# Шумоподавление для SaluteJazz: Делаем ваши беседы более приятными

Лысова Яна

DL Engineer @ SberDevices

SBER DEVICES

# План

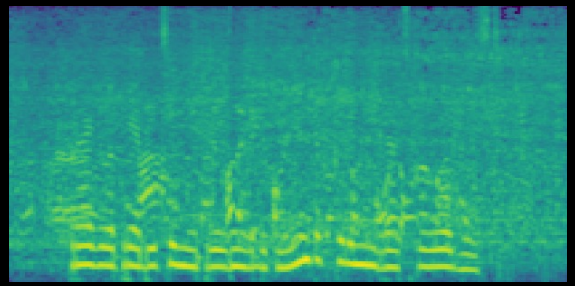
Сформулируем задачу шумоподавления

Выбираем архитектуру нейронной сети и меняем ее согласно цели

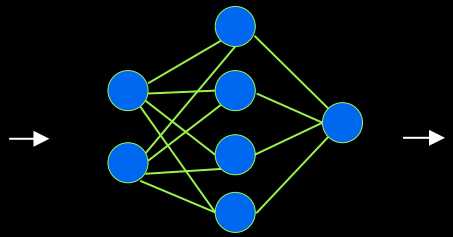
Получаем результаты на тестовых датасетах

Сравниваем наше решение со сторонним продуктом

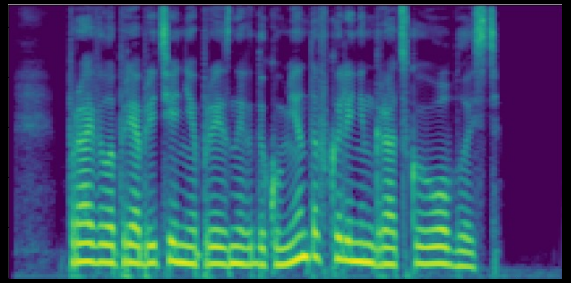
# Задача шумоподавления



сигнал с шумом



нейронная сеть



сигнал без шума

# Задача шумоподавления

- Нейронная сеть работает быстро,
  сохраняет качество речи

- Шумоподавление работает даже в тяжелых условиях,
  когда шума много

- Пользователи слышат только чистые голоса друг друга
  без искажений и прерывистости

# Как выбирали архитектуру сети

- Выдает высокое качество речи
- Модель потенциально портируется в веб и мобайл

# Как выбирали архитектуру сети

- Выдает высокое качество речи
- Модель потенциально портируется в веб и мобайл
- Работает в real-time
- Lookahead (контекст в будущее) < 32 мс

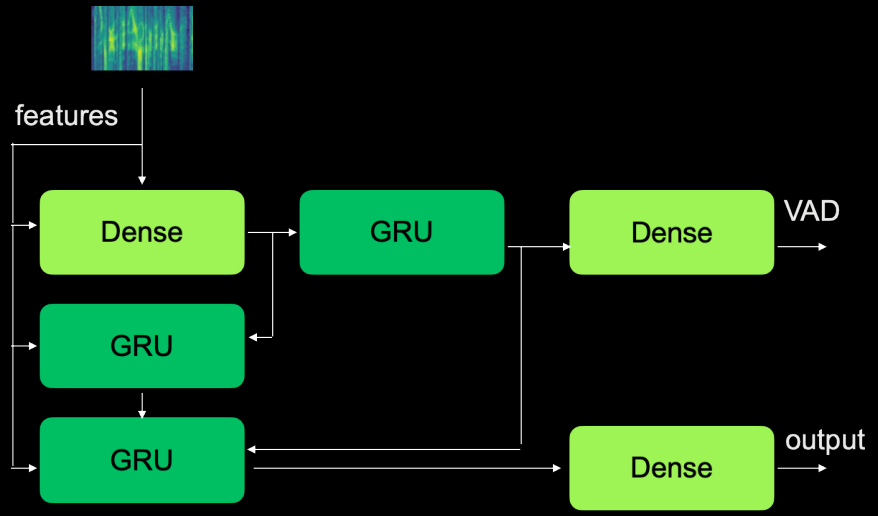
# Как выбирали архитектуру сети

- Выдает высокое качество речи
- Модель потенциально портируется в веб и мобайл
- Работает в real-time
- Lookahead (контекст в будущее) < 32 мс
- Real-Time Factor (скорость инференса) << 1.0
- Footprint (размер модели) =< 10 Мб

