

А.В. Панкова, А.М. Романенков

## РЕСТАВРАЦИЯ АУДИОПОТОКА С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ И АЛГОРИТМОВ РАСПОЗНАВАНИЯ РЕЧИ

**Аннотация.** Проведено исследование в области шумоподавления в речевом сигнале. Идентификация речи из зашумленного сигнала происходит с помощью скрытых марковских моделей изолированных слов. Предложен способ очистки речевого сигнала от посторонних шумов с применением фильтра Винера на основе эталонных слов как альтернатива методам спектрального вычитания и винеровской фильтрации с использованием образца шума. Полученные результаты отражены в программной реализации исследуемого способа фильтрации.

**Ключевые слова:** распознавание речи, скрытые марковские модели, реставрация речи, шумоподавление, фильтр Винера.

A.V. Pankova, A.M. Romanenkov

## AUDIO STREAM RESTORATION USING SPEECH RECOGNITION METHODS AND ALGORITHMS

**Abstract.** A study has been carried out in the field of noise suppression in a speech signal. Identification of speech from a noisy signal occurs with the help of hidden Markov models of isolated words. A method for cleaning a speech signal from extraneous noise using a Wiener filter based on reference words is proposed as an alternative to spectral subtraction and Wiener filtering methods using a noise sample. The results obtained are reflected in the software implementation of the studied filtration method.

**Keywords:** speech recognition, hidden Markov models, speech restoration, noise reduction, Wiener filter.

### *Введение*

Улучшение качества передаваемого аудиосигнала является главной задачей для цифровой фильтрации зашумленных сигналов, поэтому разработка методов очистки сигнала от шума является актуальным направлением исследований, например, при анализе записей с посторонними шумами и помехами в целях судебной экспертизы, восстановления архивных аудиозаписей. Кроме того, всё чаще приходится передавать информацию в звуковом формате, находясь при этом в условиях с большим количеством источников постороннего шума. В данной статье рассматривается один из методов шумоподавления, первостепенной задачей которого является реставрация речевого сигнала.

Результатом реставрации аудиосигнала должно стать улучшение качества речи, присутствующей в сигнале, – повышение разборчивости и удаление шумов, влияющих на восприятие речи. В данной работе рассматриваются оцифрованные речевые сигналы с частотой дискретизации 16 кГц, к которым искусственно добавлен стационарный шум.

Для распознавания речи используется метод скрытых марковских моделей (далее – СММ), на основе которого получено фонетическое представление сигнала. Процедура распознавания речи является главным этапом в процессе фильтрации записи.

**Панкова Арина Викторовна**

магистрант, Российский государственный университет имени А.Н. Косыгина (Технологии. Дизайн. Искусство), Москва. Сфера научных интересов: Алгоритмы и структуры данных. Электронный адрес: [ari.pankova@yandex.ru](mailto:ari.pankova@yandex.ru)

**Романенков Александр Михайлович**

кандидат технических наук, доцент, доцент кафедры математики, Московский авиационный институт (НИИУ), Москва; старший научный сотрудник Федерального исследовательского центра «Информатика и управление», Российская академия наук, Москва. Сфера научных интересов: математическое моделирование; численные методы; методы и средства обработки и анализа данных; программирование и разработка программного обеспечения. Автор 70 опубликованных научных работ. ORCID: 0000-0002-0700-8465, Scopus Author ID: 57196480014, ResearcherID: AАН-9530-2020  
Электронный адрес: [romanaleks@gmail.com](mailto:romanaleks@gmail.com)

Результат распознавания речи (преобразование устной речи в текст) используется в качестве оценки работы процедуры с использованием метода СММ. Для устранения шумов применяется фильтр Винера. Сочетание фильтра Винера и СММ для улучшения качества речевого сигнала уже применялось в исследованиях (см. например, [1]), однако базировалось на разработке СММ-шума.

*Распознавание речи*

Процедура распознавания речи включает четыре предварительных этапа [2]: обнаружение речевой активности (Voice Activity Detection, VAD), сканирование сигнала перекрывающимися окнами и выделение фреймов, вычисление речевых признаков каждого фрейма (в данной работе признаками речевого сигнала являются 13 мел-частотных кепстральных коэффициентов), классификация векторов признаков по методу  $k$ -средних.

В данной работе используются изолированные дикторозависимые слова для иллюстрации механизма шумоподавления с использованием СММ. Каждому слову соответствует одна модель. СММ выбрана лево-правая с количеством состояний, зависящим от общего числа фонем каждого слова. Таким образом, количество состояний модели, например, для слова «ноль», равно трем. Выбранный способ определения состояний не является универсальным и приводит к ошибкам распознавания в условиях сильной вариативности произнесения слов и для больших словарей. Лучших результатов распознавания можно добиться при фиксированном количестве состояний для всех моделей. Однако задача данной работы заключается не в построении точной системы распознавания, а в проведении исследования в области реставрации сигнала на основе методов распознавания речи.

