

## Архитектура модели

Используемая модель включает в себя последовательные этапы обработки аудиосигнала:

1. **Предобработка:** аудиосигнал преобразуется в мел-спектрограмму с использованием LogMelFilterBanks.
2. **Экстракция признаков:** далее следуют три последовательных блока:
  - 1 блок: свёртка (Conv1d) (размером 3) → ReLU → BatchNorm → MaxPooling.
  - 2 блок аналогичен
  - В третьем блоке используется AdaptiveAvgPool1d, что позволяет получить фиксированный размер на выходе.
3. **Эмбединги:** выход последнего сверточного блока преобразуется через Flatten, затем подается на линейный слой до 128 признаков, за которым снова следует BatchNorm и ReLU.
4. **Классификация:** финальный Linear слой с двумя выходами (softmax по умолчанию) классифицирует аудио по двум классам.
5. **Замеры проводились на GPU: NVIDIA RTX 3090**

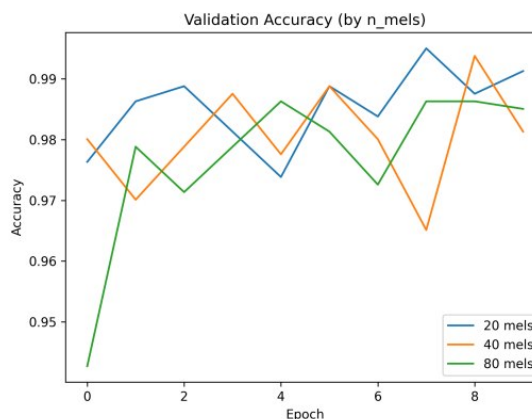
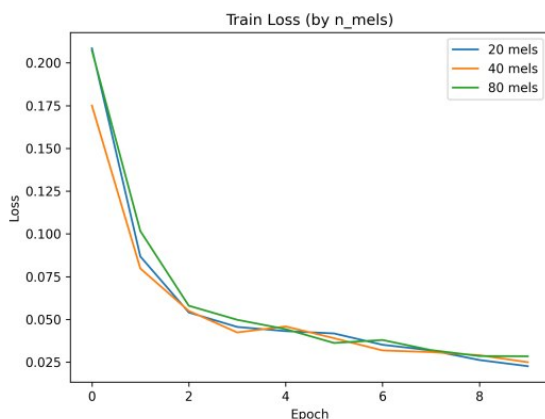
### Эксперимент 1: Подбор количества мел банков

Проверялись три значения числа мел-фильтров (n\_mels): 20, 40 и 80.

Все три конфигурации сходились с примерно одинаковой скоростью, однако точность отличалась:

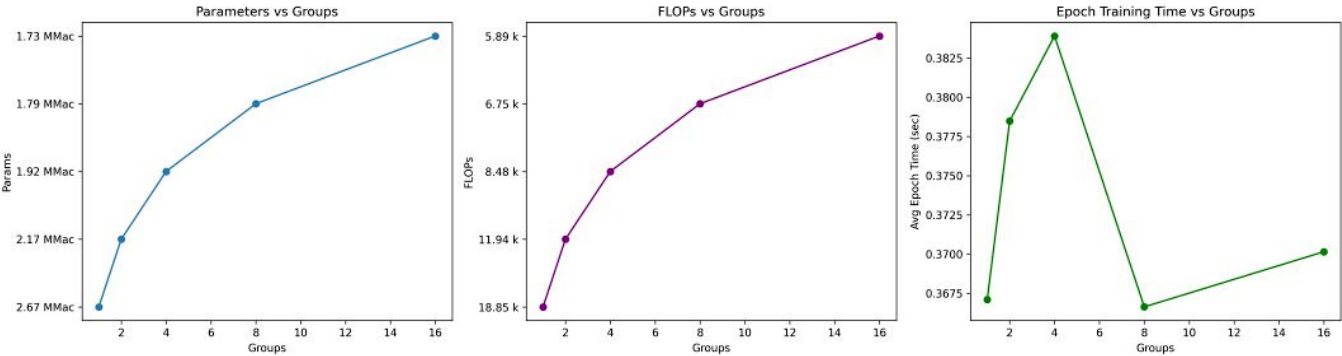
Параметр n_mels	Точность на тесте
20	0.9976
40	0.9830
80	0.9964

Таким образом, наилучший результат был достигнут при n\_mels = 20, несмотря на его меньшую размерность. Это может быть связано с лучшей обобщающей способностью модели на менее детализированном спектре.



## Эксперимент 2: Влияние количества групп в свёртках (grouped convolutions)

Был проведён тест с изменением параметра groups в слоях Conv1d: от 1 (без группировки) до 16 (максимальная разрядность).



### Общие наблюдения:

- С ростом количества групп количество параметров и операций (MACs) сильно уменьшается.
- Точность модели остаётся высокой даже при значительном уменьшении числа параметров.
- Скорость обучения (время на эпоху) почти не меняется — около 0.38 сек.

### Сравнительная таблица:

Groups	Params (тыс.)	MACs	Max Val Acc	Epoch Time
1	18.85k	1.04M	0.9913	0.38s
2	11.94k	0.54M	0.9875	0.39s
4	8.48k	0.28M	0.9851	0.38s
8	6.75k	0.16M	0.9714	0.38s
16	5.89k	0.096M	0.9614	0.38s

### Вывод:

Оптимум между точностью и производительностью наблюдается при 4–8 группах. Использование 16 групп ведёт к наименьшему числу параметров, при этом модель всё ещё сохраняет высокую точность, что делает её подходящей для встраиваемых решений.

### Итог:

В ходе экспериментов показано, что:

- Качество сопоставимо при примерно одинаковом количестве мел банков
- Групповые свёртки позволяют резко снизить нагрузку на вычисления, сохраняя высокую точность.