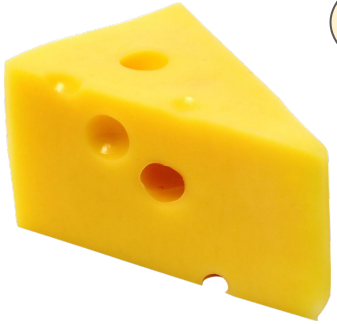
SpeechLock

Идея

Hello, I'm
Tvorozhek.
Let me in, please.

Не еш меня,
подумой

same person ✓



тыгыдык-тыгыдык
гыы я лошадка

Что тут вообще
происходит?

not same person ✘

Датасет

http://forvo.com

Скачано 7406 произношений.

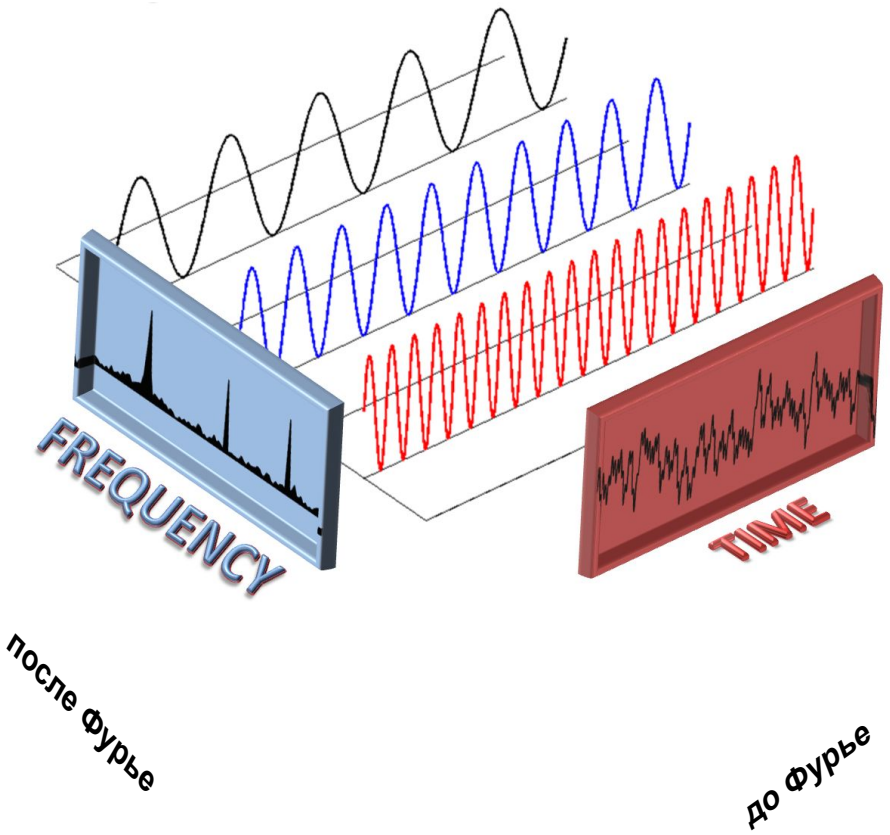
Самые популярные слова:

1	66
привет	52
здравствуйте	27
машина	26
русский	21
Китай	20
медведь	19
счастье	19
котёнок	19

