Генерация голоса с учётом индивидуальных особенностей

Курбанов Рауф Эльшад оглы

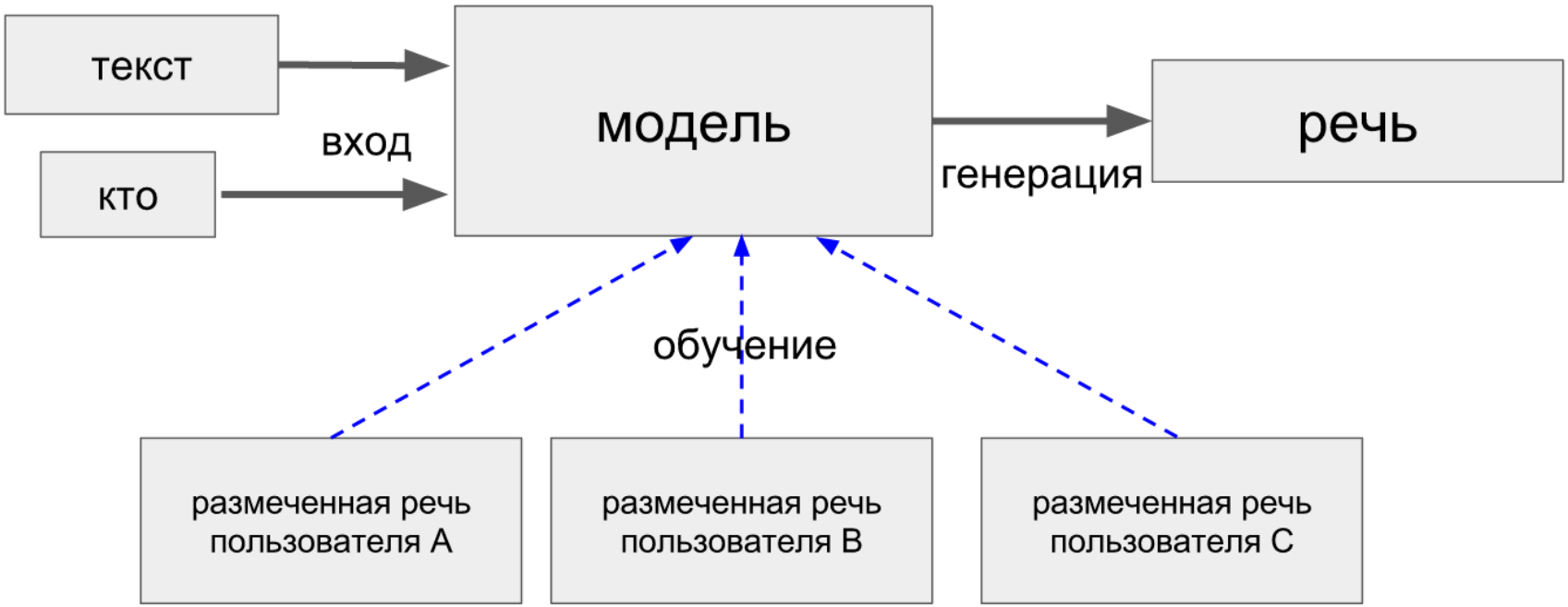
научный руководитель: А.А. Шпильман

СПб АУ НОЦНТ РАН

12 июня 2017 г.

Мотивация

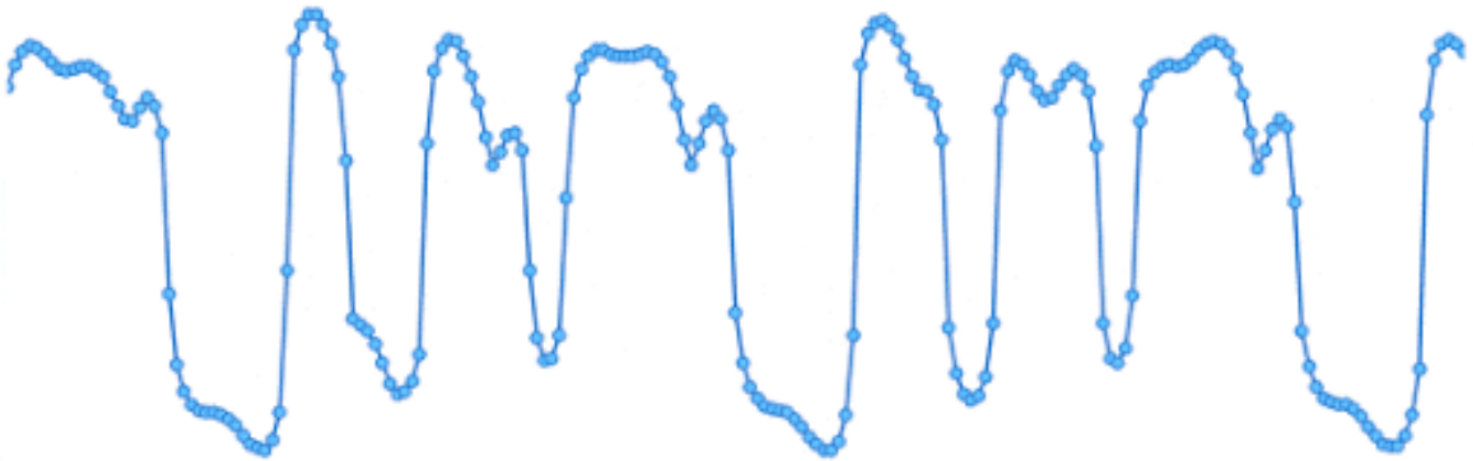
- Система, генерирующая персонализированный голос