*Фильтрация*

Качество шумоподавления сигнала определяется двумя критериями: отсутствием помех и разборчивостью речи в отреставрированном звуковом сигнале. Одним из самых простых методов удаления шумов в сигнале является спектральное вычитание [3]. На основе данного метода получается спектр чистой речи путем вычитания из зашумленного сигнала спектра шума. Однако полученная оценка чистой речи не является оптимальной и может содержать музыкальный шум. Оптимальная оценка чистого речевого сиг-

Реставрация аудиопотока с использованием методов и алгоритмов распознавания речи

нала получается путем вычисления минимальной среднеквадратичной ошибки спектра (Minimum Mean Square Error, MMSE) [4–6] полезной составляющей сигнала и спектра очищенной речи. Оптимальный фильтр Винера обеспечивает минимальную среднеквадратическую разность данных спектров и позволяет получить наиболее близкий спектр чистой речи к полезному сигналу.

*Математическое описание фильтра Винера*

Зашумленный сигнал  $x(n)$  можно представить во временной области в виде суммы полезной  $y(n)$  составляющей сигнала и стационарного шума  $d(n)$  :

$$x(n) = y(n) + d(n). \tag{1}$$

Для перевода сигнала в частотную область вычисляется прямое дискретное преобразование Фурье (далее – ДПФ), и зашумленный сигнал представляется в виде спектра  $X(\omega)$  :

$$X(\omega) = Y(\omega) + D(\omega), \tag{2}$$

где  $Y(\omega)$  – спектр полезной составляющей сигнала;  $D(\omega)$  – спектр шума.

Далее необходимо избавиться от комплексной составляющей сигнала. Для этого необходимо вычислить модуль комплексного числа обеих частей выражения (2). Затем находится математическое ожидание  $M$  кратковременной спектральной мощности сигнала:

$$M(|X(\omega)|^2) = M(|Y(\omega)|^2) + M(|D(\omega)|^2) + 2M(|Y(\omega)D(\omega)|). \tag{3}$$

Предполагается, что шум и полезный сигнал не коррелируют, поэтому значение выражения  $2M(|Y(\omega)D(\omega)|)$  можно принять равным нулю. Математическое ожидание кратковременной мощности сигнала является его спектральной мощностью  $P$ . Формулу (3) можно переписать в виде

$$P_x(\omega) = P_y(\omega) + P_d(\omega). \tag{4}$$

Фильтр Винера  $W$  можно рассчитать, используя отношения спектральной мощности полезного сигнала  $P_y(\omega)$  к спектральной мощности шума  $P_d(\omega)$ , следующим образом [4; 7; 8]:

$$SNR(\omega) = \frac{P_y(\omega)}{P_d(\omega)}; W(\omega) = \frac{SNR(\omega)}{1 + SNR(\omega)}. \tag{5}$$

Для оценки чистого сигнала необходимо применить фильтр из выражения (5) к зашумленному спектру сигнала:

$$\hat{P}_y(\omega) = W(\omega)P_x(\omega)phase(X(\omega)), \tag{6}$$

где  $phase[X(\omega)]$  – фаза зашумленного сигнала.

Фаза очищенного сигнала остается неизменной, поскольку предполагается, что фазовые искажения на слух не воспринимаются. После вычисления обратного дискретного преобразования Фурье (далее – ОДПФ) получаем сигнал  $\hat{y}(n)$ , разность которого с полезным сигналом  $y(n)$  минимальна.

Чтобы реализовать фильтр Винера, необходимо знать либо спектр шума, либо спектр полезного сигнала (4). Образ шума можно получить, вычислив усредненный спектр сигнала, в котором нет полезного сигнала, например, если взять несколько участков с шумом перед непосредственно произнесением речи в зашумленном сигнале. Однако при достаточно низком соотношении SNR в результате винеровской фильтрации возникает музыкальный шум, который негативно сказывается на восприятии звукового сигнала [4]. Поэтому в данной статье предлагается вычислить спектр эталонного слова  $P_s$  с использованием усреднения участков фонем на основе изменения спектральных характеристик или алгоритма Витерби. А затем усредненный спектр сравнить со спектром зашумленного сигнала с помощью формулы

$$\hat{P}_d(\omega) = P_x(\omega) - P_s(\omega), \text{ если } \begin{cases} P_x(\omega) > P_s(\omega) \\ P_x(\omega) - P_s(\omega) < \alpha, \end{cases} \quad (7)$$

