Яндекс.ГолосовойВвод

Зона экрана спикера

Контент не располагать, желтый квадрат удалить после дизайна слайда

Поиск музыки по напеванию

Максим Семенов, erasedwalt@

Проблематика

Зона экрана спикера

Контент не располагать, желтый квадрат удалить после дизайна слайда

Реализовать поиск песни по напеванию её мелодии (query by humming/singing).

Небольшое напевание песни
(возможно без конкретных слов)

Nirvana — Smells Like Teen Spirit

Результат

Метрика для измерения качества работы алгоритма

Зона экрана спикера

Контент не располагать, желтый квадрат удалить после дизайна слайда

Будем предполагать, что на запрос алгоритм выдает ранжированный список песен.

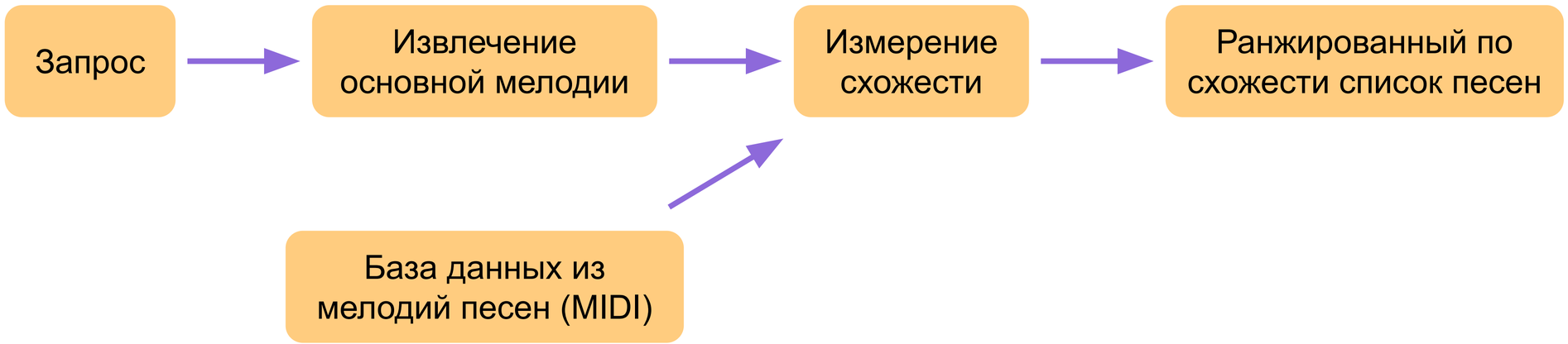
где n — количество песен в датасете, — i-ое напевание, a — позиция соответствующей песни по порядку ранжирования.

Существующие статьи по теме

Зона экрана спикера

Контент не располагать, желтый квадрат удалить после дизайна слайда

Большинство статей предлагают решать задачу по следующей схеме:

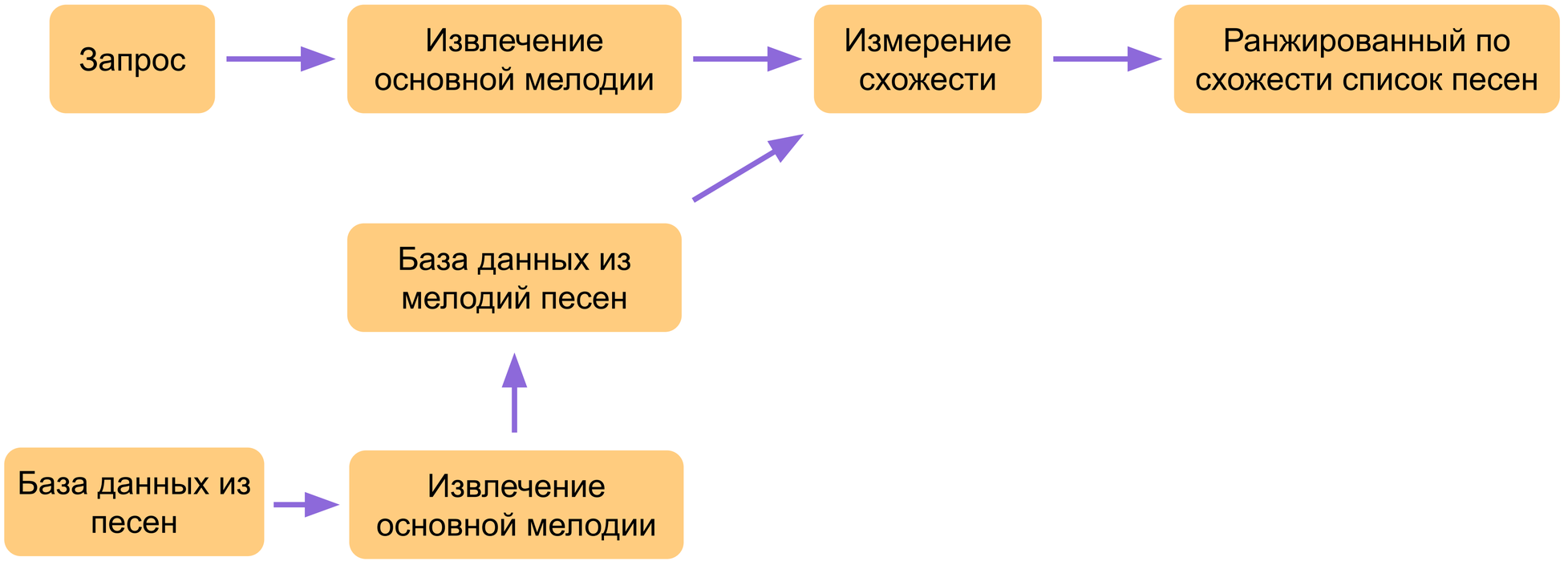


Существующие статьи по теме

Зона экрана спикера

Контент не располагать, желтый квадрат удалить после дизайна слайда

Для реальных условий схема была переделана таким образом:



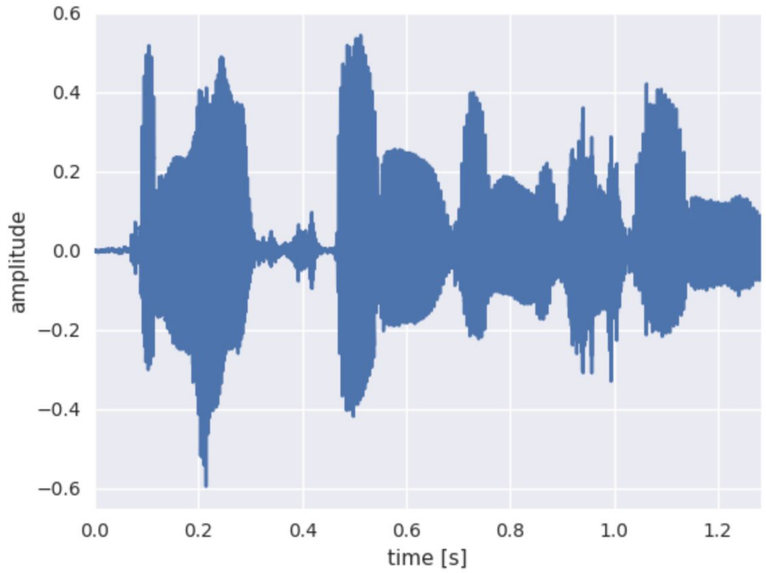
Извлечение основной мелодии

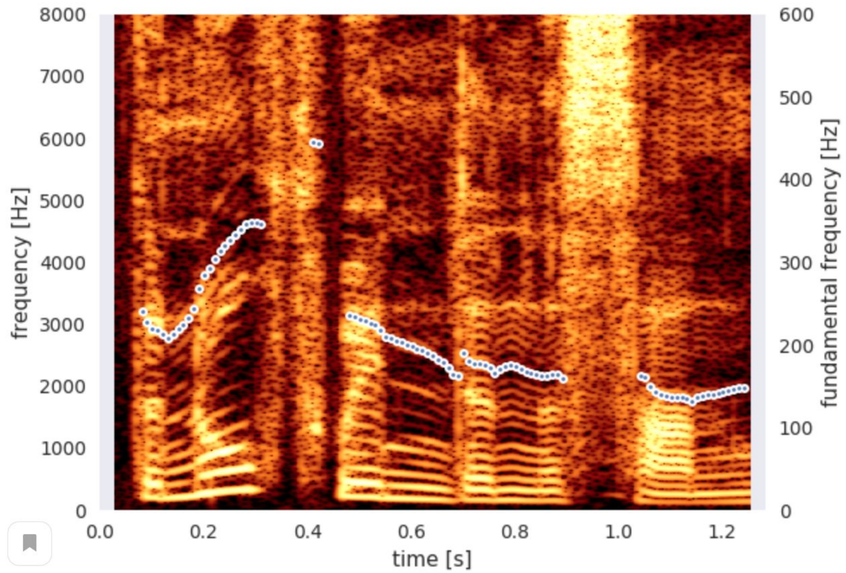
Зона экрана спикера

Контент не располагать, желтый квадрат удалить после дизайна слайда

Для извлечения основной мелодии предлагается использовать алгоритмы, вычисляющие фундаментальную (основную) частоту.

В экспериментах были использованы алгоритмы YiN, pYiN, Praat, Kaldi Pitch и Crepe.



