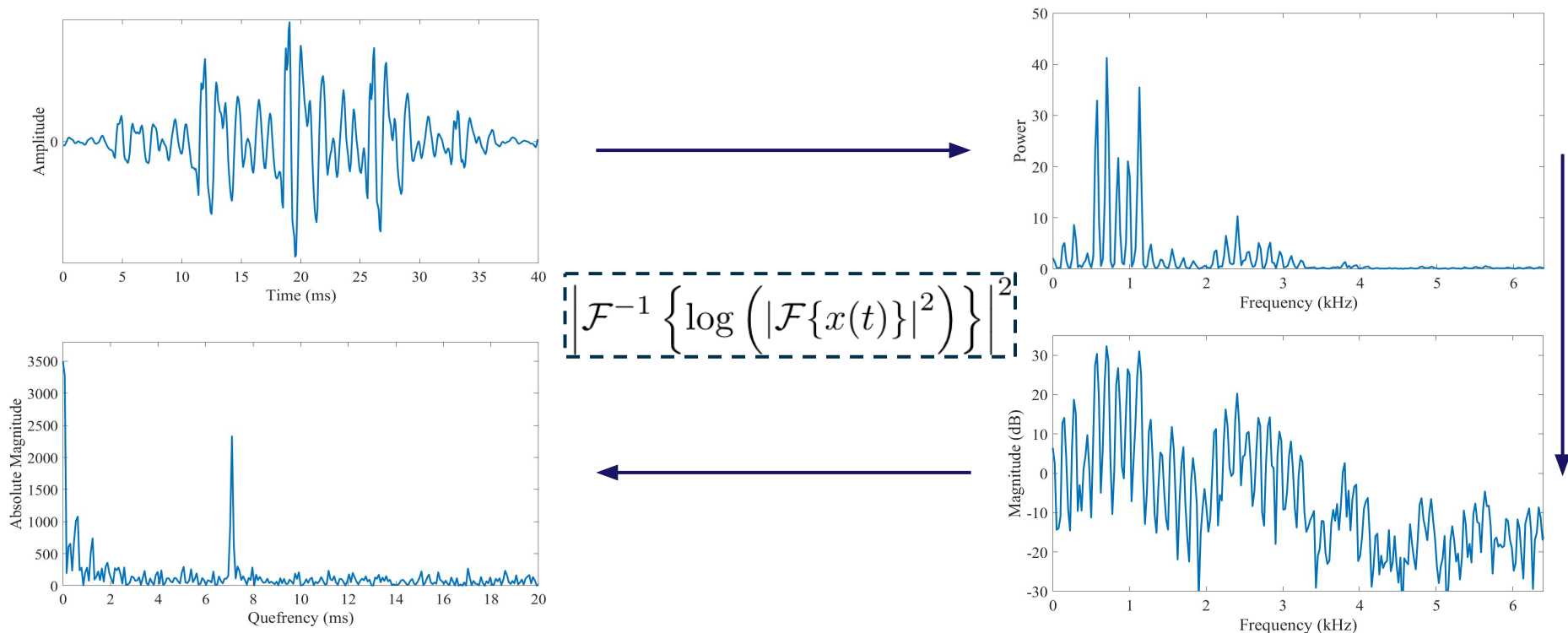


Filip Szympliński

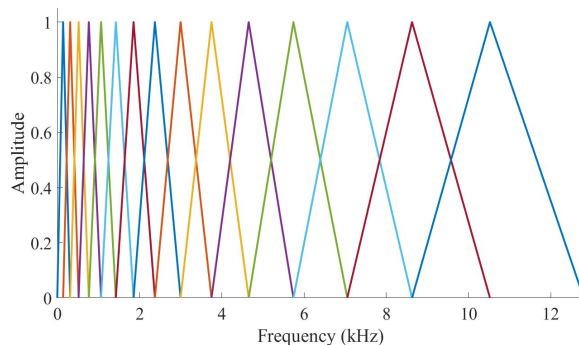
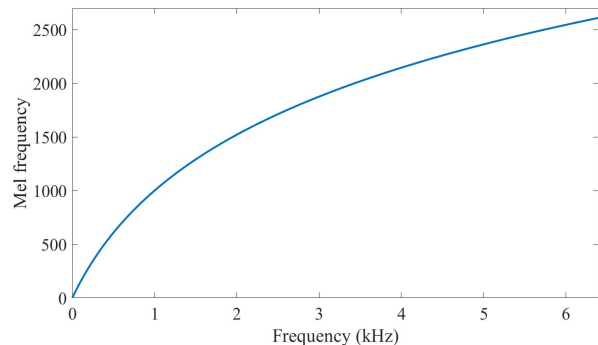
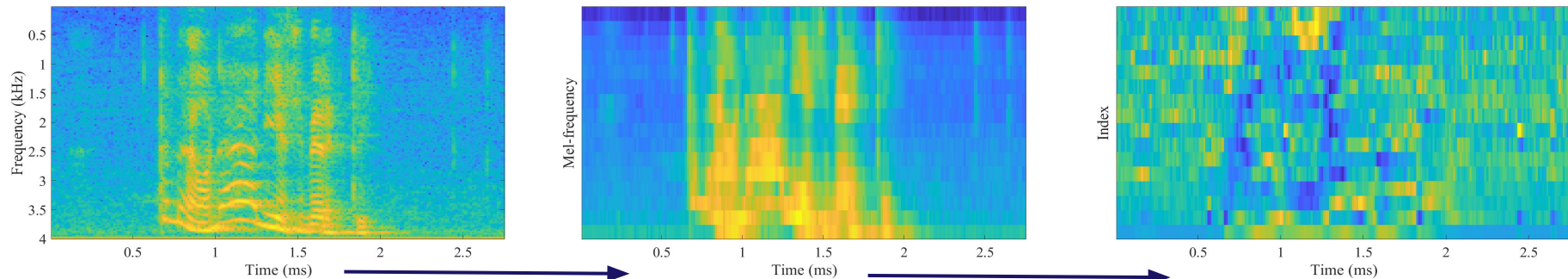