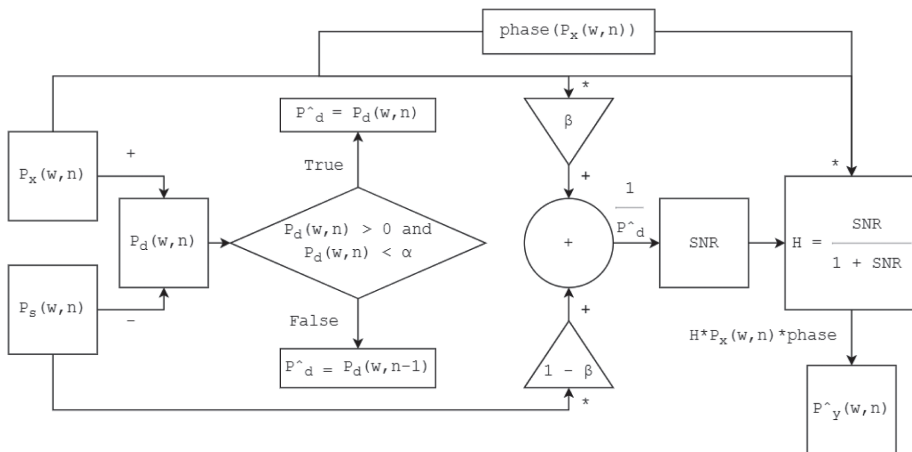
где  $\hat{P}_d(\omega)$  – некоторая оценка шумового сигнала.

Разница  $P_x(\omega) - P_s(\omega)$  в формуле (7) дает сигнал, в котором присутствует шум и остаточный полезный сигнал от  $P_x(\omega)$ . Однако предполагается, что разница между полезной составляющей зашумленного сигнала и усредненного эталонного слова минимальна, поэтому в результате получается некоторый образ шума. Параметр  $\alpha$  выступает в качестве порогового значения средней мощности шума, который можно вычислить на участках сигнала, где отсутствует речь. SNR для фильтра Винера вычисляется следующим образом:

$$SNR(\omega) = \beta \frac{P_s(\omega)}{\hat{P}_d(\omega)} + (1 - \beta) \frac{P_x(\omega)}{\hat{P}_d(\omega)}, \quad (8)$$

где  $\beta$  – параметр сглаживания, который позволяет найти некоторый баланс между отсутствием шума в сигнале и разборчивостью речи. В данной работе параметр принят равным 0,98 [4].

Схема алгоритма вычисления фильтра Винера в частотной области представлена на Рисунке 1, где  $n$  – номер фрейма.



**Рисунок 1.** Блок-схема алгоритма фильтра Винера  
 Источник: Здесь и далее схемы и рисунки составлены авторами.

*Вычисление спектра эталонного слова*

Речевые признаки, вычисленные с помощью мел-частотных кепстральных коэффициентов, можно использовать также для определения участков фонем слова. Для этого определяется скорость изменения значений вектора из 13 коэффициентов [9]:

$$Q(n) = \frac{1}{13} \sum_{i=1}^{13} \frac{\sum_{j=-r}^r P_i[n+j] \cdot j}{\sum_{j=-r}^r j^2}, \tag{9}$$

где  $n$  – номер текущего фрейма;  $j$  – номера фреймов вокруг текущего в диапазоне  $[-r, r]$ ;  $P_i$  –  $i$ -коэффициент вектора  $P$ , в квадратных скобках которого указан номер фрейма.

Затем из данных  $n$  значений скорости изменения находятся локальные максимумы, которые задают границы фонем. Уменьшение диапазона  $r$  приведет к разделению некоторых фонем на несколько участков, так как данный метод более чувствителен к переходам речевых параметров внутри фонемы. Это свойство можно использовать для более точного усреднения спектра эталонного слова в процессе фильтрации.

Вычислить спектр эталонного слова можно также с помощью алгоритма Витерби [9]. Данный алгоритм позволяет восстановить состояние СММ, которое с наибольшей вероятностью порождает вектор признаков фонем. Спектру распознанного слова ставятся в соответствие усредненные спектры по фреймам эталонного слова, в которых состояния модели совпадают. Данный метод жестко привязан к количеству состояний СММ, поэтому результат применения алгоритма Витерби будет зависеть от выбора акустической модели слова.