Генеративные модели

Каждая точка x генерируется на основе T предыдущих

$$f(x_t) = F(x_{t-1}, \ldots, x_{t-T})$$

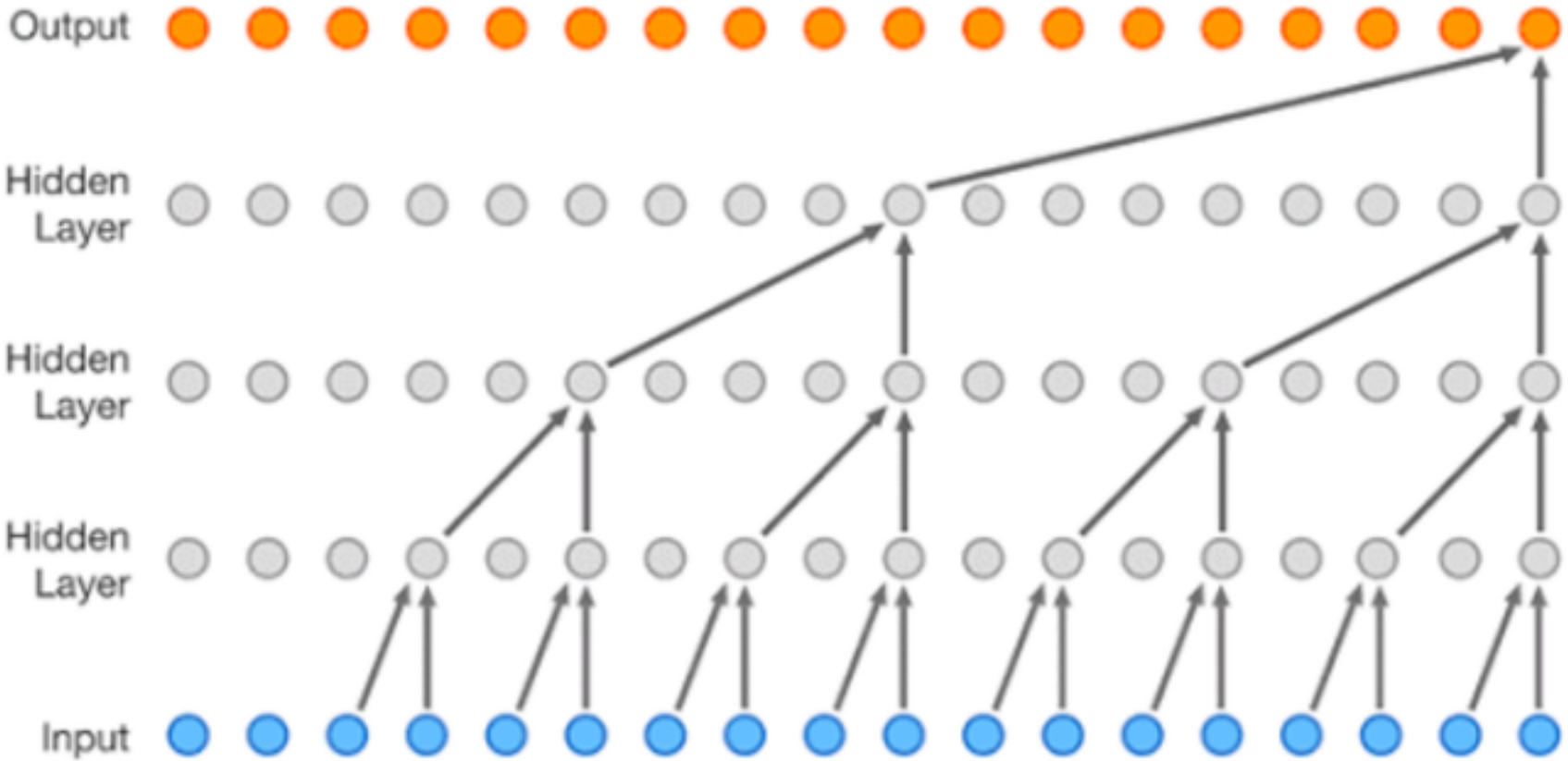


WaveNet [1]

