

# Интеллектуальная обработка речевой информации в задачах шумоочистки для средств коммуникации на объектах цифровой экономики

А. М. Алюшин

Российский экономический университет  
им. Г.В. Плеханова; Национальный исследовательский  
ядерный университет “МИФИ”  
Москва, Россия  
alyshin@list.ru

Н. М. Леонова<sup>1</sup>, А. Д. Модяев<sup>2</sup>

Национальный исследовательский ядерный университет  
“МИФИ”  
Москва, Россия  
<sup>1</sup>nMLEONOVA@mephi.ru, <sup>2</sup>ADMODYAEV@mephi.ru

**Аннотация.** Обоснована актуальность разработки методических и технических средств шумоочистки акустических сигналов. Дан анализ применяемых на практике программных средств шумоочистки, выявлены их основные недостатки. В работе рассматривается подход, базирующийся на интеллектуальной цифровой обработке зашумленного акустического сигнала с неизвестными параметрами шума и помех. Подход реализует методику распознавания параметров шума и помех при обработке двумерных изображений, получаемых путем трансформации зашумленного акустического сигнала в динамические сонограммы. Это позволяет распознать и классифицировать различные типы помех, а также их параметры на основе использования специальных инструментов обработки графической информации. Рассмотрена технология автоматического анализа зашумленной акустической информации в рамках предложенного подхода.

**Ключевые слова:** интеллектуальная обработка; распознавание и классификация источников шума; речевая подпись

## I. ВВЕДЕНИЕ

В современном цифровом мире вопрос повышения качества, в частности, разборчивости и естественности звучания, зашумленных звуковых сигналов не теряет своей актуальности, так как звуковая составляющая – неотъемлемая часть этого цифрового мира [1–5]. На сегодняшний день существует не такое большое количество программ (как в России, так и в мире), которые имеют такую возможность, как автоматическое очищение звуковых файлов от различного рода помех. Среди них можно выделить такие программы, как «Adobe Audition», «Sound Cleaner», «iZotope RX», которые успешно справляются с шумами простой сложности, например, импульсным шумом.

Но когда речь заходит о сложных помехах и шумах, имеющих различную физическую природу, эффективность данных программ значительно снижается, а то и вообще

---

Исследование выполнено по гранту Российского научного фонда (проект №19-71-30008) в Российском экономическом университете им. Г.В. Плеханова

отсутствует. Типичным примером в этом плане является такой тип помех, как “речевой хор”, который предполагает наличие большого количества различных дикторов. Задача шумоочистки в этом случае заключается в выделении из общего акустического потока одного единственного полезного сигнала. Проблема усугубляется тем обстоятельством, что разборчивый, громкий и внятный голос на переднем плане является шумом, а еле слышный, заикающийся и слабо разборчивый голос на заднем плане является полезной информацией, которую мы хотим выделить. Данную задачу ни одна из существующих систем шумоочистки не решает на должном уровне.

Целью исследования является разработка технологии решения задач шумоочистки такого рода.

## II. СОСТОЯНИЕ ИССЛЕДОВАНИЙ В ДАННОЙ ОБЛАСТИ

Основной принцип работы классических программ, как правило, заключается в последовательном переборе рекомендованных фильтров с изменяющимися параметрами (например, изменение размера и формы апертуры при использовании медианного фильтра), и сравнение полученных новых результатов с имеющимися.

Основная проблема при решении такого рода задач заключается в том, что обычно затруднительно, либо просто невозможно правильно задать программе настройки для осуществления фильтрации полезной информации. Также, как правило, у пользователя отсутствует правильное понимание того, что нужно делать.

Типичный состав и последовательность выполняемых основных операций для существующих программ шумоочистки представлены на рис. 1. Как правило, это итерационный процесс, предполагающий последовательное применение фильтров с различными характеристиками.

Для простых шумов, например, узкополосных низкочастотных, либо высокочастотных решение вопросов о целесообразности применения того, либо иного типа фильтра решается достаточно просто. Например, для указанных случаев – применением соответственно

фильтров высоких и низких частот. Аналогично обстоит дело с выбором параметров таких фильтров в зависимости от значений шумовых сигналов.