*Результаты применения фильтра Винера*

Вычисление спектра эталонного слова проводилось на основе изменения спектральных характеристик сигнала, так как данный способ дает лучшие результаты при реставрации речи в сравнении с алгоритмом Витерби для акустических моделей, используемых в данной работе. В качестве оценки восстановленной речи из зашумленного сигнала было выбрано отношение сигнал/шум, которое вычисляется с помощью среднеквадратической ошибки (Mean Square Error, MSE):

$$SNR_n = 10 \log_{10} \frac{MSE_{x(n)}}{MSE_{x(n)-\hat{y}(n)}}, \tag{10}$$

где выражение  $x(n) - \hat{y}(n)$  является остаточным шумом;  $MSE_{x(n)}$  и  $MSE_{x(n)-\hat{y}(n)}$  – среднеквадратические ошибки зашумленного сигнала и остаточного шума соответственно.

Исследование проводилось на десяти тестовых записях двух дикторов для трех уровней зашумленности гауссовым белым шумом (15 SNR, 10 SNR, 5 SNR). Сравнивались три метода шумоподавления:  $M_1$  – метод спектрального вычитания [3],  $M_2$  – винеровская фильтрация с образцом шума [4] и  $M_3$  – винеровская фильтрация с использованием эталонных слов при трех значениях параметра  $r$ , определяющих границы фонем.

В таблице представлены усредненные оценки результатов эксперимента.

Оценка SNRn результатов трех методов фильтрации при различных SNR, дБ\*

SNR	$M_1$	$M_2$	SNR <sub>n</sub>		
			$M_3$		
			$r=7$	$r=5$	$r=3$
15	15,9198	17,2454	20,3213	20,0912	18,0354
10	11,4294	13,7250	16,9477	16,5606	14,4110
5	7,0635	9,9801	14,3082	13,8506	10,7291

\*Источник: составлено авторами.

Как следует из таблицы, при использовании метода  $M_3$ , предложенного в данной работе, качество записи заметно улучшается после фильтрации для всех уровней зашумления по сравнению с методами  $M_1$  и  $M_2$ . Однако наилучшие результаты получаются при параметре  $r = 7$ . Дальнейшее уменьшение диапазона  $r$  приводит к ухудшению качества восстановленной записи.

На Рисунке 2, а–в, представлены спектрограммы для исходной, зашумленной и восстановленной записей при различных типах зашумления.

#### Программная реализация фильтра Винера

Для обработки речевого сигнала в данной работе реализована программа в виде веб-приложения. Ядро приложения написано на языке программирования Python 3.7 с использованием библиотек NumPy и Scikit-learn. Для пользовательского интерфейса приложения был выбран язык JavaScript с использованием библиотеки Dugraph. Подобная архитектура приложения обеспечивает быструю реализацию пользовательской части и платформонезависимость. Ядро на языке Python может быть в дальнейшем адаптировано под другие задачи или конвертировано в подключаемую библиотеку для встраивания в другие прикладные решения. Для механизма взаимодействия между пользовательским интерфейсом и ядром была выбрана реализация в виде RESTfulAPI (Representational State Transfer Application Programming Interface), являющаяся стандартом де-факто в области веб-разработок. В качестве основы для построения API выбран фреймворк FastAPI для языка Python.

Основной интерфейс программы состоит из трех html-страниц. На главной странице (Рисунок 3) изображен график аудиосигнала.

С левой стороны располагается область навигации, которая содержит навигатор по аудиофайлам. Раздел Filtering в области навигации открывает страницу для реставрации сигнала (Рисунок 4).

Раздел с отчетами содержит историю изменений, связанных с процедурой распознавания речи. В область навигации также включены этапы обработки речевого сигнала. Опционально можно выставить SNR и добавить к текущему аудиосигналу белый гауссовский шум. Основным функционалом главной страницы является реализация механизма определения речевой активности (Рисунок 5).

Над графиком аудиосигнала расположена функциональная панель, на которой указан путь к текущему файлу и расположен простой аудиоплеер. Также на функциональной панели расположены две кнопки для выделения участков с речью и шумом. Границы речи/шума сбрасываются при нажатии кнопки Reset.