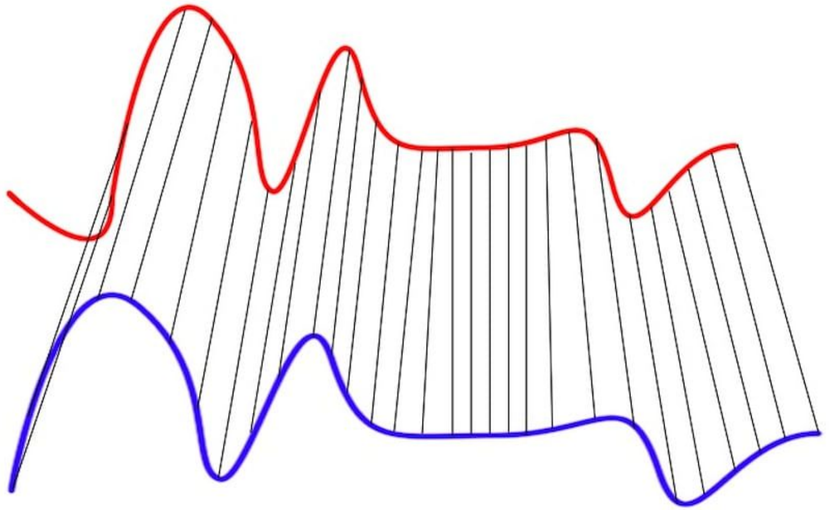


Измерение схожести мелодий

Зона экрана спикера

Контент не располагать, желтый квадрат удалить после дизайна слайда

Схожесть двух временных рядов предлагается измерять с помощью DTW (Dynamic Time Warping) и Wasserstain distance. Первый алгоритм более точный, но долгий по сравнению со вторым.



DTW

Проблемы

Зона экрана спикера

Контент не располагать, желтый квадрат удалить после дизайна слайда

- Не все люди хорошо умеют петь, из-за чего тональность напевания может не совпадать с тональностью песни.

- Некоторые алгоритмы по извлечению основной частоты работают долго.

Результат

С различными гиперпараметрами и алгоритмами на 135 песнях получился MRR от 0.1 до 0.15.

Более современный подход

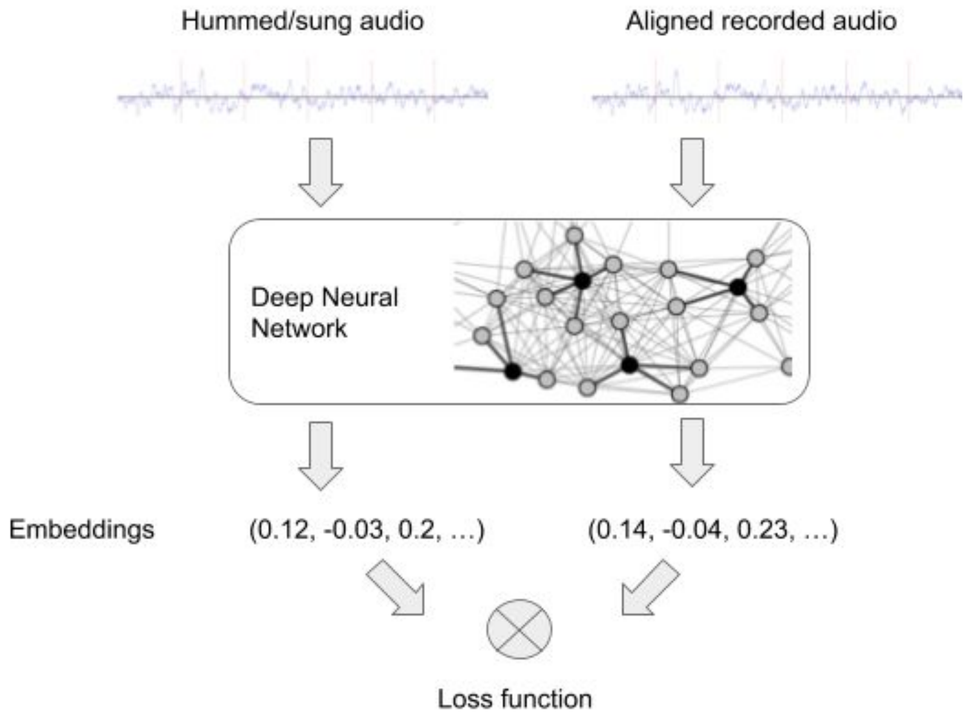
Зона экрана спикера

Контент не располагать, желтый квадрат удалить после дизайна слайда

Система поиска по напеванию реализована в Google. Статьи про это у них нет, но есть два поста в их блоге.

Для создания эмбеддингов используется сиамская сеть.

Подробностей обучения нет, но известно, что было использовано небольшое улучшение Triplet Loss’a.



Эксперименты

Для обучения глубокой сети оказалось мало данных. Порядка 4 тысяч напеваний и около 150 соответствующих им песен. Но попробовать похожий с Google подход хотелось.

Лучший результат на около 30 песнях получился 0.18 MRR.

Что делать дальше

Зона экрана спикера

Контент не располагать, желтый квадрат удалить после дизайна слайда

Так как данных не хватает, мы заказали разметку в Толоке.

Попробовать решить задачу по определению каверов песен, эта задача похожа на нашу, и по ней есть современные статьи, надеемся почерпнуть оттуда идеи.

Зона экрана спикера

Контент не располагать, желтый квадрат удалить после дизайна слайда

Спасибо за внимание

Максим Семенов

Стажер-разработчик


erasedwalt@yandex-team.ru


@derpsx