Jak podejść do problemu?



Preprocessing danych – MFCC



Preprocessing danych – MFCC

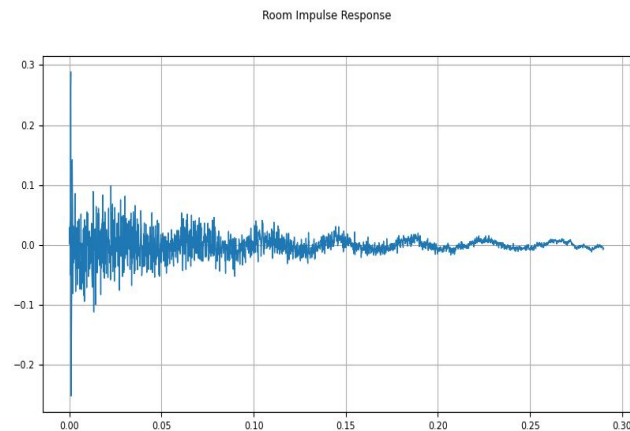
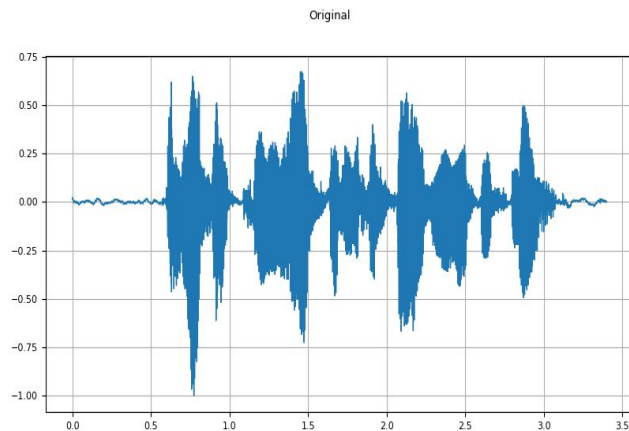