Первая попытка

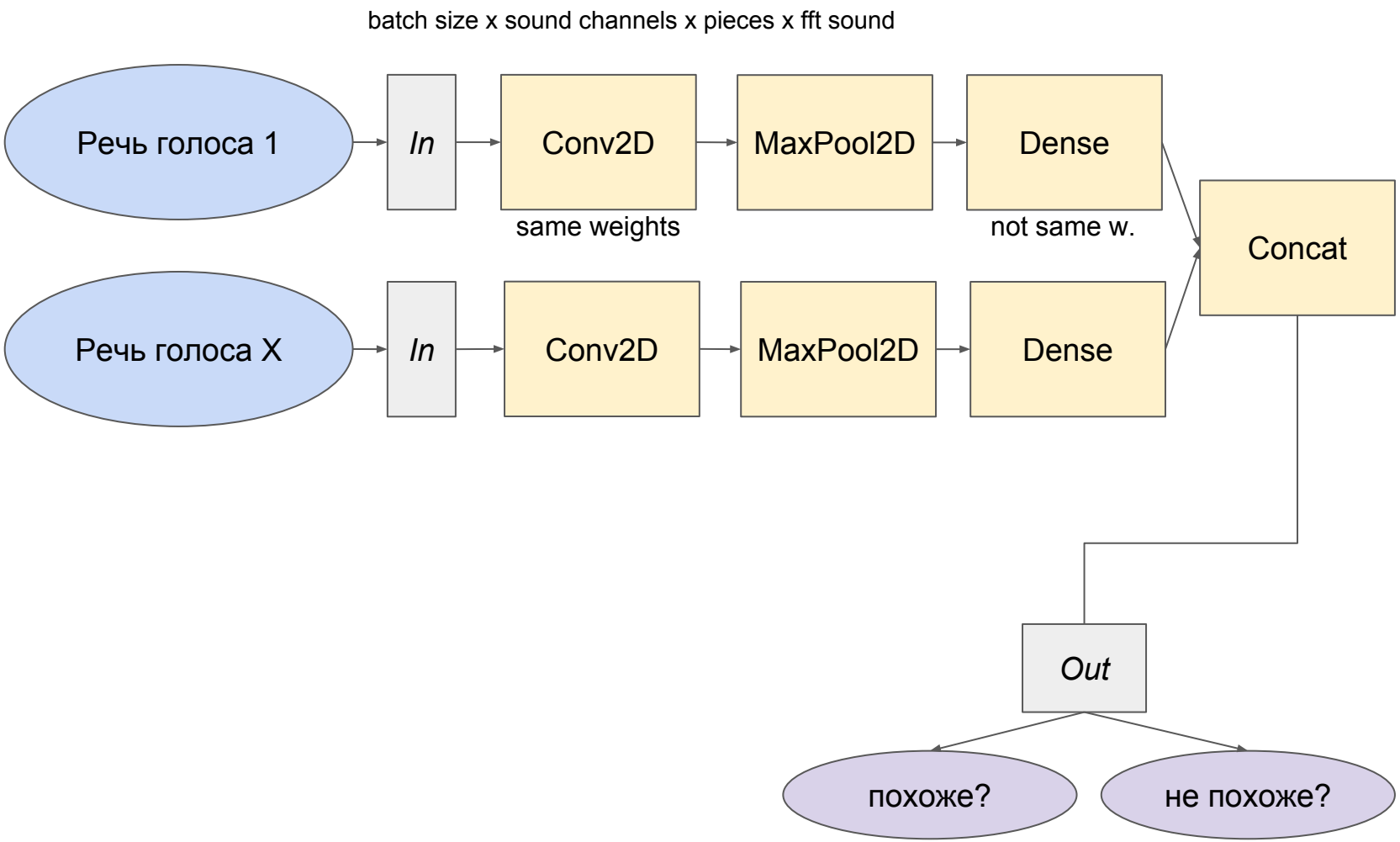
Предобработка голосов

1. Убрали все звуки длиною меньше секунды
2. Выбрали случайную секунду
3. Нарезали на кусочки по 25 мс с пересечениями по 15 мс
4. Применили преобразование Фурье к каждому кусочку