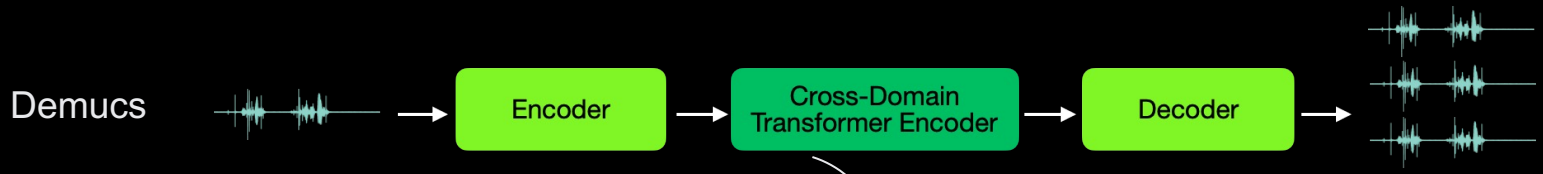
# Сравнение моделей

RNNoise

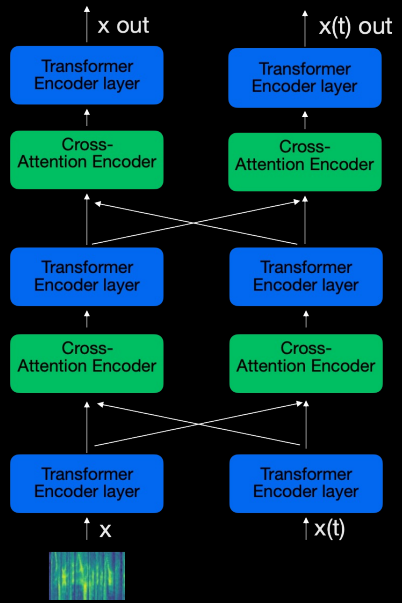
- Быстрая
- Работает с данными 48кГц
- Footprint < 1Мб
- Lookahead = 100мс
- RTF = 0.1
- Не справляется с шумами во время речи



# Сравнение моделей



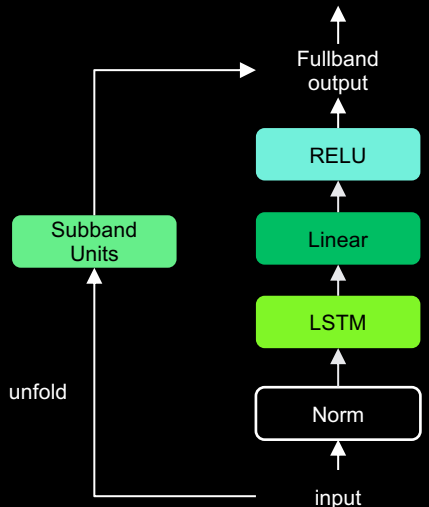

- Большой размер модели
- Footprint = 108Мб
- Агрессивно очищает речь, возникают прерывания
- Lookahead = 0мс
- RTF = 1.49
- Сложная переработка модели для real-time