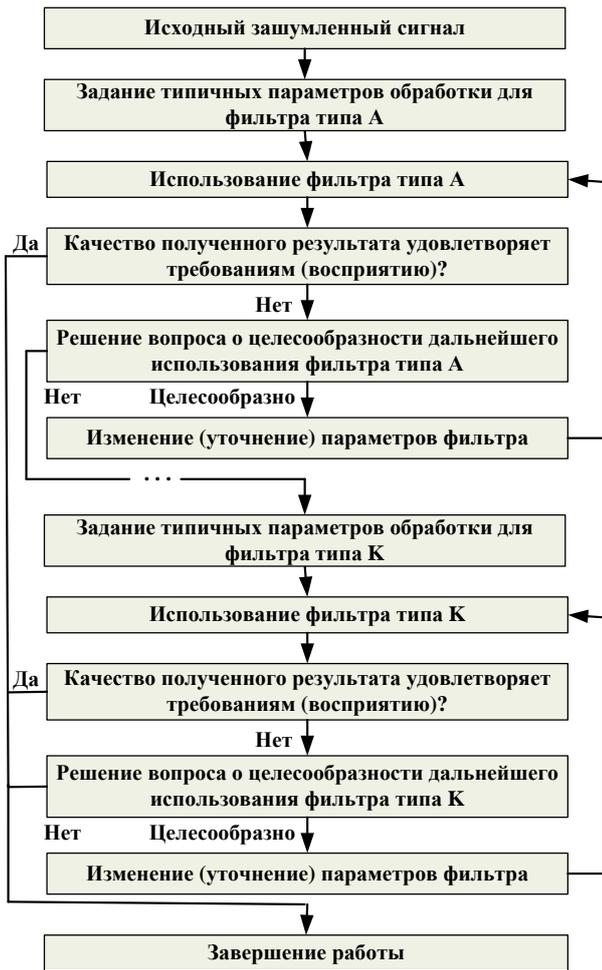


Рис. 1. Алгоритм работы программ шумоочистки сигналов

Для улучшения такого общепринятого показателя качества, как SNR (Signal to Noise Ratio) при наличии шумов и помех различной физической природы практическое использование рассмотренного выше подхода малоэффективно.

### III. ПРЕДЛАГАЕМЫЙ ПОДХОД

Сущность предлагаемого подхода заключается в использовании специальных графических инструментов, дающих возможность осуществить редактирование в ручном, либо в полностью автоматическом режиме графического представления зашумленного акустического сигнала. Для осуществления визуализации исходного акустического сигнала предлагается использовать его преобразование в частотную область с последующим графическим представлением в виде сонограммы [6–10]. Получение акустической информации в виде изображения в одном из стандартных графических форматов позволяет осуществить его очистку (редактирование) с помощью специализированных инструментов.

Общий принцип работы разработанных алгоритмов и программного обеспечения шумоочистки представлен на рис. 2.

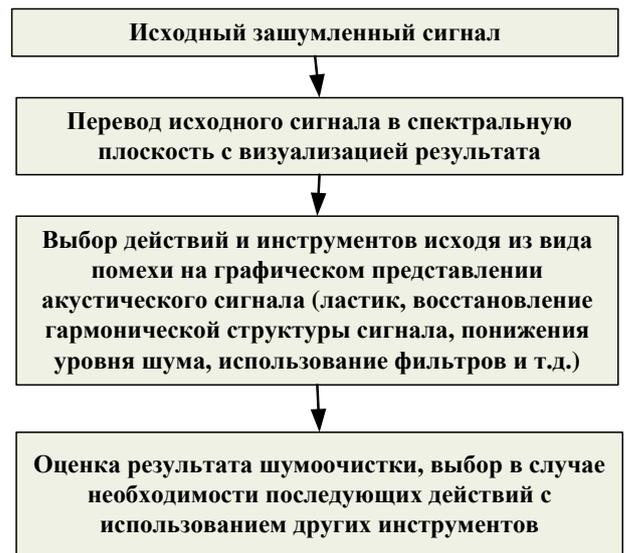


Рис. 2. Алгоритм работы разработанного программного комплекса «ДАК»

Основная идея разработанных алгоритмов шумоочистки заключалась в реконструкции и восстановлении по оставшейся в спектрограмме полезной смеси характеристик незамаскированных «следов» участков речевых вокализмов (вокализм – несущая частота и амплитуда элементарных узкополосных процессов – вейвлетов), их гармонической и формантной структуры с последующим синтезом по вновь созданным спектрально-временным описаниям новой волновой формы речеподобного сигнала, близкого по звучанию к исходному.

