

Автоматическая настройка усиления микрофона

Студент:

Голованов Сергей

Руководители:

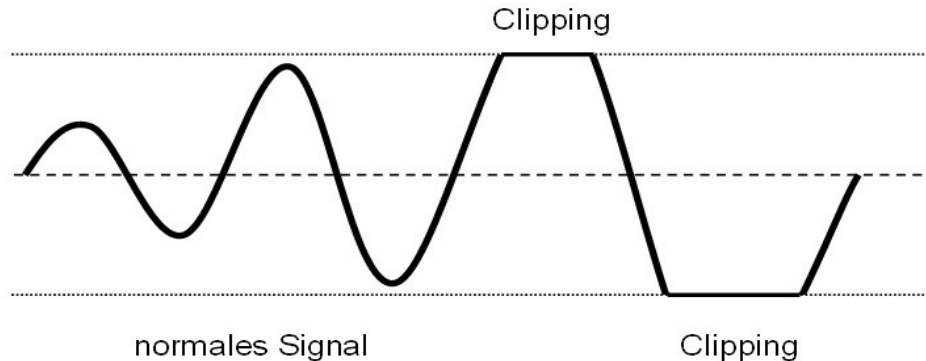
Служаев Евгений Дмитриевич

Шпильман Алексей Александрович

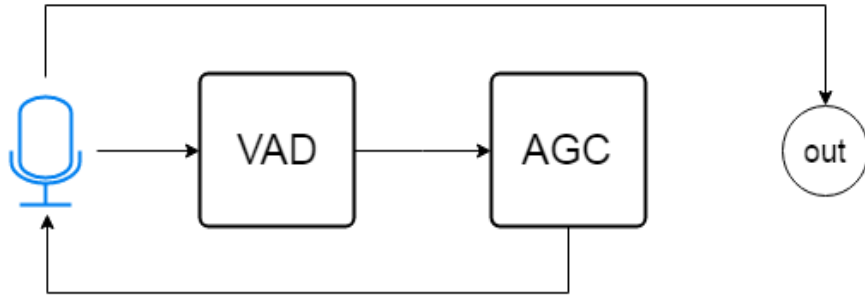
Мотивация

Качество работы большинства систем, обрабатывающих речь, во многом зависит от “громкости” входного сигнала.

- Высокая “громкость” может вызвать обрезку амплитуд (клиппинг)
- Низкая “громкость” затрудняет обнаружение речи и снижает ее разборчивость



Описание задачи



Automatic gain controller - поддержание “громкости” речи на заданном уровне

Voice activity detector - детекция речи и удаление шумов

Цель: повысить качество работы имеющегося модуля AGC

Задачи:

- Повысить характеристики VADa
- Повысить адаптацию к изменениям громкости речи и устойчивость к всплескам амплитуд

Новый VAD

Готовые реализации:

- Протестировано несколько существующих реализаций (WebRTC, Bob.spear, Baidu, VoiceBox, Bowon Lee's vad).
- Их основной недостаток - невозможность отсеять высоко-амплитудный шум.
- Наиболее оптимальный, WebRTC VAD, взят для использования.

[Репозиторий](#) - исходный код WebRTC VAD
[GMM](#) - описание алгоритма GMM

WebRTC VAD:

