

# Проект «Voice2Vec»

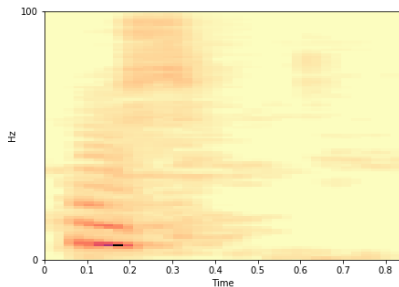
Таскынов Ануар, 517 группа

Курс «Современные методы распознавания и синтеза речи»

9 июня 2018 г.

Хочется иметь возможность заполучить вектора для голосов людей, чтобы распознавать одинаковые голоса.

Данные взяты из сайта [forvo.com](https://forvo.com)  
Скачано около 8000 произношений.

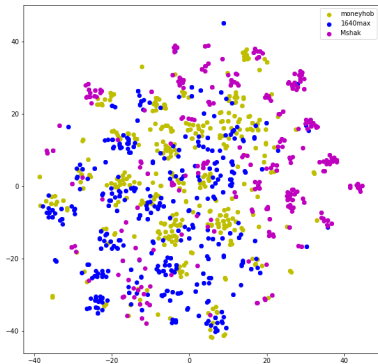


Пользователь `yulia_m`, произношение  
слова «кус».

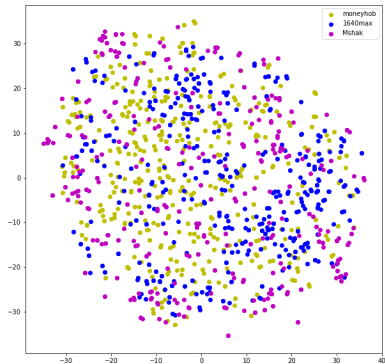


Пусть  $f(x_i^a)$ ,  $f(x_i^b)$  — эмбединги для речи  $i$ -го человека,  $f(y_i)$  — эмбединги для речи других людей (не  $i$ -го). Тогда функция потерь будет следующей:

$$L = \frac{1}{N} \sum_{i=1}^N \left[ \left\| f(x_i^a) - f(x_i^b) \right\|_2^2 - \left\| f(x_i^a) - f(y_i) \right\|_2^2 + \alpha \right]_+.$$



Векторное представление.



Развёрнутый mel.

*Florian Schroff, Dmitry Kalenichenko, James Philbin.* FaceNet: A Unified Embedding for Face Recognition and Clustering, 2015.