Реставрация аудиопотока с использованием методов и алгоритмов распознавания речи

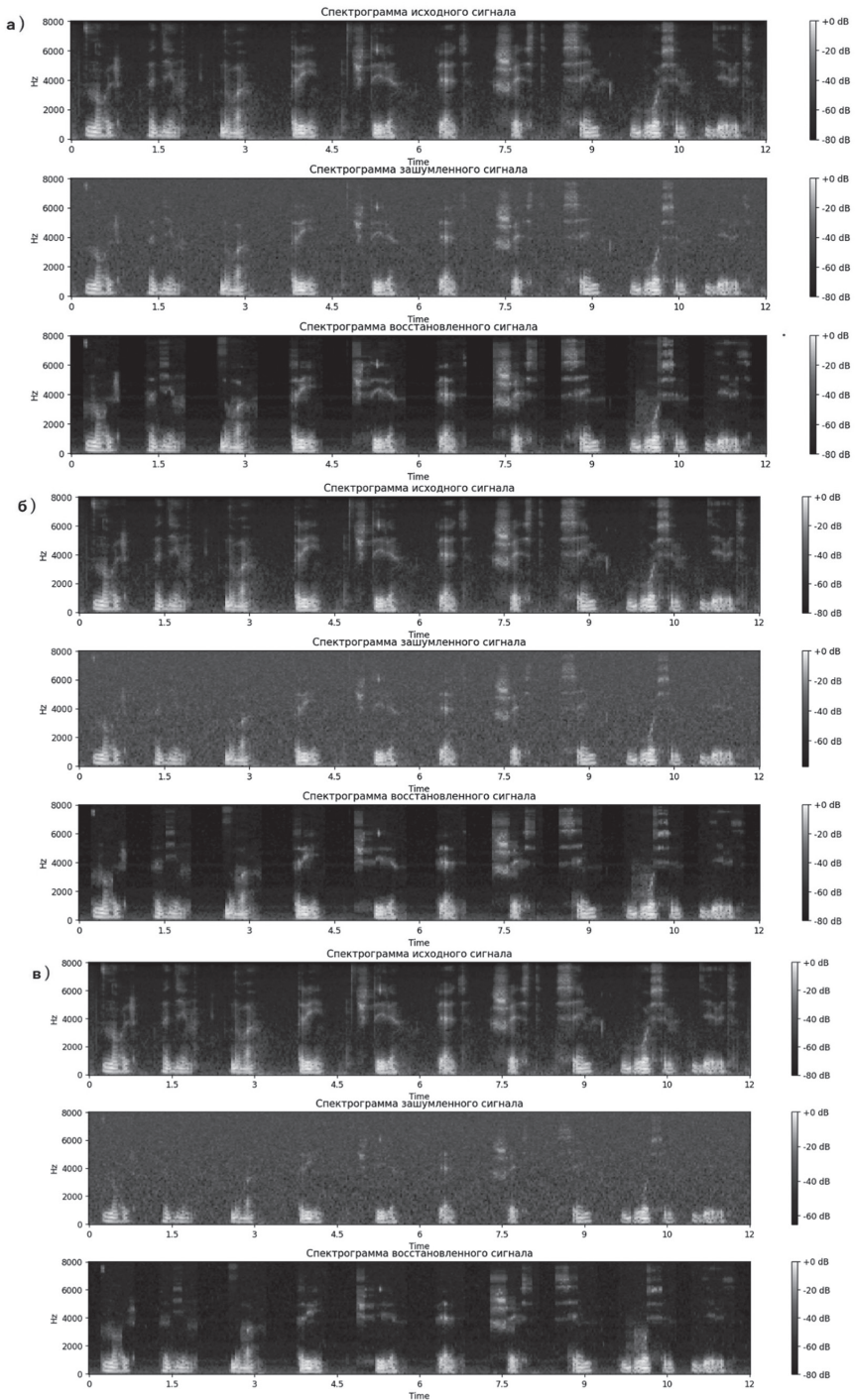


Рисунок 2. Сравнение спектрограмм исходного и восстановленного сигнала при уровне шума 15 SNR (а), 10 SNR (б) и 5 SNR (в)