$$2595 \log_{10} \left(1 + \frac{f}{700} \right)$$

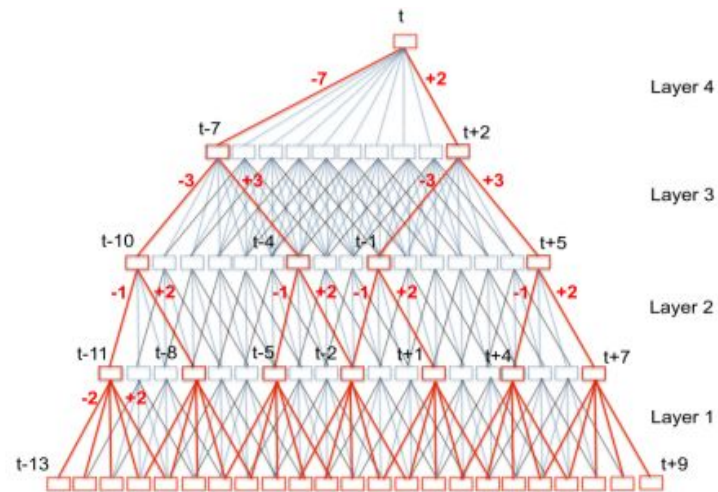
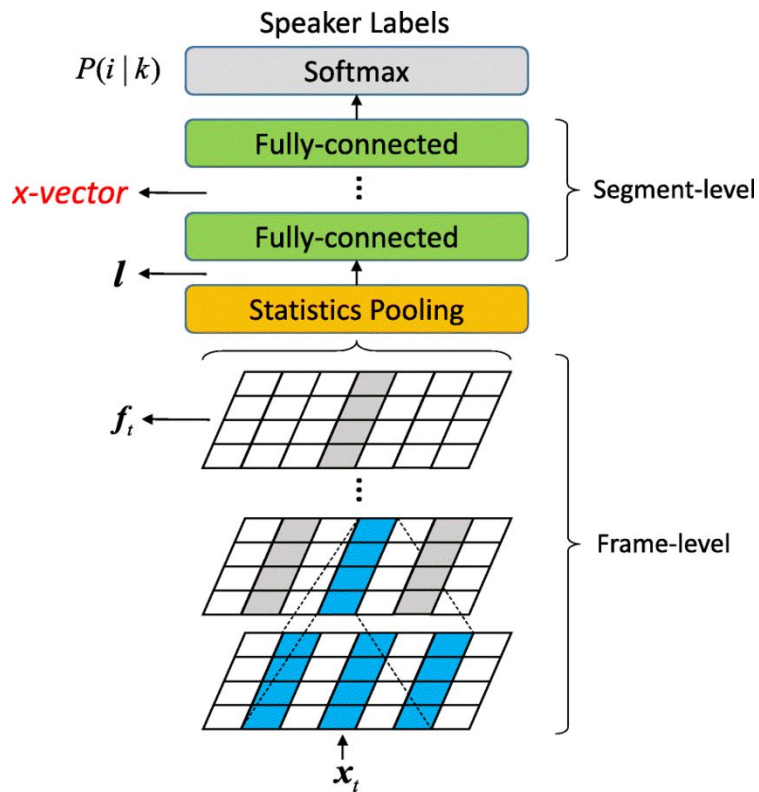
$$u_k = \sum_{h=f_{k-1}+1}^{f_{k+1}-1} w_{k,h} |x_h|^2$$

Preprocessing danych – data augmentation

- Dodawanie sumów (biały, różowy, gaussowski, itd.)
- Dodawanie dźwięków w tle
- Room Impulse Response (RIR)
- Przysłanianie części częstotliwości/fragmentu sygnału
- Przyspieszenie/spowolnienie sygnału
- ...