https://arxiv.org/abs/2211.08553

# Сравнение моделей

FullSubNet

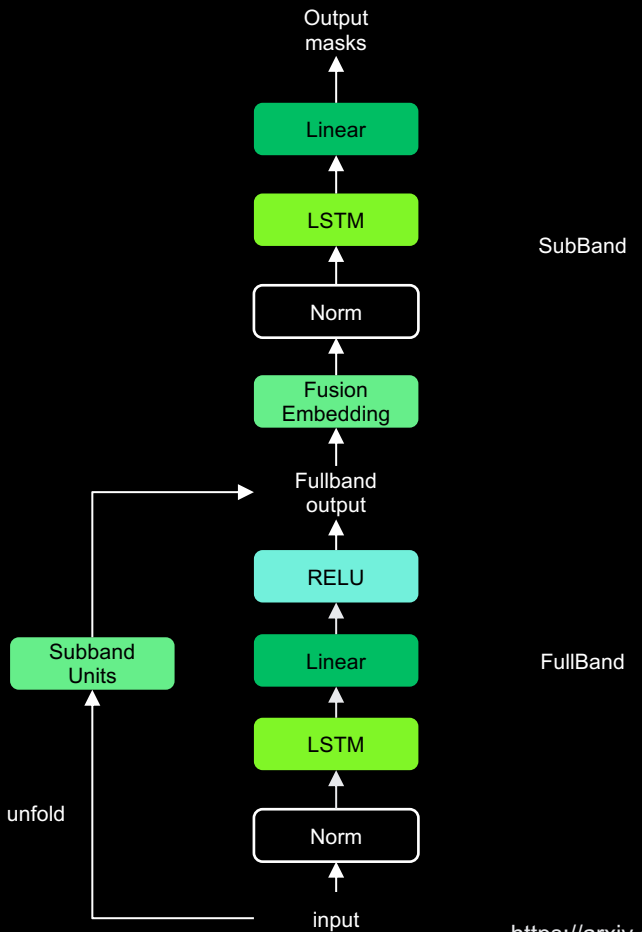



FullBand

https://arxiv.org/abs/2010.15508

# Сравнение моделей

FullSubNet

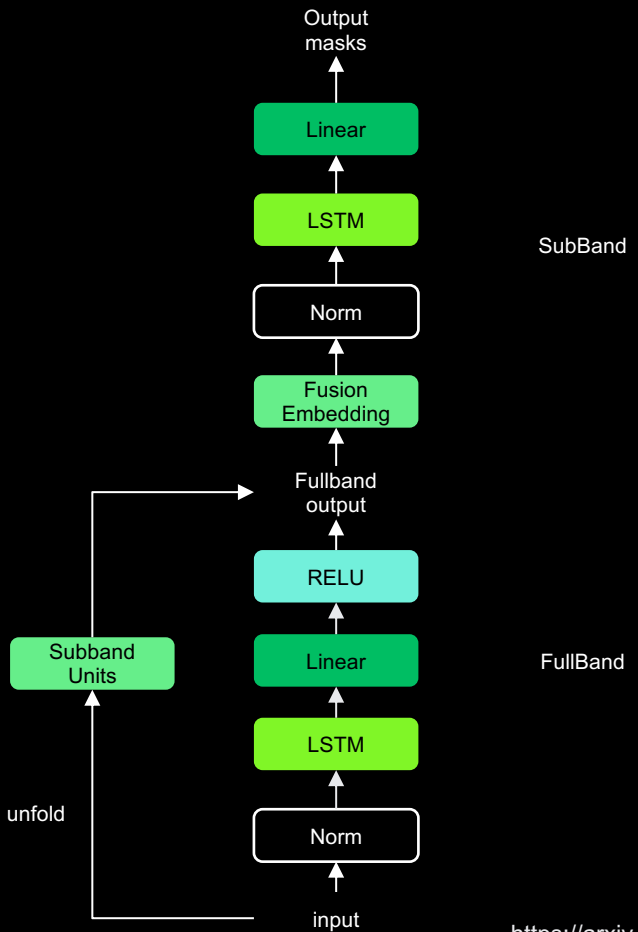



https://arxiv.org/abs/2010.15508

# Сравнение моделей

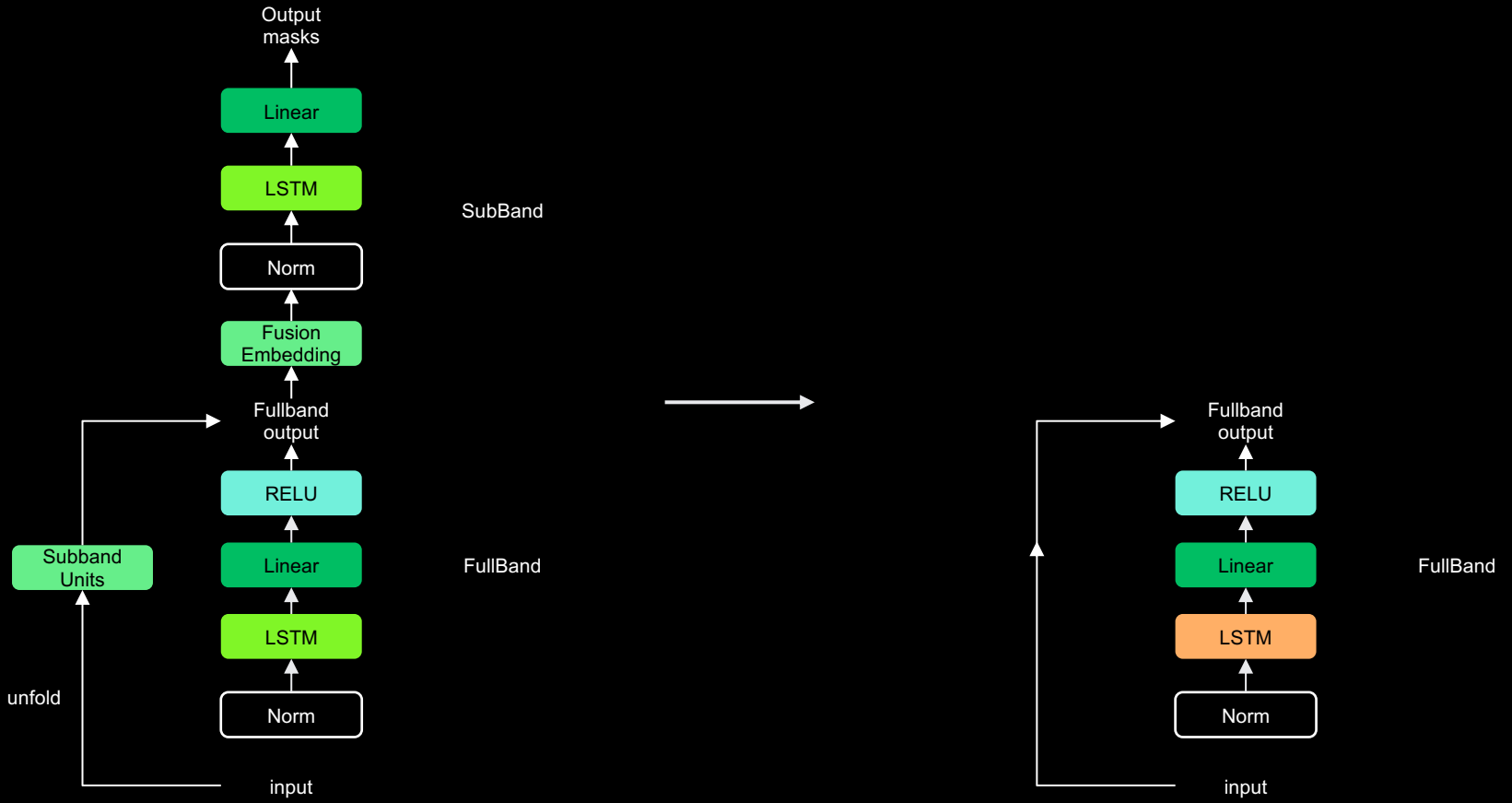
FullSubNet

- Рассчитана на работу с данными 16кГц
- RTF = 0.6
- Не вносит деформацию в голос
- Footprint модели 14Мб
- Lookahead модели 32мс