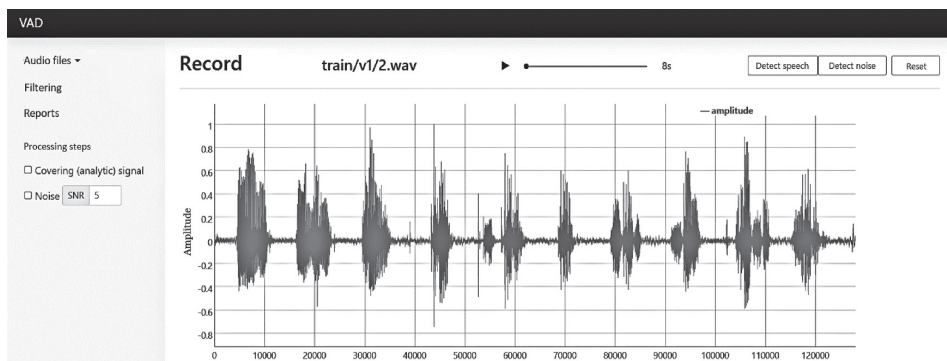


Рисунок 3. Главная страница

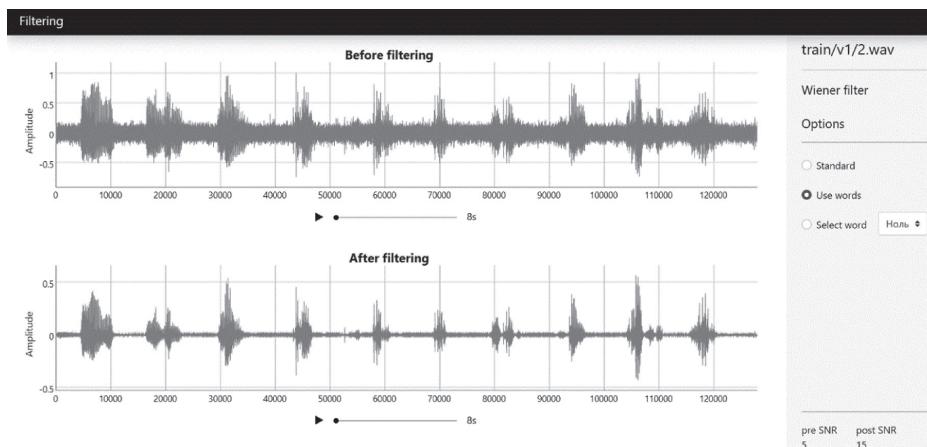


Рисунок 4. Страница для шумоподавления сигнала

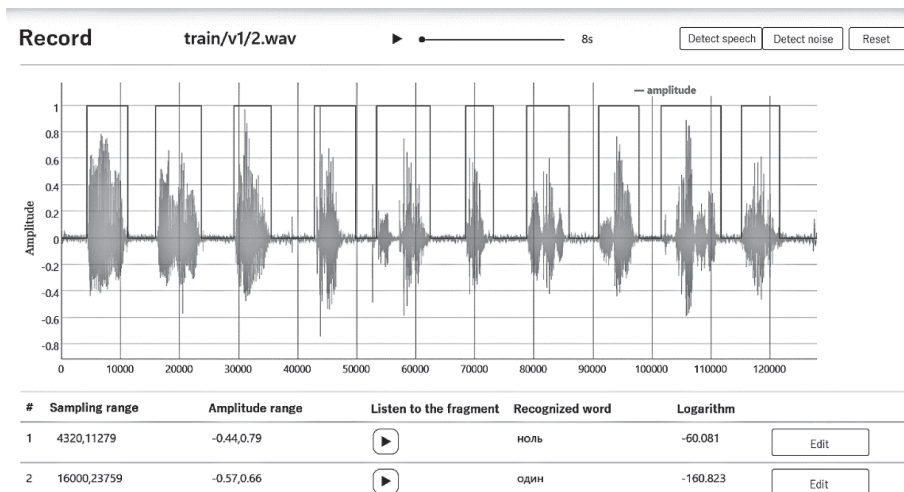
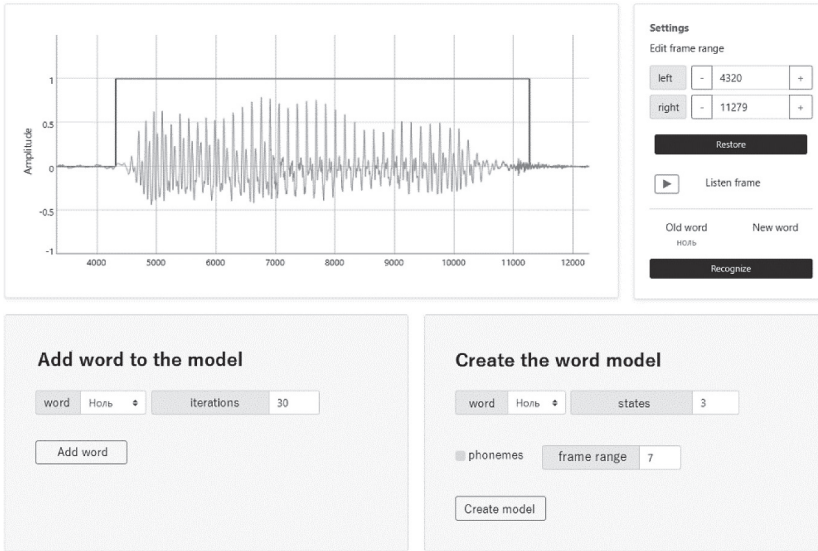


Рисунок 5. Определение речевой активности



## Реставрация аудиопотока с использованием методов и алгоритмов распознавания речи

При нажатии кнопки Detect Speech внизу графика появляется таблица с распознанными словами на участках выделенной речи (Рисунок 6).