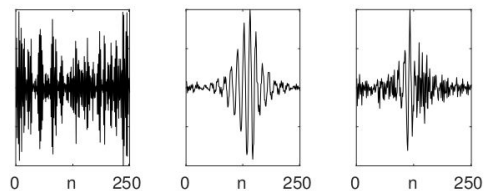


x-vector

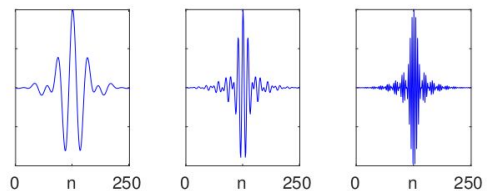


Inne metody embeddingu dźwięku

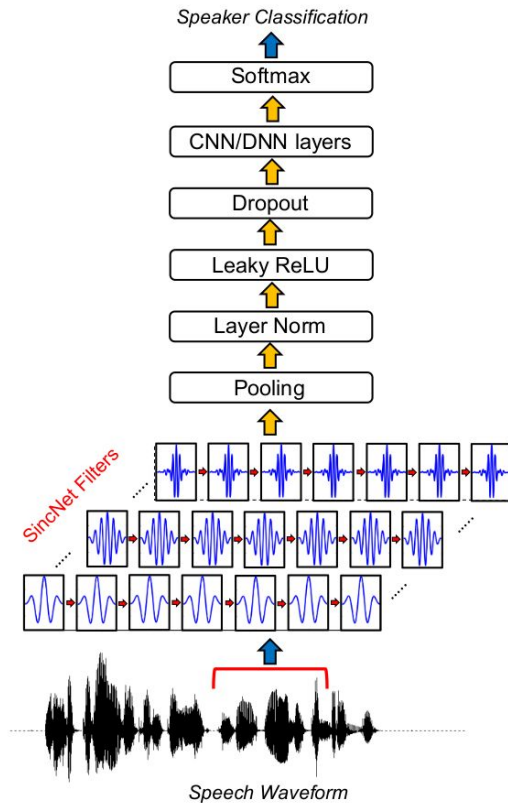
- i-vector
- CNN
- SincNet
- ...



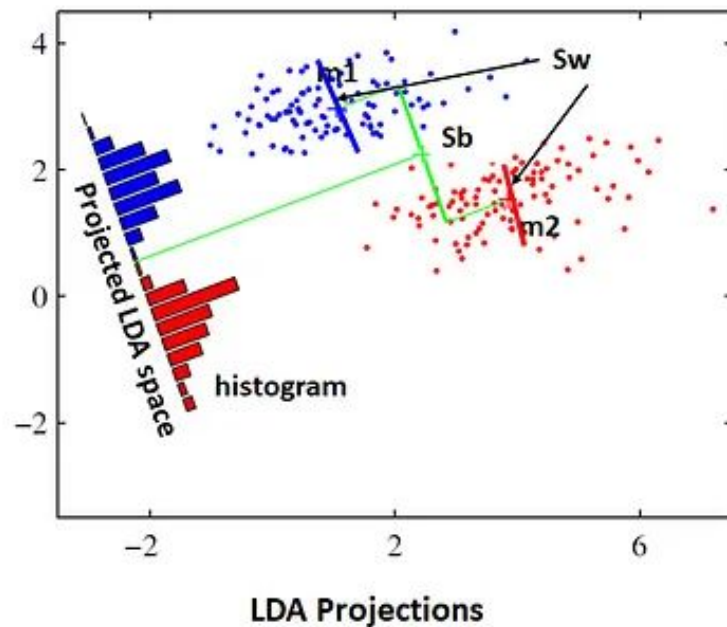
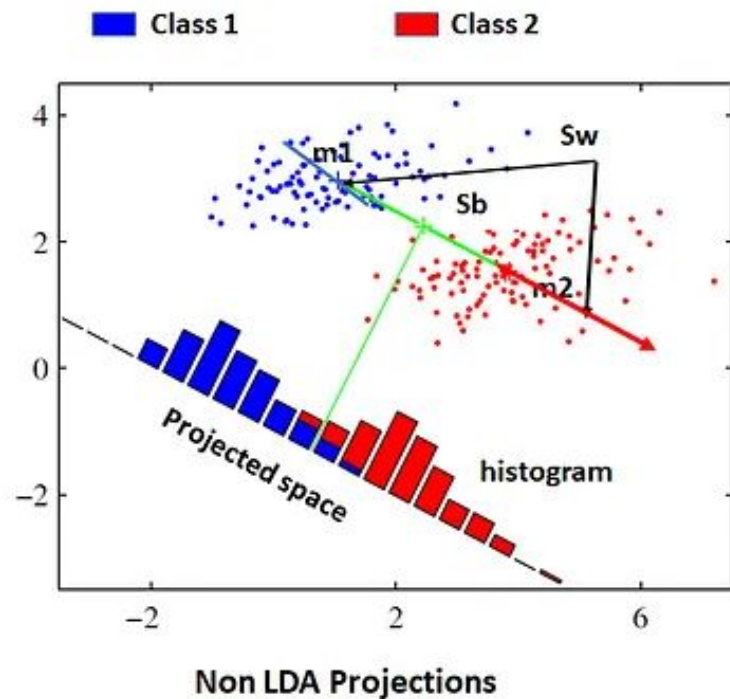
(a) CNN Filters