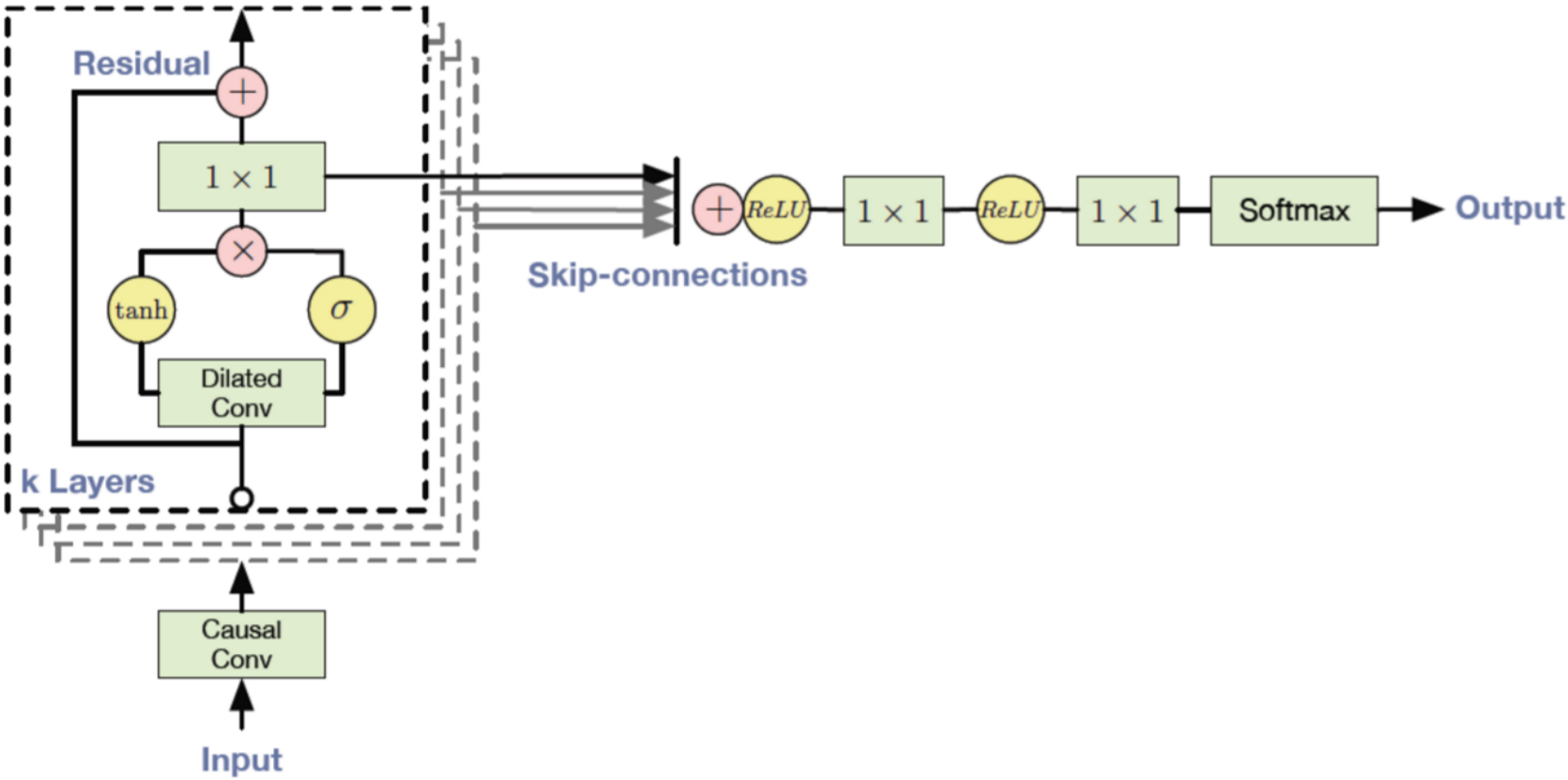
- Нейросетевая архитектура от Google Brain
- Новый ”state of the art” в генерации голоса
- Вычислительно лёгкая архитектура на основе дырвях свёрток

[1] Aäron van den Oord, Sander Dieleman, Heiga Zen et al. - 2016

Дырявые свёртки



Архитектура WaveNet



Генерация по условию

Для генерации по условию данные в WaveNet передаются по следующим каналам:

- Сырые данные
  - Временной ряд, цифровое представление голоса по времени
- Локальное условие
  - Временной ряд, той же длины что и данные. Качество, изменяющееся по времени
- Глобальное условие
  - Качество говорящего, не зависящее от времени. Не меняет своего значения в процессе обучения/генерации

Задачи

Цель: Реализовать модель для генерации голоса с учётом особенностей

Задачи:

- Реализовать WaveNet максимально придерживаясь описания из статьи:
  - Генерацию голоса без условия
  - Генерацию голоса по тексту
  - Генерацию голоса по тексту с условием
- Разработать признаки для генерации голоса
- Получить результаты генерации

Данные

Корпус VCTK

- 109 голосов
- 400 предложений каждый
- текст прилагается, но не выровнен
- высокое качество записи, не требует фильтрации

Реализация WaveNet

- Нужна точная и полная реализация архитектуры из статьи
- Существует открытая реализация, однако модификации для генерации по условию ожидаемы в сообществе
- Модификации реализованы в виде ответвления от самого поплуярного репозитория `https://github.com/Rauf-Kurbanov/tensorflow-wavenet/tree/local_conditioning`

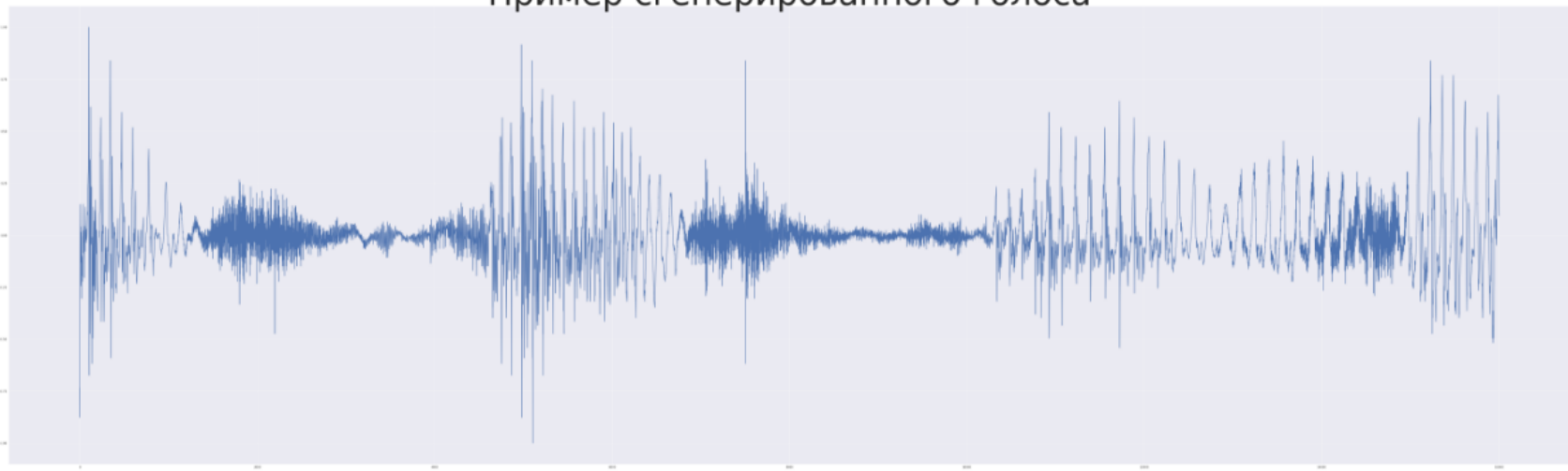
Разработанный каскад признаков

- Кодирование и выравнивание текста вдоль звука
  - ’The car’ -> ’TTTTTTThhhhhhhheeeeeee cccccaaaarrrr’
- Голос, сгенерированный открытым языковым API Yandex
- 34 признака для представления голоса
  - быстрое преобразование Фурье
  - спектральные коэффициенты
  - энергия, частота и т.д.

Примеры

Пример голоса из обучающей выборки

Пример сгенерированного голоса



Оценка качества генерации

Таблица: Сравнение натуральности генерируемых голосов

Субъективное предпочтение(%) в натуральности голоса			
Без условий	Глобальное условие: ID говорящего	Локальное условие: yandex-speech	Нет пред-почтения
27		**61**	11
34	**34**		31
	29	**56**	15

[0] https://goo.gl/forms/b32ayxpp8YdbFD373

Результаты

- Реализован WaveNet максимально придерживаясь описания из статьи
  - Доработана существующая генерация голоса без условия
  - Реализована генерация голоса по тексту
  - Реализована генерация голоса по тексту с условием
- Разработаны признаки для генерации голоса
- Получены результаты генерации

Wavenet: A generative model for raw audio / Aäron van den Oord, Sander Dieleman, Heiga Zen et al. // CoRR abs/1609.03499. — 2016.