**Рисунок 6.** Страница для редактирования слова

Таблица разделена на пять столбцов с заголовками: Sampling range, Amplitude range, Listen to the fragment, Recognized word и Logarithm. У каждого выделенного участка с речью есть промежутки отсчетов и амплитуда, модель слова и ее логарифм правдоподобия – максимальная вероятность порождения этого участка речи данной моделью. Каждый участок можно прослушать и отредактировать.

При нажатии на кнопку Edit какой-либо строки таблицы открывается страница для редактирования промежутка аудиосигнала (Рисунок 6). На данной странице отображается участок с речью, диапазон которого можно менять либо в области настроек, либо с помощью сочетания нажатия левой кнопки мыши и клавиши (Ctrl сдвигает левую, Alt – правую границу диапазона на позицию мыши). Выбранный диапазон можно сбросить при нажатии кнопки Restore. Если алгоритм VAD неточно определил участок речи, слово может быть распознано неверно. Поэтому для диапазона, установленного вручную, можно еще раз распознать речь с помощью кнопки Recognize, и в поле New word отобразится новое слово. Также на данной странице есть возможность добавить слово к существующей СММ. Это значит, что набор признаков, соответствующий этому участку слова, будет добавлен в последовательность наблюдений, а затем будет выполнен пересчет матриц вероятностей состояний и наблюдений выбранной модели с помощью алгоритма Баума – Велша [10]. Этот механизм добавления слова удобен для сбора обучающих данных. Кроме того, можно на основе выделенного слова создать новую СММ. Для этого необходимо выставить количество скрытых состояний модели или довериться программе, которая автоматически определяет количество состояний с помощью метода определения границ фонем по скорости изменения спектральных характеристик. Чтобы отобразить участки фонем на графике, необходимо поставить галочку напротив phonemes. По умолчанию диапазон фреймов установлен равным 7.

Таким образом, создание новых СММ в данной программной реализации обеспечивает простой способ расширения словаря.

Таким образом, данная реализация программного решения в виде веб-приложения создает удобную для исследования среду обработки и визуализации звуковых файлов. Построение графиков речевых сигналов и их реставрация являются основным функционалом приложения. В качестве опциональных возможностей доступны функции добавления шумов, обнаружения, выделения и редактирования промежутков речи.

#### *Заключение*

В ходе исследования был проведен сравнительный анализ трех методов шумоподавления на основе спектрального вычитания, фильтра Винера с использованием образца шума и разработанного в данной работе алгоритма с усредненным спектром эталонного слова. Результаты эксперимента показали, что фильтрация Винера с усредненным спектром дает лучшие показатели по сравнению с двумя другими методами, обеспечивая баланс между отсутствием шумов и разборчивостью речи. На основе результатов можно утверждать, что предложенный метод фильтрации при дальнейшем исследовании и развитии может показать хорошие результаты. Необходимо учитывать, что представленный в данной работе способ реставрации не является самым простым из возможных, поскольку для достижения желаемого результата предполагает несколько этапов обработки речи, а также требует предварительной подготовки речевой базы. Преимуществом данного метода является то, что отреставрированная запись почти не содержит музыкальных шумов. Однако нужно отметить, что речь на отреставрированной записи будет звучать глухо при достаточно низком отношении сигнал/шум по сравнению с исходной записью. Можно потенциально добиться улучшения качества речи на отреставрированной записи при применении СММ, разработанных на основе не целых слов, а последовательности фонем. Описанная технология имеет хорошие перспективы для анализа звуковых записей в криминалистике, реставрации звуковых записей в сфере профессиональной звуковой обработки, например в киноиндустрии, а также при записи голосовых сообщений в условиях с большим количеством источников посторонних шумов.

#### *Литература*

1. Sameti H., Sheikhzadeh H., Li Deng, Brennan R.L. HMM-Based Strategies for Enhancement of Speech Signals Embedded in Nonstationary Noise // IEEE Transactions on Speech and Audio Processing. 1998.Vol. 6. No. 5. P. 445–455. DOI: 10.1109/89.709670
2. Парамонов П.А. Методы, алгоритмы и устройства распознавания речи в ассоциативной осцилляторной среде : дис. ... канд. техн. наук: 05.13.05 / НИУ «МЭИ». М., 2015. 147 с.
3. Boll S.F. Suppression of acoustic noise in speech using spectral subtraction // IEEE Transactions on Acoustics, Speech and Signal Processing. 1979. Vol. 27. No. 2. P. 113–120. DOI: 10.1109/TASSP.1979.1163209
4. Ephraim Y., Malah D. Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator // IEEE Transactions on Acoustics, Speech Signal Processing. 1984. Vol. asp-32. No. 6. P. 1109–1121. DOI: 10.1109/TASSP.1984.1164453
5. Lim J.S., Oppenheim A.V. Enhancement and Bandwidth Compression of Noisy Speech // Proceeding of the IEEE. 1979. Vol. 67. No. 12. P. 1586–1604. DOI: 10.1109/PROC.1979.11540
6. Abd El-Fattah, M.A., Dessouky, M.I., Abbas, A.M., Diab, S.M., El-Rabaie, E.M., Al-Nuaimy, W., Alshebeili,