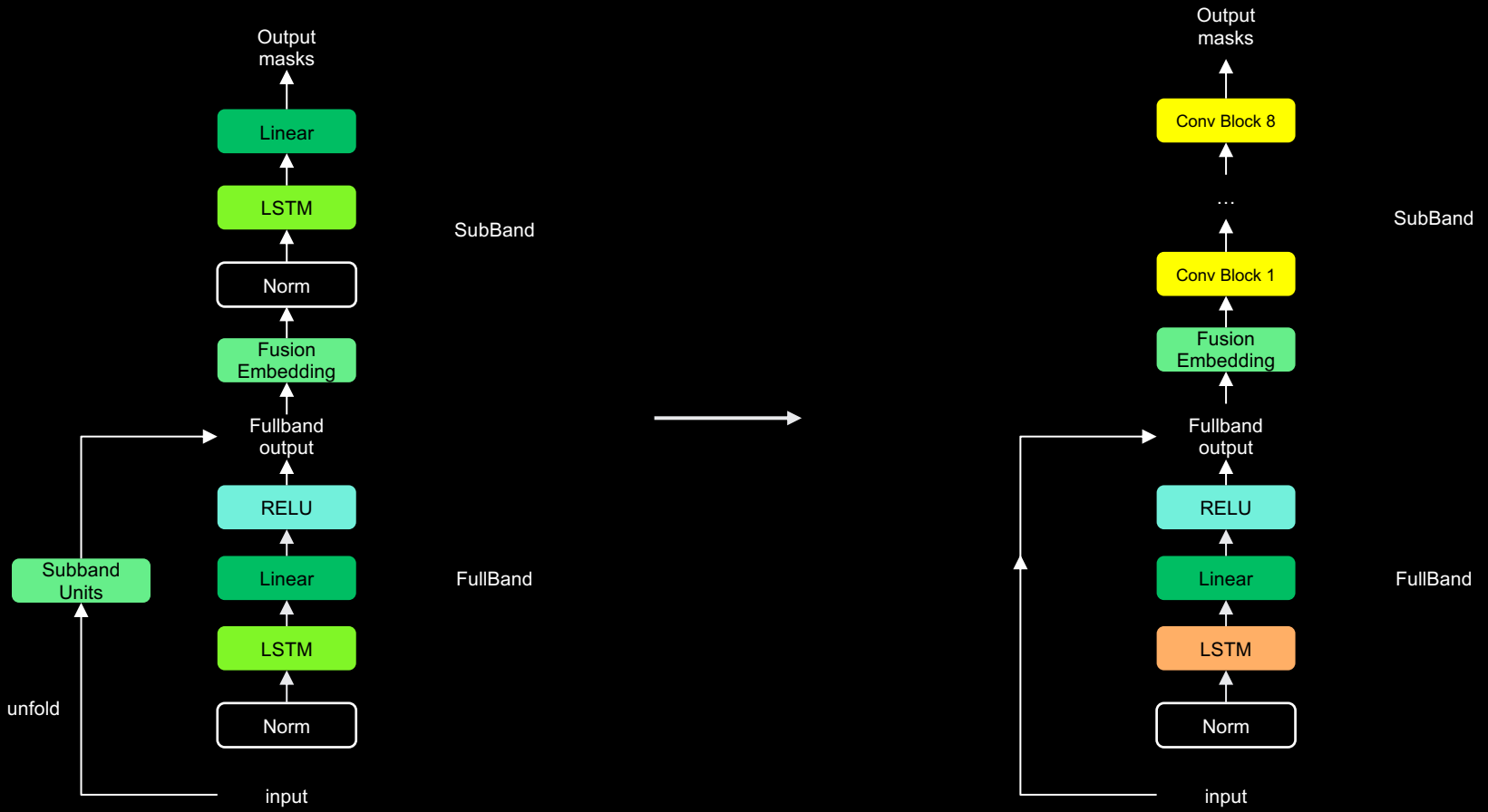



https://arxiv.org/abs/2010.15508

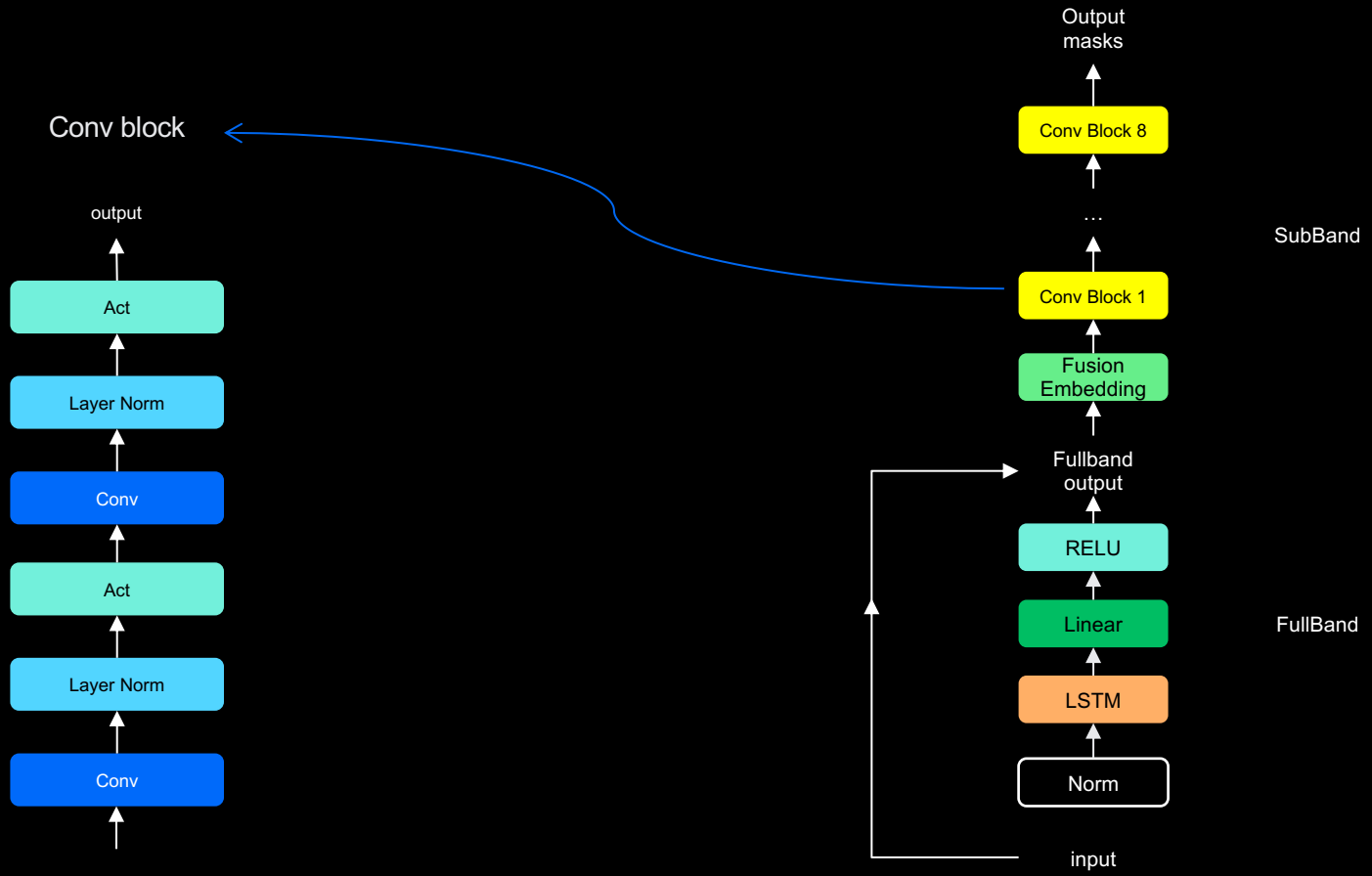
# Изменение архитектуры FullSubNet



# Изменение архитектуры FullSubNet



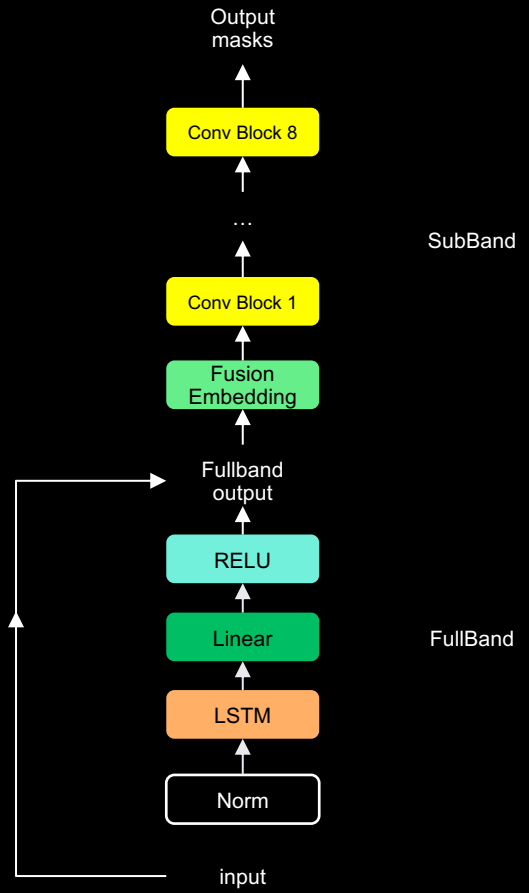
# Изменение архитектуры FullSubNet