Для повышения эффективности процесса спектрально-временной реставрации искаженной шумами и помехами речи предлагалось использовать заранее собранную голосовую базу данных известного диктора, чья речь использовалась для создания полезной смеси.

На первом этапе шумоочистки с помощью комплекса «ДАК» строится изображение спектрограммы звукового сигнала по разработанному усовершенствованному алгоритму кратковременного Фурье-анализа. Так, для соотношения -10 дБ результат представлен на рис. 3.

На втором этапе графическая спектрограмма смеси сигнала с шумом подвергается различным видам обработки и, прежде всего, сегментированию. Данная операция дает возможность выделить области изображения, подлежащие очистке графическими инструментами.

На третьем этапе осуществляется очистка изображения от помех, которые находятся между «следами» различных слов и/или фраз. Для этого используется набор стандартных и специализированных инструментов редактирования изображений «ДАК», например, «Ластик»

и другие. На данном этапе также осуществляется очистка от графической информации, которая не принадлежала к разряду полезной

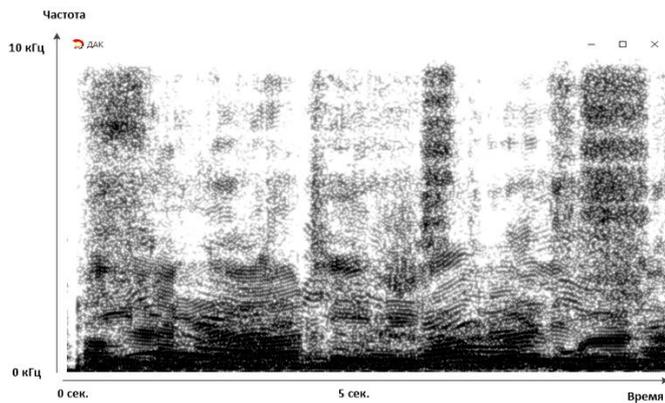


Рис. 3. Получение спектрограммы шумового сигнала

. В данном случае это информация, которая не относится к человеческой речи, например, представленной в виде обертонов. Для этого используются инструменты графического редактирования «Порог», «Ластик», «Дорисовка» (реконструкция существующих выделенных гармоник), «Удаление» (удаление участков определенной площади, к примеру, без признаков речи или бесполезной речи).

На четвертом этапе по специально разработанному алгоритму происходило восстановление гармонической структуры речевого сигнала по оставшимся следам. Типичный результат восстановленной гармонической структуры представлен на рис. 4. Исходный фрагмент гармонической структуры с частично потерянной информацией показан на рис. 4а. Восстановленная гармоническая структура с синтезированными обертонами показана на рис. 4б.

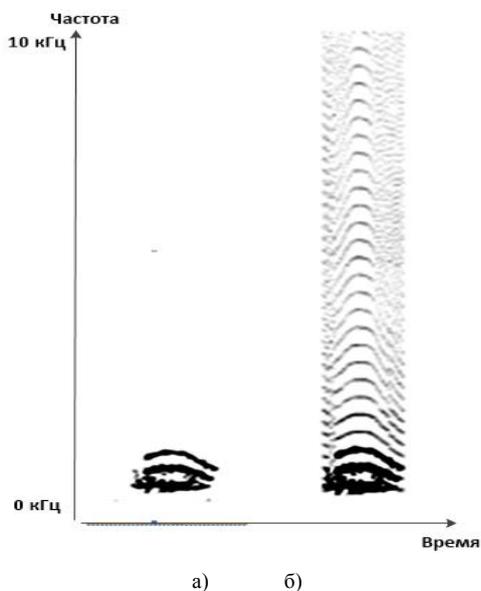


Рис. 4. Восстановление гармонической структуры.

