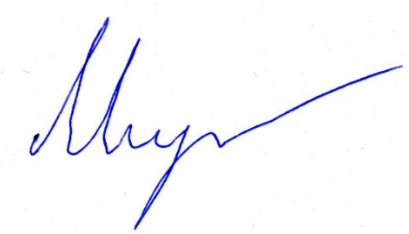
Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский Нижегородский государственный
университет им. Н.И. Лобачевского»

На правах рукописи



Миронов Никита Андреевич

Пространственная обработка речевых сигналов на фоне интенсивных распределенных помех

Специальность 01.04.06 – Акустика

Диссертация на соискание ученой степени кандидата
физико-математических наук

Научный руководитель,
доктор физико-математических наук, доцент
Канаков Владимир Анатольевич

Нижний Новгород – 2020

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Актуальность темы исследования и степень ее разработанности

Задача разделения акустических источников и выделения полезного сигнала из акустической смеси решается много десятилетий [1-5]. Человеческий мозг демонстрирует феноменальный результат по обработке сложной шумовой акустической обстановки и способен выделять нужный сигнал при большом числе мешающих источников. Но в области цифровой обработки сигналов – это сложнейшая задача.

Техника выделения голоса одного человека из смеси голосов может найти применение в огромном числе речевых приложений, а также в работе органов внутренних дел и служб безопасности. Практическая реализация предложенного решения может быть осуществлена в аэропортах, вокзалах железнодорожного транспорта и в других местах массового скопления граждан для регистрации акустической обстановки и выделении речи лиц, склонных к организации массовых беспорядков и совершению террористических актов.

Современные решения по выделению голоса в меняющейся динамической обстановке связаны с применением микрофонных решеток (МР) [6-8], которые имеют ряд преимуществ по сравнению с одноканальными аудиосистемами. В связи с различными сферами применения микрофонных решеток способы разделения акустических колебаний, алгоритмы обработки и геометрия таких систем крайне разнообразны [9-10].

Эффективность микрофонных решеток в задачах выделения речи целевого диктора определяется возможностью реализовать пространственную фильтрацию акустических сигналов. Большинство алгоритмов, описанных в научно-технической литературе, работают в частотной области, основаны на использовании узкополосного приближения и осуществляют обработку речевого сигнала отдельно в каждом поддиапазоне частот. Введение оптимальных весовых комплексных коэффициентов в каждый сигнал соответствующей полосы частот позволяет максимизировать целевую функцию пространственной фильтрации,

например, отношение сигнал/помеха для целевого диктора.

Однако такой способ пространственной фильтрации вносит дополнительные частотные искажения в полезный сигнал. Кроме того, число отсчетов сигнала в каждой полосе частот на длительности интервала анализа становится крайне малым, что ухудшает выполнение приближения независимости отсчетов и снижает эффективность статистических методов обработки [10].

С другой стороны, базовый алгоритм пространственной фильтрации, известный как «delay-and-sum» и реализуемый во временной области, лишен этих недостатков, прост в реализации и практически не требует затрат процессорного времени на обработку [11]. Базовый алгоритм может быть реализован и в частотной области, но его эффективность снижается в сравнении с временной областью для определенных задач [12]. Единственным недостатком алгоритма «delay-and-sum» является относительно высокий средний уровень боковых лепестков реализуемого фильтра пространственных частот, что проявляется снижением эффективности подавления большого числа распределенных в пространстве источников помех. Известные способы подавления боковых лепестков фильтра пространственных частот, основанные на вычислении оптимальных частотно-зависимых комплексных весовых коэффициентов, в настоящее время реализуются в частотной области обработки сигналов, так как использование быстрого преобразования Фурье позволяет значительно ускорить процедуру частотной фильтрации.

На основании изложенного можно констатировать, что широко применяемые в настоящее время алгоритмы пространственной обработки речевых сигналов в частотной области достигли некоторого предела качества выделения речи целевого диктора из акустической смеси, тогда как потенциальные возможности оптимальных алгоритмов, реализованных во временной области, остаются не исследованными.

Реализация алгоритма оптимальной пространственной фильтрации речевых сообщений на фоне пространственно-распределенных источников помех во временной области, с использованием полной полосы частот без разбиения в реальном масштабе времени, позволит сочетать достоинства алгоритма «delay-and-

sum» и оптимальных методов пространственной фильтрации в частотной области.