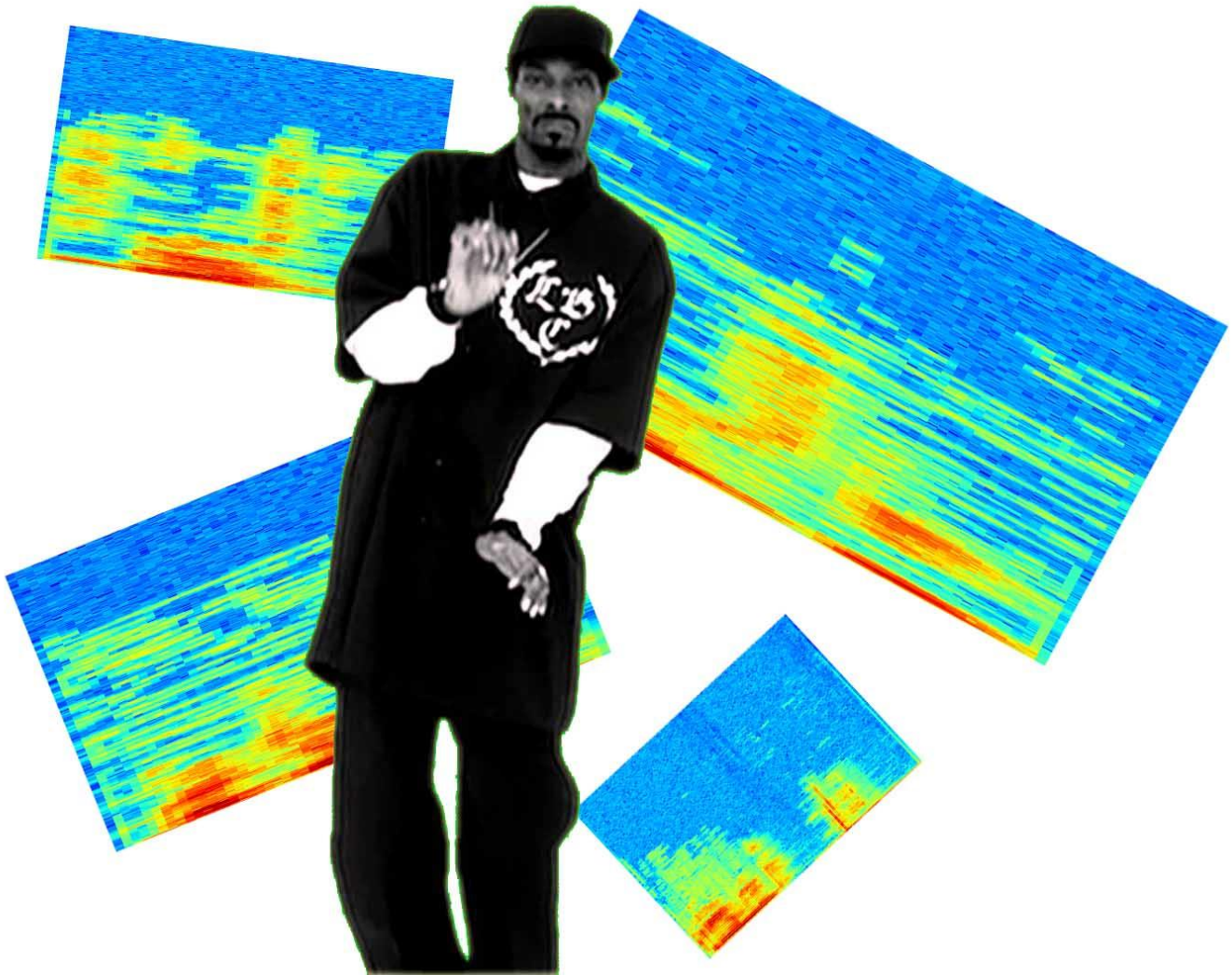


Преобразование Фурье позволяет получить “фичи” голосов

Структура нейросети



Вторая попытка

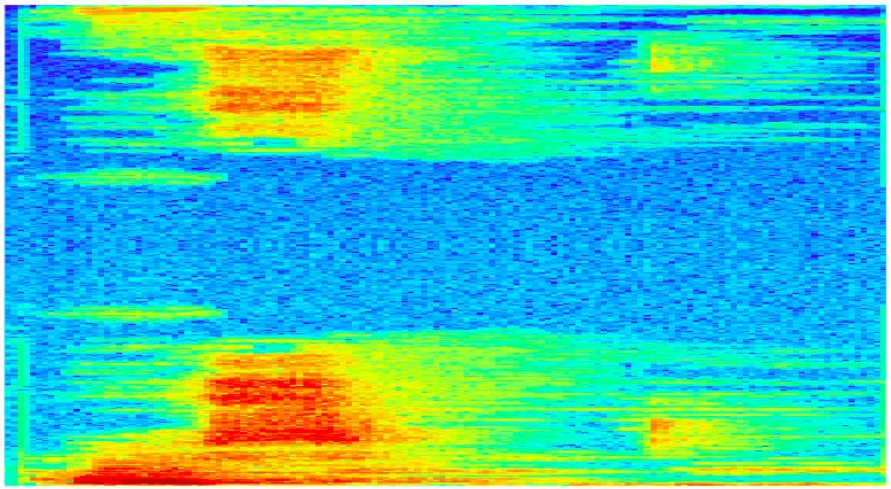