# Характеристики модели для данных 24кГц

Важные замечания:

- Causal свертки с группами
- Уменьшение LSTM
- 384 frequency bin
- Входные данные – спектрограмма
- MSE + SI-SDR loss



# Характеристики модели для данных 24кГц

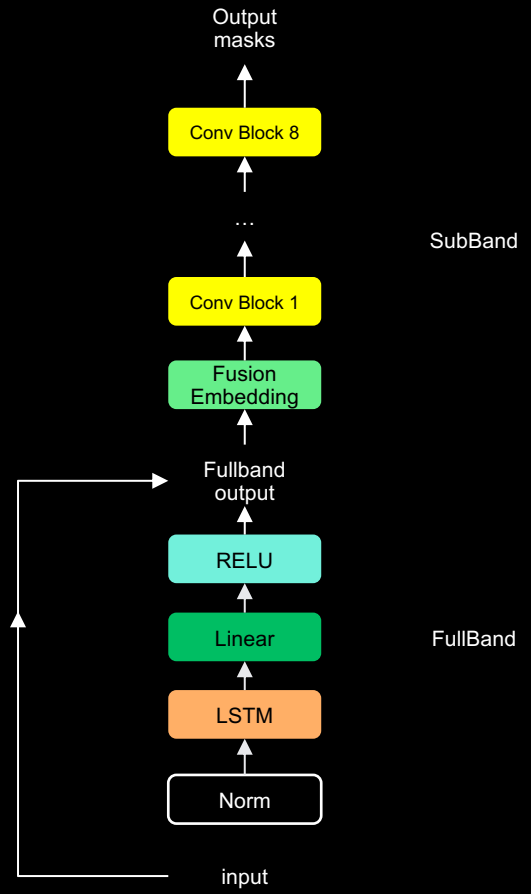
Важные замечания:

- Causal свертки с группами
- Уменьшение LSTM
- 384 frequency bin
- Входные данные – спектрограмма
- MSE + SI-SDR loss

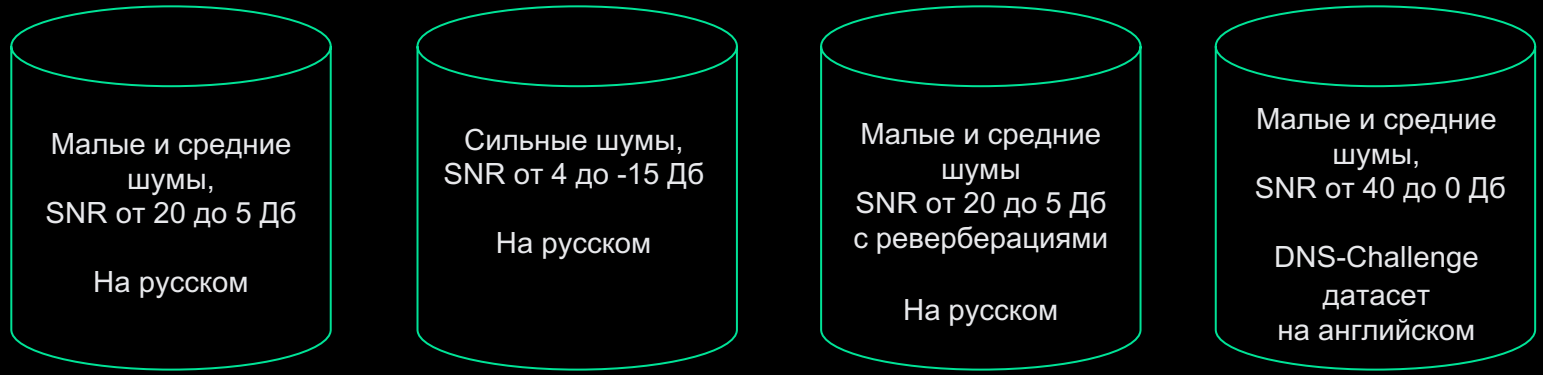
Итоги:

- Lookahead **= 0мс**
- Real Time Factor [0.08; 0.2]*
- Footprint [5.7; 20] Мб

*2,6 GHz 6-Core Intel Core i7 MacOS Ventura 13.2.1 and with tflite framework

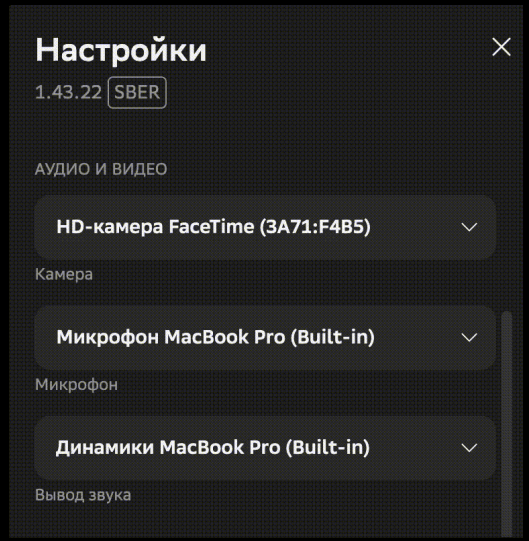


# Тестовые датасеты

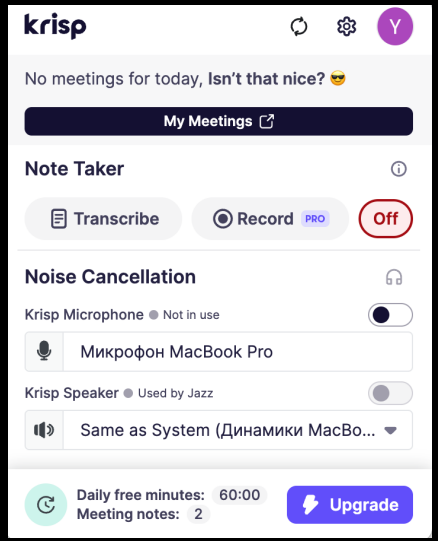


# Сравнение с Krisp

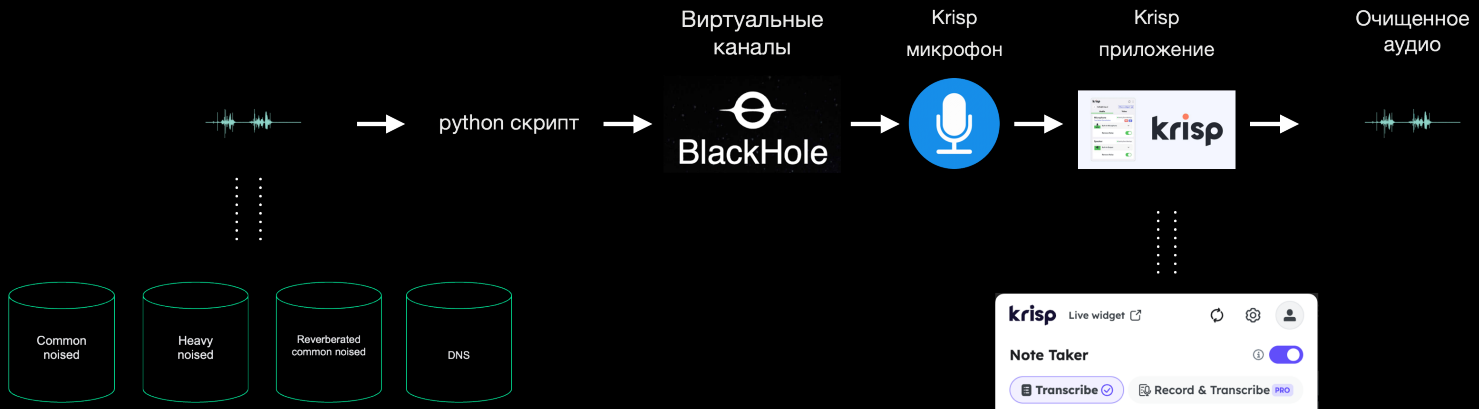
Интерфейс SaluteJazz




Интерфейс Krisp



# Сравнение с Krisp




```python
import sounddevice as sd
import soundfile as sf
import librosa
import numpy as np
import os

in_directory = 'path_to_noised_files'
out_directory = 'path_for_enhanced_audio'
for filename in os.listdir(in_directory):
    if filename.endswith(".wav"):
        data, fs = librosa.load(os.path.join(in_directory, filename), sr=48000, dtype='float32')
        sd.default.samplerate = fs
        sd.default.channels = 1
        sd.default.dtype = 'float32'
        myrecording = sd.playrec(data, device=(4,1), blocking=True) #4 - krisp mic, 1 - BlackHole
        sf.write(os.path.join(out_directory, filename), myrecording, fs)
```

# Сравнение с Krisp

| | Десктоп, Мб | Мобайл, Мб | Веб, Мб |
|---|---|---|---|
| Krisp 32кГц | 48 | 5.6 | 5.6 |
| SaluteSpeech 24кГц | 20 (fp16) | 10 (fp16) | 6.7 (int8) |

# Метрики

| Модель | Размер, Мб | STOI | WB_PESQ | SI-SDR | Инференс, мс |
|---|---|---|---|---|---|
| Rnnoise | 0.5 (fp32) | 0.720 | 2.130 | -24.200 | - |
| Бейзлайн | 40 (fp32)<br>20 (fp16) | 0.907 | 2.733 | 17.408 | 2.61 |
| Бейзлайн с нормированными данными и гибридной loss функцией | 40 (fp32)<br>20 (fp16) | **0.914** | **2.785** | **19.099** | 1.18 |
| Модель фильтербанки | 20 (fp32)<br>10 (fp16) | 0.909 | 2.750 | 17.484 | 2.12 |

STOI: разборчивость

Диапазон: от 0 до 1

WB_PESQ: качество речи, близость к оригиналу, отсутствие артефактов

Диапазон: от -0.5 до 4.5

SI-SDR: степень искажений по сравнению с эталонным сигналом

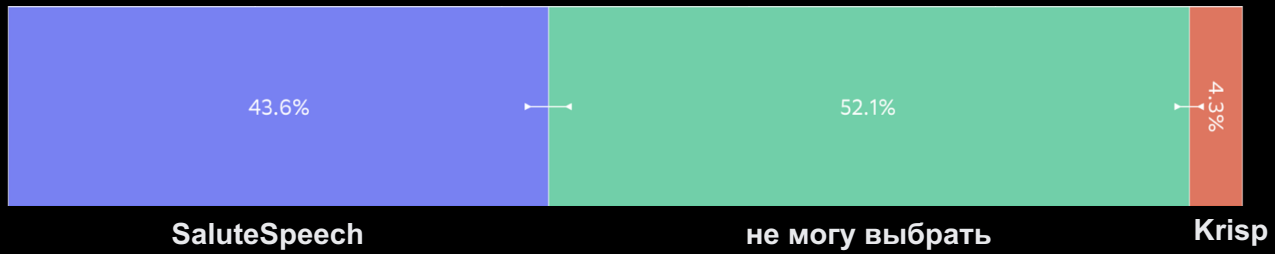
Диапазон: от отрицательных до положительных величин

# Субъективная оценка качества звука

- 2 датасета, очищенные SaluteSpeech и Krisp, размещены на платформе Яндекс Задания
- Респонденты попарно сравнивают аудиозаписи и выбирают,

  какая из аудио лучше по качеству и по наличию в ней шумов
- Результаты выводятся в диаграмме

Результаты SBS на датасете со средним уровнем шумов




SBER DEVICES

# Инференс на тестовых датасетах

| | Krisp | | | | | |
|---|---|---|---|---|---|---|
| | Сильные шумы | | | | | |
| SaluteSpeech | Small 5.6Мб | | | Large 40Мб | | |
| Web 6.7Мб | **39.7** | 27.1 | 33.1 | | | |
| Mobile 10Мб | **44.3** | 37.3 | 18.4 | **24.3** | 53.8 | 21.9 |
| Large 20Мб | | | | **35.1** | 51.0 | 13.9 |