На последнем этапе восстановленное по гармонической структуре изображение спектрограммы подвергается вейвлет-синтезу для получения звукового сигнала. В результате получается очищенный звуковой сигнал высокого качества.

#### IV. ИНТЕЛЛЕКТУАЛЬНАЯ ОБРАБОТКА ИНФОРМАЦИИ

Разработанные средства шумоочистки реализуют ряд интеллектуальных технологий на этапе обработки графической информации:

- распознавание графических объектов, характерных для шумовых составляющих;
- классификация выделенных графических объектов с шумовыми признаками;
- выбор оптимального пути шумоочистки для выделенных объектов различных классов;
- обучение на примерах с использованием программно реализованных нейросетей;
- синтез недостающих графических участков поврежденных вокализмов в соответствии с результатами их распознавания.

Для распознавания и классификации объектов двумерной сонограммы целесообразно использовать методические и технические средства, разработанные для решения задач распознавания графических образов. В работе для этой цели исследовалась эффективность применения средств распознавания графических образов на основе алгоритма Виолы-Джонса [11, 12].

Достоинствами данного алгоритма и его многочисленных модификаций являются:

- возможность использования для описания широкой номенклатуры графических объектов так называемых примитивов Хаара [11];
- возможность реализации процедуры поиска графических объектов с заданными признаками в окне произвольных размеров;
- возможность естественной интеграции фрагментов обучаемых нейросетей в состав выполняемых процедур и функций;
- наличие в настоящее время хорошо развитой методической и алгоритмической базы для работы с графическими объектами различных классов, например, с изображением лица человека и техническими объектами;
- возможность создания специализированной базы графических образов типовых акустических помех, представленных в виде двумерных сонограмм в соответствии с рассмотренным ранее подходом.

На рис. 5 представлен пример распознавания и классификации с помощью алгоритма Виолы-Джонса областей сонограммы, содержащей характерные признаки широкополосных импульсных помех. Для автоматического выделения границ графических областей

сложной формы был реализован итерационный процесс поиска объектов с заданными параметрами в плавающем окне с уменьшающимися (в два раза на каждой последующей итерации) размерами – 1-6.

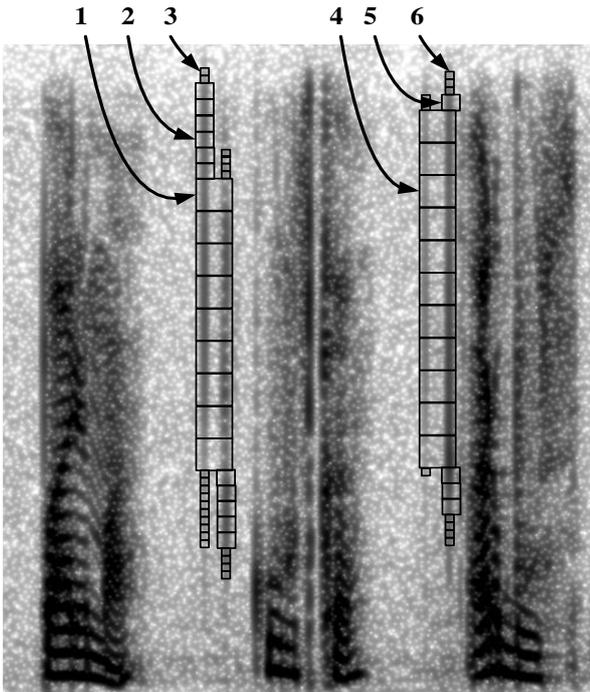


Рис. 5. Распознавание широкополосных импульсных помех

На рис. 6 представлен результат распознавания областей с белым шумом (1-2 – окна различных размеров).