Спектрограммы

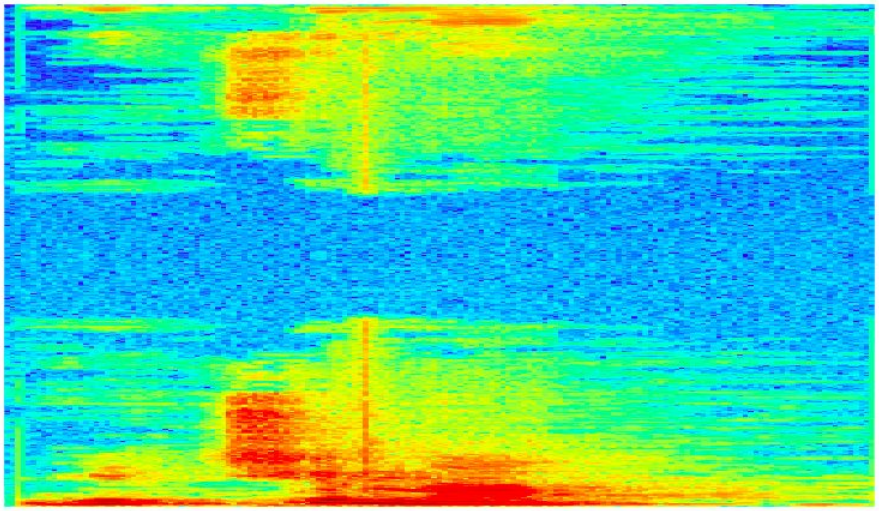
По горизонтали время.

По вертикали частоты.

Цвет означает амплитуды, соответствующие определенной частоте.