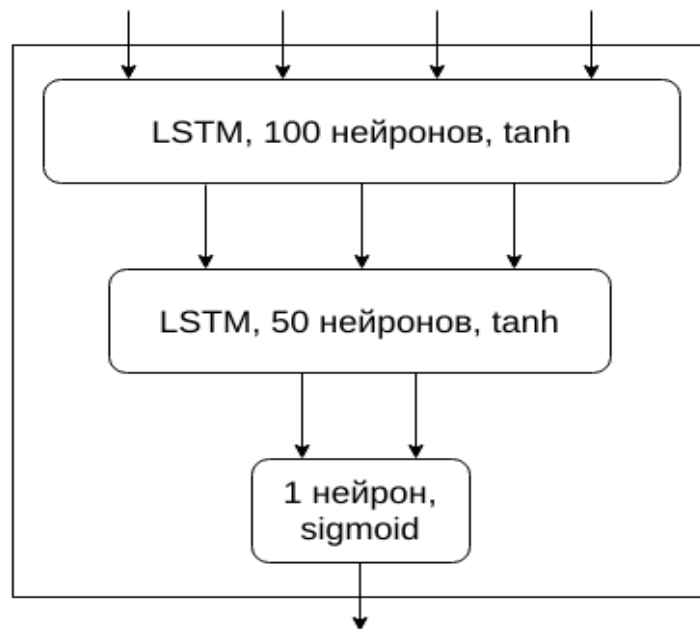
- Используются фреймы по 40 мс
- Для каждого фрейма вычисляется энергия в 6-ти частотных поддиапазонах и общая энергия
- В случае, если общая энергия выше заданного порога, применяется алгоритм Gaussian mixture models с энергией в поддиапазонах в качестве признаков
- Параметры GMM обновляются в процессе работы

Удаление высоко-амплитудных шумов

При классификации фрейма VADом как речь, он дополнительно проверяется рекуррентной нейронной сетью (RNN).

RNN:

- Входная матрица размерности 4x70 (учитывается 3 предыдущих фрейма и текущий)
- Loss - бинарная кросс энтропия
- Алгоритм оптимизации - adam



[LSTM](#) - описание LSTM слоя

[Adam](#) - описание алгоритма оптимизации adam

Удаление высоко-амплитудных шумов (признаки)

Для каждого фрейма вычисляется вектор признаков из 70 элементов:

- Gammatone frequency cepstral coefficients (50 элементов)
- Доля энергии в барк поддиапазонах спектра (13 элементов)
- Энтропия спектра
- Zero-crossing rate
- Доминирующая частота
- Максимум производной, до максимального значения сигнала
- Коэффициент плоскостности огибающей сигнала
- High frequency content спектра
- Pitch salience спектра

Удаление высоко-амплитудных шумов (обучение)

Dataset: речь взята из [VoxForge database](#), шумы - из [IEEE AASP Challenge](#), дополнительно шумы были наложены на речь.

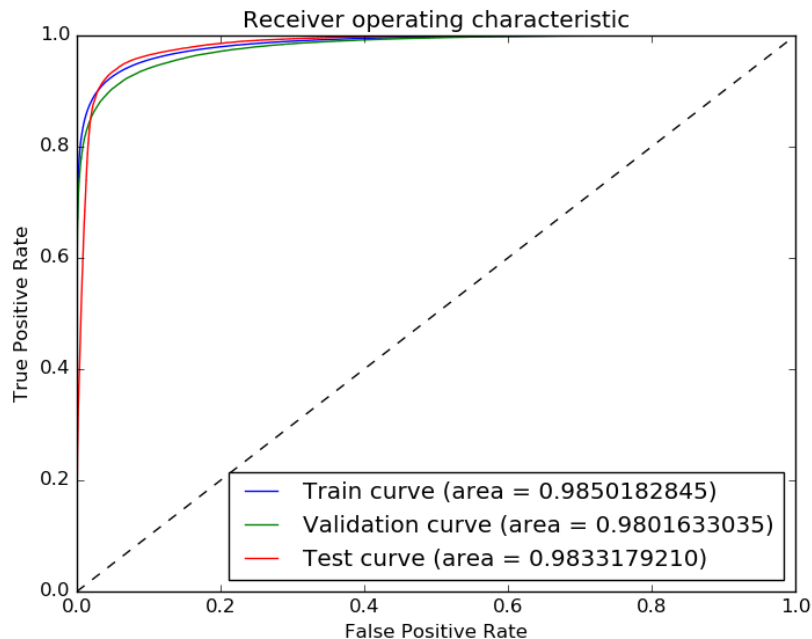
Train dataset:

- ~ 1000000 фреймов речи
- ~ 800000 фреймов шума

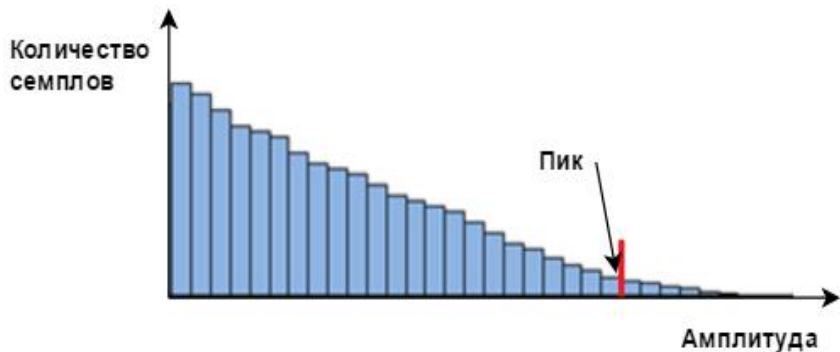
20 % - validation, 80% - train

Test dataset:

- ~ 75000 фреймов речи
- ~ 50000 фреймов шума



Алгоритм настройки усиления



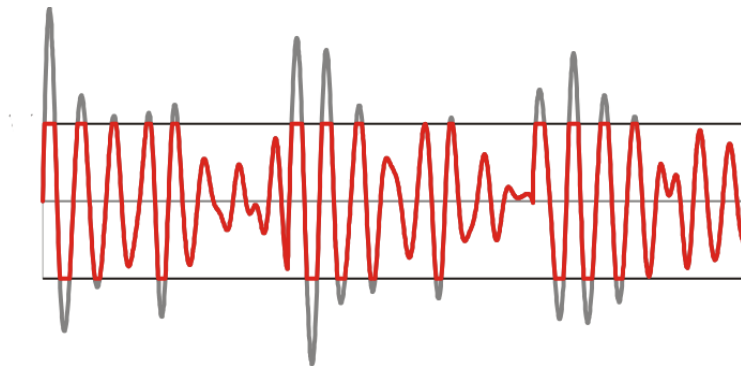
Амплитуды сигнала, нормированы относительно максимально возможного значения

При каждом обновлении уровня усиления:

- Находится пик амплитуд за последние 150 мс
- Амплитуда рассматривается, если количество семплов справа, включая текущую амплитуду, превышает порог
- Значение пика ограничено снизу
- Пик обновляется сглажено

Алгоритм настройки усиления

Проверяется наличие клиппинга



Если есть клиппинг:

- Незначительно увеличивается величина пика
- Уровень клиппинга обновляется на значение текущего уровня усиления
- Уровень клиппинга ограничивается снизу

Через некоторый промежуток времени, в случае отсутствия обрезки сигнала, уровень клиппинга медленно поднимается

Алгоритм настройки усиления

Вычисляется уровень усиления:

- Уровень усиления обновляется сглажено
- Уровень усиления не превышает уровень клиппинга
- Шаг изменения уровня усиления ограничен (разные шаги для повышения и понижения)
- Уровень усиления ограничивается максимальным и минимальным значениями

Тестирование

ideal - нормализованная речь продолжительностью 35 мин

simulation - ideal, домноженный на кусочно постоянную функцию (дополнительно могут накладываться шумы)

algo - алгоритм, нормализующий речь на заданный уровень

error_simulation = mean(|ideal - simulation|)

error_algo = mean(|ideal - algo(simulation)|)

advantage = 100 * (error_simulation - error_algo) / error_simulation

Метрика	Наличие шума	Старый подход	Подход из Webrtc	Новый подход без удаления шумов	Новый подход с удалением шумов
advantage, %	нет	81	79	85	84
	есть	54	72	76	82

Спасибо за внимание