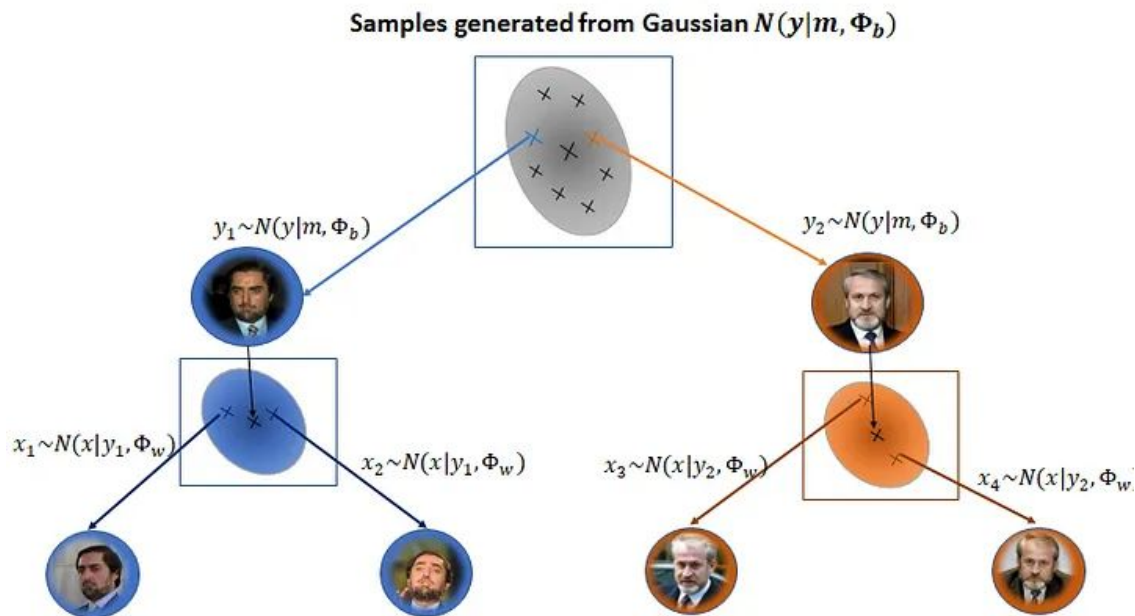
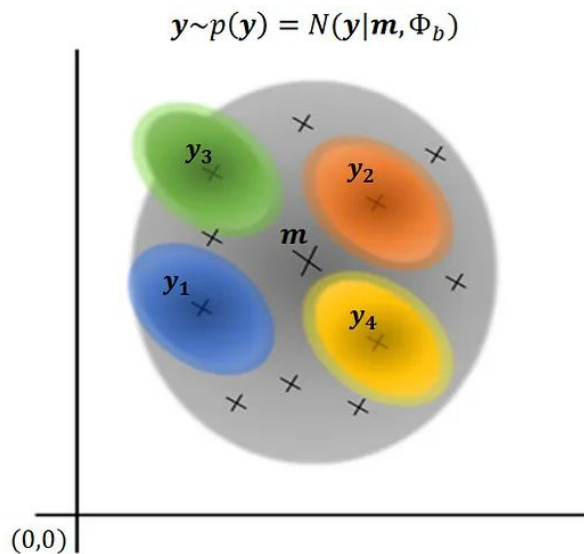
(b) SincNet Filters