- S.A., Abd El-Samie, F.E. Speech enhancement with an adaptive Wiener filter // *International Journal of Speech Technology*. 2013. Vol. 17. Pp. 53–64. DOI: <https://doi.org/10.1007/s10772-013-9205-5>
7. Парфенов В.И., Жуков М.М., Кривцов Е.А. Повышение эффективности шумоподавления при ви-неровской фильтрации речевых сигналов // *Вестник Воронежского института МВД России. Компьютерные и информационные науки*. 2019. № 4. С. 137–145. EDN MUVADS.
8. Bingyin Xia, Changchun Bao. Wiener filtering based speech enhancement with Weighted Denoising Auto-encoder and noise classification // *Speech Communication*. 2014. Vol. 60. P. 13–29. DOI: 10.1016/j.specom.2014.02.001
9. Dusan S., Rabiner L. On the relation between maximum spectral transition positions and phone boundaries // *Proc. Interspeech 2006*, paper 1317-Mon3CaP3. DOI: 10.21437/Interspeech.2006-230
10. Rabiner L.R. A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition // *Proceedings of the IEEE*. 1989. Vol. 77. No. 2. P. 86–120. DOI: 10.1109/5.18626

### References

1. Sameti H., Sheikhzadeh H., Li Deng, Brennan R.L. (1998) HMM-Based Strategies for Enhancement of Speech Signals Embedded in Nonstationary Noise. *IEEE Transactions on Speech and Audio Processing*. Vol. 6. No. 5. Pp. 445–455. DOI: 10.1109/89.709670
2. Paramonov P.A. (2015) *Metody, algoritmy i ustroistva raspoznavaniya rechi v assotsiativnoi ostsillyatornoi srede* [Methods, algorithms and speech recognition devices in an associative oscillatory environment : Cand. Sci. Thesis (Technical, specialty 05.13.05)]. Moscow : Moscow Power Engineering Institute Publishing. 147 p. (In Russian).
3. Boll S.F. (1979) Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech and Signal Processing*. Vol. 27. No. 2. Pp. 113–120. DOI: 10.1109/TASSP.1979.1163209
4. Ephraim Y., Malah D. (1984) Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator. *IEEE Transactions on Acoustics, Speech Signal Processing*. Vol. asp-32. No. 6. Pp. 1109–1121. DOI: 10.1109/TASSP.1984.1164453
5. Lim J.S., Oppenheim A.V. (1979) Enhancement and Bandwidth Compression of Noisy Speech. *Proceeding of the IEEE*. Vol. 67. No. 12. Pp. 1586–1604. DOI: 10.1109/PROC.1979.11540
6. Abd El-Fattah, M.A., Dessouky, M.I., Abbas, A.M., Diab, S.M., El-Rabaie, E.M., Al-Nuaimy, W., Alshebeili, S.A., Abd El-Samie, F.E. (2013). Speech enhancement with an adaptive Wiener filter. *International Journal of Speech Technology*. Vol. 17. Pp. 53–64. DOI: <https://doi.org/10.1007/s10772-013-9205-5>
7. Parfenov V.I., Zhukov M.M., Krivtsov E.A. (2019) Noise reduction enhancement under Wiener filtering. *Vestnik of Voronezh Institute of the Ministry of Interior of Russia*. No. 4. Pp. 137–145 (In Russian).
8. Bingyin Xia, Changchun Bao (2014) Wiener filtering based speech enhancement with Weighted Denoising Auto-encoder and noise classification. *Speech Communication*. Vol. 60. Pp. 13–29. DOI: <https://doi.org/10.1007/s10772-013-9205-5>
9. Dusan S., Rabiner L. (2006) On the relation between maximum spectral transition positions and phone boundaries. *Proc. Interspeech 2006*, paper 1317-Mon3CaP3. DOI: 10.21437/Interspeech.2006-230
10. Rabiner L.R. (1989) A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition. *Proceedings of the IEEE*. Vol. 77. No. 2. Pp. 86–120. DOI: 10.1109/5.18626