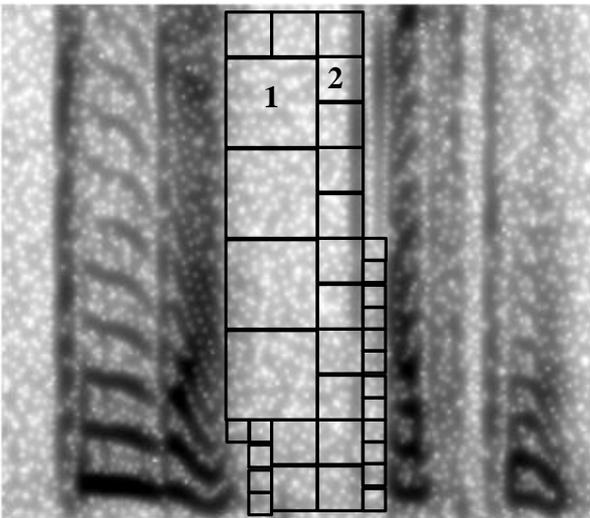


Рис. 6. Распознавание областей с белым шумом

## V. ЗАКЛЮЧЕНИЕ

Таким образом, разработанная технология с использованием графические изображения узкополосных динамических спектрограмм дает возможность осуществить шумоочистку акустической информации при наличии шумов различной физической природы. Интеллектуальные средства шумоочистки позволяют осуществить автоматический режим шумоочистки.

## СПИСОК ЛИТЕРАТУРЫ

- [1] Rath T., Manmatha R. Word image matching using dynamic time warping // Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2003. Vol. 2. P. 521–527.
- [2] Zhu G., Zheng Y., Doermann D., Jaeger S. Multi-Scale structural saliency for signature detection // Proc. IEEE Conf. Computer Vision and Pattern Recognition. 2007. P. 1–8.
- [3] Fang B., Leung C.H., Tang Y.Y., Tse KW., Kwok P.C.K., Wong Y.K. Off-Line signature verification by the tracking of feature and stroke positions // Pattern Recognition. 2003. Vol. 36. No. 1. P. 91–101. DOI: 10.1016/S0031-3203(02)00061-4
- [4] Zheng Y., Li H., Doermann D. Machine printed text and hand writing identification in noisy document images // IEEE Trans. Pattern Analysis and Machine Intelligence. 2004. Vol. 26. No. 3. P. 337-353. DOI: 10.1109/TPAMI.2004.1262324
- [5] Secure Hash Standard (SHS). Federal information processing standards publication, FIPS PUB 180-4. 2015. 31 p. <http://dx.doi.org/10.6028/NIST.FIPS.180-4>
- [6] Дворянкин С.В. Речевая подпись. М. РИО. 2003. 184 с.
- [7] Alyushin M.V., Alyushin A.M., Kolobashkina L.V. Human face thermal images library for laboratory studies of the algorithms efficiency for bioinformation processing.// Proc. of the 11th IEEE International Conference on Application of Information and Communication Technologies, AICT 2017. DOI: <http://dx.doi.org/10.26583/bit.2017.2.01>
- [8] Alyushin A.M. Biologically inspired physical model of the vocal tract for the tasks of recognition of the current psycho-emotional state of a person // Advances in Intelligent Systems and Computing. 2020. Vol. 948, P. 15–21
- [9] Vorobiov V.I. Inter component phase processing of speech signals for their recognition and identification of announcers // XVIII Session of the Russian Acoustical Society, Taganrog, September 11-15. 2006. P. 529–523.
- [10] Козлов Ю.Е. Подходы к определению надежности мультимодальной трехмерной динамической подписи // Безопасность информационных технологий. 2018. № 1. С. 74–80. DOI: <http://dx.doi.org/10.26583/bit.2018.1.07>
- [11] Alyushin M.V., Alyushin V.M., Kolobashkina L.V. Optimization of the data representation integrated form in the viola-jones algorithm for a person's face search // Procedia Computer Science. 2018. Vol. 123. P. 18–23.
- [12] Kolobashkina L.V., Alyushin M.V. Analysis of the possibility of the neural network implementation of the Viola-Jones algorithm //Advances in Intelligent Systems and Computing. 2020. Vol. 948. P. 232–239