LDA



PLDA



PLDA

$$V^T \Phi_w V = I$$

$$V^T \Phi_b V = \Psi$$

By defining, $A = V^{-T}$, we obtain

$$\Phi_w = AA^T$$

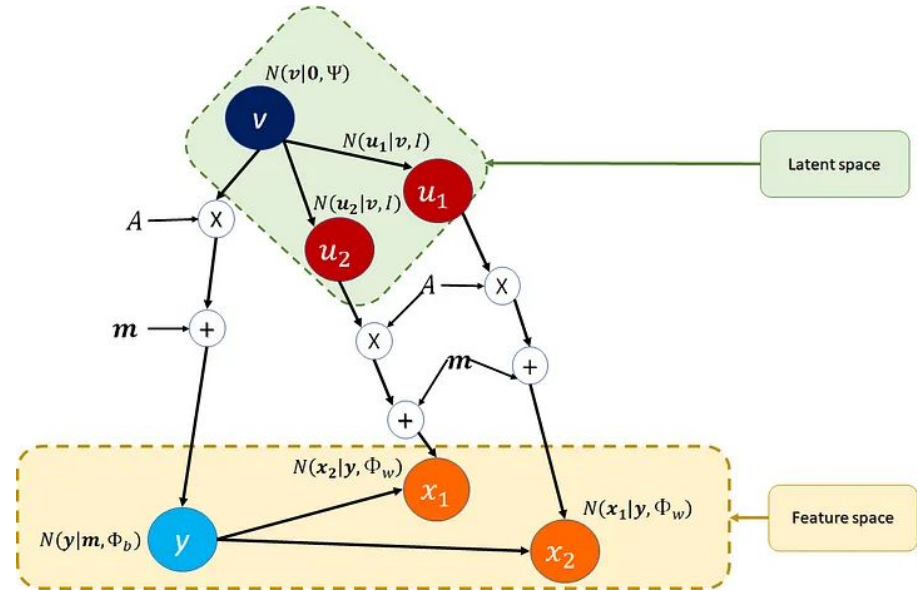
$$\Phi_b = A\Psi A^T$$

$$\mathbf{u} \sim N(\cdot | \mathbf{v}, I)$$

$$\mathbf{v} \sim N(\cdot | \mathbf{0}, \Psi)$$

$$\mathbf{y} = \mathbf{m} + A\mathbf{v}$$

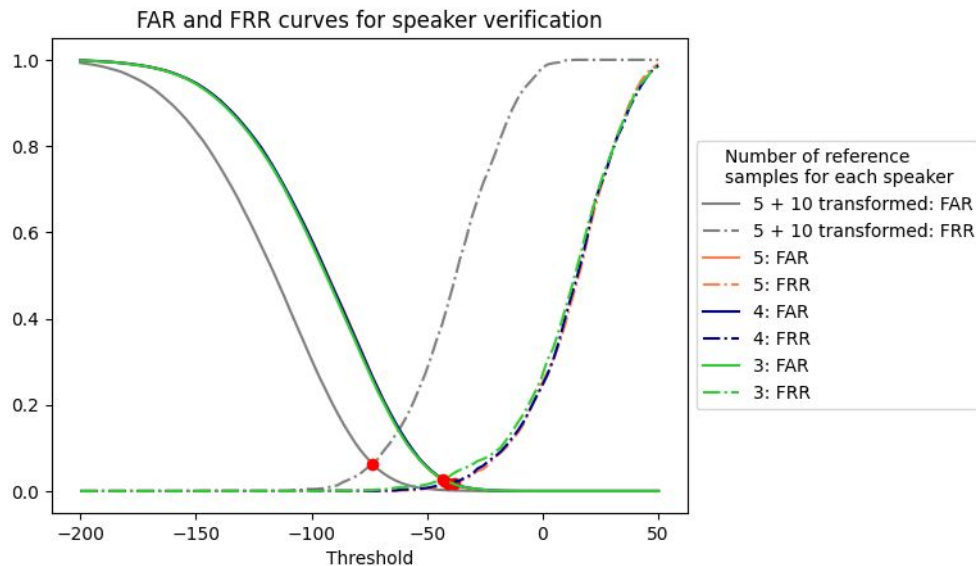
$$\mathbf{x} = \mathbf{m} + A\mathbf{u}$$



Ewaluacja systemu do weryfikacji głosu

$$FAR = \frac{\text{liczba próbek niewłaściwie zaakceptowana}}{\text{liczba próbek nienależąca do danej osoby}}$$

$$FRR = \frac{\text{liczba próbek niepoprawnie odrzucona}}{\text{liczba próbek należąca do danej osoby}}$$



Literatura i dodatkowe materiały

- **MFCC:**
 - <https://wiki.aalto.fi/display/ITSP/Cepstrum+and+MFCC>
 - https://www.youtube.com/watch?v=4_SH2nfbQZ8
- **AUDIO DATA AUGMENTATION:**
 - https://pytorch.org/audio/master/tutorials/audio_data_augmentation_tutorial.html
- **x-vector:**
 - <https://www.mathworks.com/help/audio/ug/speaker-recognition-using-x-vectors.html>
 - David Snyder, Daniel Garcia-Romero, Daniel Povey, and Sanjeev Khudanpur.
“Deep Neural Network Embeddings for Text-Independent Speaker Verification”
In Proc. Interspeech 2017, pages 999–1003, 2017
- **PLDA:**
 - <https://towardsdatascience.com/probabilistic-linear-discriminant-analysis-plda-explained-253b5effb96>
 - Sergey Ioffe. “Probabilistic linear discriminant analysis” In European Conference on Computer Vision, 2006.
- **Kanał na YT o analizie i przetwarzaniu dźwięku:**
 - <https://www.youtube.com/@ValerioVelardoTheSoundofAI>