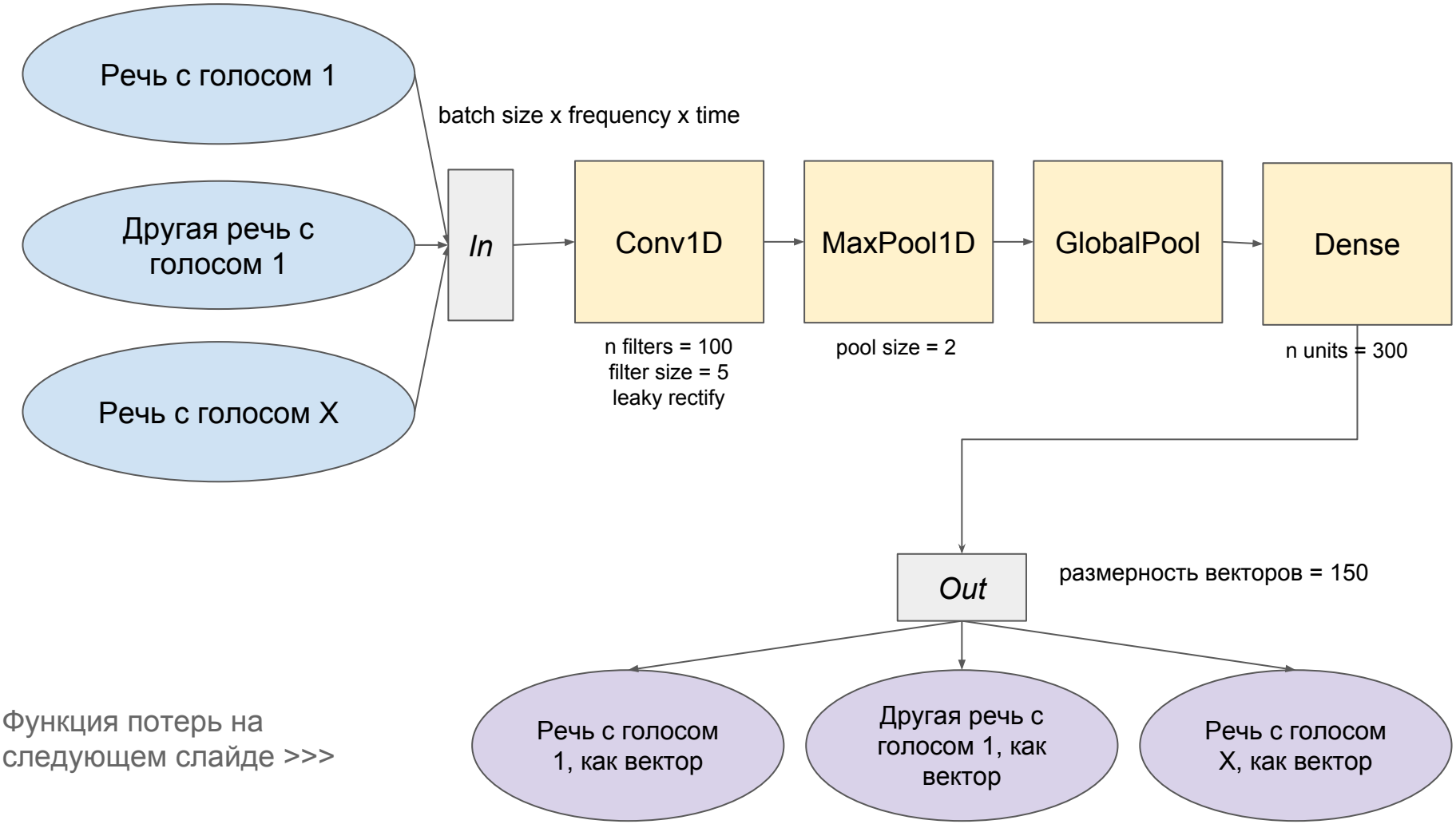


yulia_m: кусь



yulia_m: ветчина

Структура нейросети



Функция потерь на следующем слайде >>>

Функция потерь

$$L = max(f(x_1, y_1) + \alpha, 0) - max(f(x_1, x_2) + \alpha, 0)$$

где x_1 – речь голосом 1, x_2 – другая речь голосом 1, y_1 – речь голосом X,
f – функция различия между векторами (например, евклидово расстояние),
α – константа, $\alpha > 0$, $\alpha << 1$

Функция предсказания

$$P(x, y) = \cos(x, y) = \frac{(x, y)}{\|x\| \|y\|}$$

где x – речь голосом 1, y – речь голосом X

64%

AUC ROC

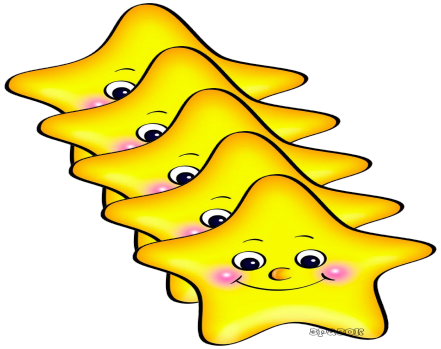
Похожие работы

FaceNet: A Unified Embedding for Face Recognition and Clustering https://arxiv.org/pdf/1503.03832v3.pdf

Recommending music on Spotify with deep learning http://benanne.github.io/2014/08/05/spotify-cnns.html

Ссылки

github.com/xenx/speech



Поставь звездочку!

Презентация на гитхабе pdf