**SaluteSpeech** **не могу выбрать** **Krisp**

SBER DEVICES

# Инференс на тестовых датасетах

| | Krisp | | | | | |
|---|---|---|---|---|---|---|
| | Средние шумы | | | | | |
| SaluteSpeech | Small 5.6Мб | | | Large 40Мб | | |
| Web 6.7Мб | **42.0** | 41.3 | 16.7 | | | |
| Mobile 10Мб | **47.2** | 44.7 | 8.0 | **38.0** | 53.6 | 8.4 |
| Large 20Мб | | | | **43.6** | 52.1 | 4.3 |

**SaluteSpeech** **не могу выбрать** **Krisp**

SBER DEVICES

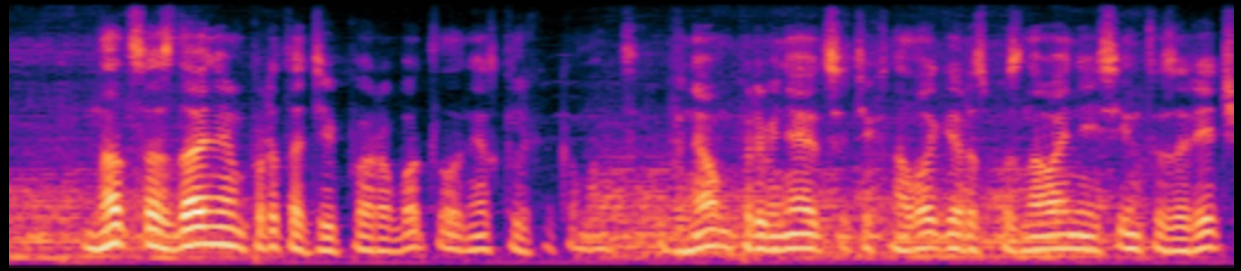
# Инференс на тестовых датасетах

| | Krisp | | | | | |
|---|---|---|---|---|---|---|
| | Средние шумы и реверберация | | | | | |
| SaluteSpeech | Small 5.6Мб | | | Large 40Мб | | |
| Web 6.7Мб | 25.3 | 35.4 | **39.3** | | | |
| Mobile 10Мб | 31.1 | 33.6 | **35.0** | **29.9** | 51.0 | 19.3 |
| Large 20Мб | | | | **38.3** | 24.6 | 27.1 |

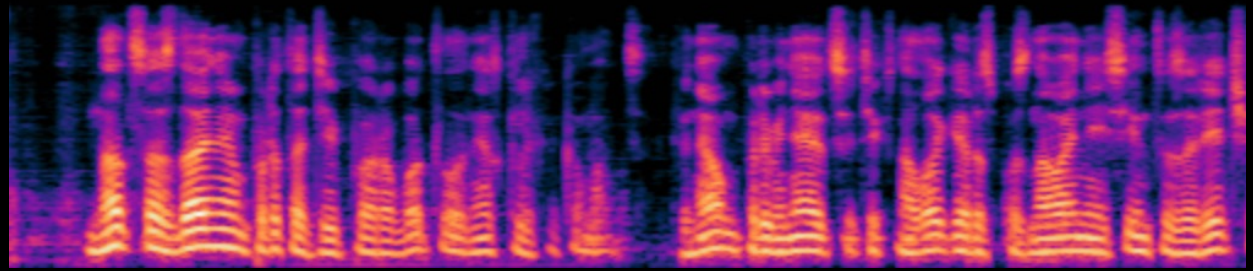
**SaluteSpeech** **не могу выбрать** **Krisp**

# Примеры аудиозаписей

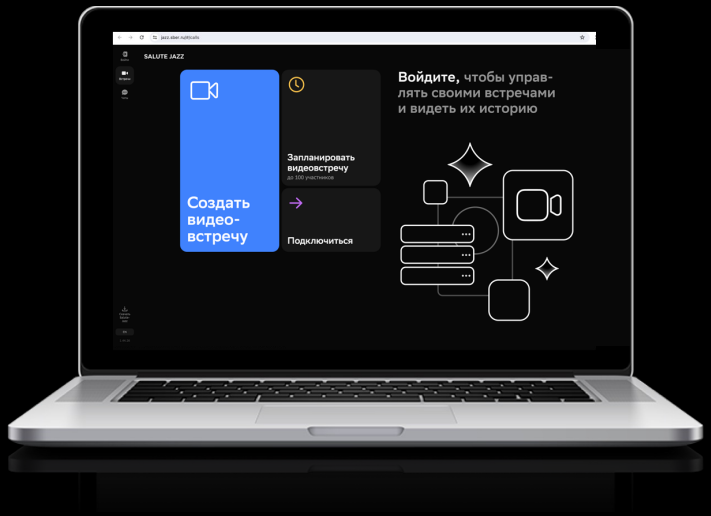
**Аудиозапись со средним шумом**



**Аудиозапись очищенная SaluteSpeech**



# Платформы




Уже доступно:
Windows, Linux, MacOS,
Android, Web

SBER DEVICES

yvlysova@sberbank.ru

t.me/ianalys