Наличие практической потребности в разработке алгоритма по выделению речевого сообщения из помех от сторонних источников речи во временной области с классом качества, обеспечивающим понимание передаваемой речи и соответствующим современному уровню технического прогресса, обусловили необходимость и актуальность решения задач, рассматриваемых в диссертации.

Цель диссертации состоит в разработке алгоритма обработки речевого сигнала микрофонной решеткой во временной области, позволяющего выделять речевые сообщения из любой точки пространства наблюдения с максимальным отношением сигнал/помеха, независимо от взаимного расположения целевого диктора и других дикторов, являющихся источниками речевых помех.

Для достижения указанной цели в диссертации необходимо было решить **следующие задачи**:

1. Провести анализ существующих методов разделения акустических сигналов.
2. Провести анализ существующих алгоритмов обработки сигналов микрофонными решетками.
3. Разработать алгоритм обработки речевого сигнала микрофонной решеткой во временной области, максимизирующий отношение сигнал/помеха на выходе решетки.
4. Провести численный эксперимент по выделению речевых сообщений из помех многопозиционной акустической системой микрофонов.
5. Провести исследование эффективности работы предложенного алгоритма в условиях, максимально приближенных к реальным.

Научная новизна диссертации состоит в следующем:

1. Предложен метод пространственной фильтрации речевых сигналов во временной области, основанный на введении временных задержек, зависящих от пространственных координат, и расчете оптимальных весовых коэффициентов

микрофонной решетки.

2. Предложена оптимальная конфигурация микрофонной решетки в плоскости размещения источников звука для открытого пространства наблюдения, ограниченного периметром.

3. Разработан алгоритм обработки речевого сигнала микрофонной решеткой во временной области, позволяющий выделять речевые сигналы из любой точки пространства наблюдения с классом качества, обеспечивающим понимание передаваемой речи.

Теоретическая значимость

Описан теоретический подход к обработке речевых сигналов микрофонной решеткой во временной области, обеспечивающий максимизацию выходного отношения сигнал/помеха за счет введения временных задержек и адаптивного формирования вектора оптимальных весовых коэффициентов микрофонов на интервалах стационарности.

Практическая значимость

Разработанный алгоритм обработки речевых сигналов микрофонной решетки, размещенной по периметру акустической сцены, позволяет выделять сигналы источников, расположенных в любой точке акустической сцены, с максимальным отношением сигнал/помеха.

Предложенный в работе алгоритм обработки сигналов, реализованный во временной области, устойчив к реверберации звука в помещении, может быть применим для выделения слабых сигналов на фоне более мощных распределенных в пространстве источников помех, реализуем в режиме реального времени.

По результатам диссертационного исследования имеется возможность разработки устройства для выделения речевых сигналов из смеси широкополосных помех от пространственно-разнесенных источников из любой точки пространства наблюдения с максимальным отношением сигнал/помеха.

Методология и методы исследования

Теоретические и экспериментальные исследования базируются на использовании методов физической акустики, вычислительной математики, методов статистического анализа, методов математического и компьютерного моделирования.

Личный вклад соискателя состоит в:

- получении аналитических выражений для обработки речевого сигнала микрофонной решеткой во временной области на интервалах стационарности;
- разработке компьютерного комплекса по выделению речевых сообщений из помех для микрофонной решетки, реализующего обработку сигналов предложенным оригинальным алгоритмом;
- обработке результатов численного эксперимента.

Защищаемые положения:

1. Разделение зарегистрированных микрофонной решёткой нескольких синхронных речевых сообщений может быть осуществлено на основе адаптивного алгоритма цифровой обработки сигналов, основанного на введении временных задержек и расчете циклической оценки оптимального на интервалах стационарности вектора действительных весовых коэффициентов для сигналов от микрофонов в составе решётки. При использовании N микрофонов в составе решётки достигается увеличение отношения сигнал/помеха выделенного речевого сообщения не менее чем в N раз.

2. Разработанный алгоритм цифровой обработки сигналов решётки микрофонов, размещенных эквидистантно по периметру контролируемого помещения, позволяет выделять речевые сообщения целевых дикторов вне зависимости от их взаимного расположения в любом месте акустической сцены с максимальным отношением сигнал/помеха.

3. Компьютерное моделирование работы предложенного алгоритма цифровой обработки сигналов микрофонной решётки показало, в заданных условиях модели, его реализуемость в режиме реального времени, работоспособность при выделении слабых полезных речевых сигналов на фоне более мощных мешающих речевых сообщений при отношении сигнал/помеха более минус 20 дБ, способность выделения речи движущегося по известной траектории диктора, устойчивость к эффекту реверберации звука в контролируемом помещении.

Степень достоверности полученных результатов подтверждается согласованностью принятых при теоретическом анализе моделей акустической обстановки с общеизвестными принципами физической акустики, применением апробированных в практической радиолокации алгоритмов цифровой обработки сигналов, максимизирующих отношение сигнал/помеха на выходе решетки, использованием большого количества реальных фонограмм при проведении компьютерного моделирования работы алгоритма в условиях, максимально приближенных к реальным.

Апробация

По материалам данной работы были сделаны доклады на XVIII научной конференции по радиофизике, посвященной дню радио, Н. Новгород, ННГУ, 12-16 мая 2014 года, XVII Международной конференции «Цифровая обработка сигналов и ее применение DSPA-2015», Москва, 25-27 марта 2015 года, XXI Международной научно-практической конференции «Информационные системы и технологии», Н. Новгород, НГТУ, 15-17 апреля 2015 года, XXII Международной научно-технической конференции «Информационные системы и технологии», Н. Новгород, НГТУ, 22 апреля 2016 года, Международной научно-практической конференции «Наука XXI века: открытия, инновации, технологии», Смоленск, 30 апреля 2016 года, XXI Нижегородской сессии молодых ученых (естественные, математические науки), Княгинино, НГИЭУ, 2016 года, XX научной конференции

по радиофизике, посвященной 110-летию со дня рождения Г.С. Горелика, Н. Новгород, ННГУ, 12-20 мая 2016 года, XV международной научно-практической конференции «Перспективы развития науки и образования», Москва, 31 марта 2017 года, Юбилейной XXIII международной научно-технической конференции, посвященной 100-летию НГТУ им. Р. Е. Алексеева «Информационные системы и технологии», Н. Новгород, НГТУ, 21 апреля 2017 года, XXI научной конференции по радиофизике, Н. Новгород, ННГУ, 15-22 мая 2017 года, XXII научной конференции по радиофизике, посвященной 100-летию Нижегородской радиолаборатории, Н. Новгород, ННГУ, 15-29 мая 2018 года, XXIII Нижегородской сессии молодых ученых (технические, естественные, математические науки), Н. Новгород, ННГУ, 22-23 мая 2018 года.

Публикации

Основные материалы диссертации изложены в 17 работах [A1-A17], 4 из которых опубликованы в журналах, включенных в перечень ВАК [A1-A4], 2 работы входят в мировые индексы цитирования (SCOPUS, Web of Science) [A1-A2].

Структура и объем диссертации

Диссертация состоит из введения, четырех глав, заключения и списка литературы. Общий объем составляет 108 страниц. В диссертации 54 рисунка, 52 формулы и 14 таблиц.

ГЛАВА 1. ВЫДЕЛЕНИЕ РЕЧЕВЫХ СИГНАЛОВ ИЗ ИХ СМЕСИ

В данной главе рассматриваются методы разделения акустических источников и способы выделения сигнала полезного источника из акустической смеси. Приведены методы с использованием одноканальных систем и многопозиционных систем – микрофонных решеток. Рассмотрены виды акустических систем, используемых в настоящее время. Рассмотрены алгоритмы обработки сигналов в микрофонных решетках.

1.1. Проблема выделения речевого сигнала из акустической смеси

> «One of our most important faculties is our ability to listen to, and follow, one speaker in the presence of others. This is such a common experience that we may take it for granted; we may call it «the cocktail party problem»…»
>
> Colin Cherry, 1957 [2]

Проблема выделения голоса одного человека из акустической смеси (Рисунок 1 [13]) получила название «Cocktail party problem» [1-5]. Термин был введен Колином Черри [1-2]. Черри определил данную проблему как психоакустический феномен, который относится к способности человека избирательно следить и распознавать один источник звука в шумной среде, где помехи представляют собой речевые сообщения от других источников звука или от других акустических источников, сигналы которых независимы.

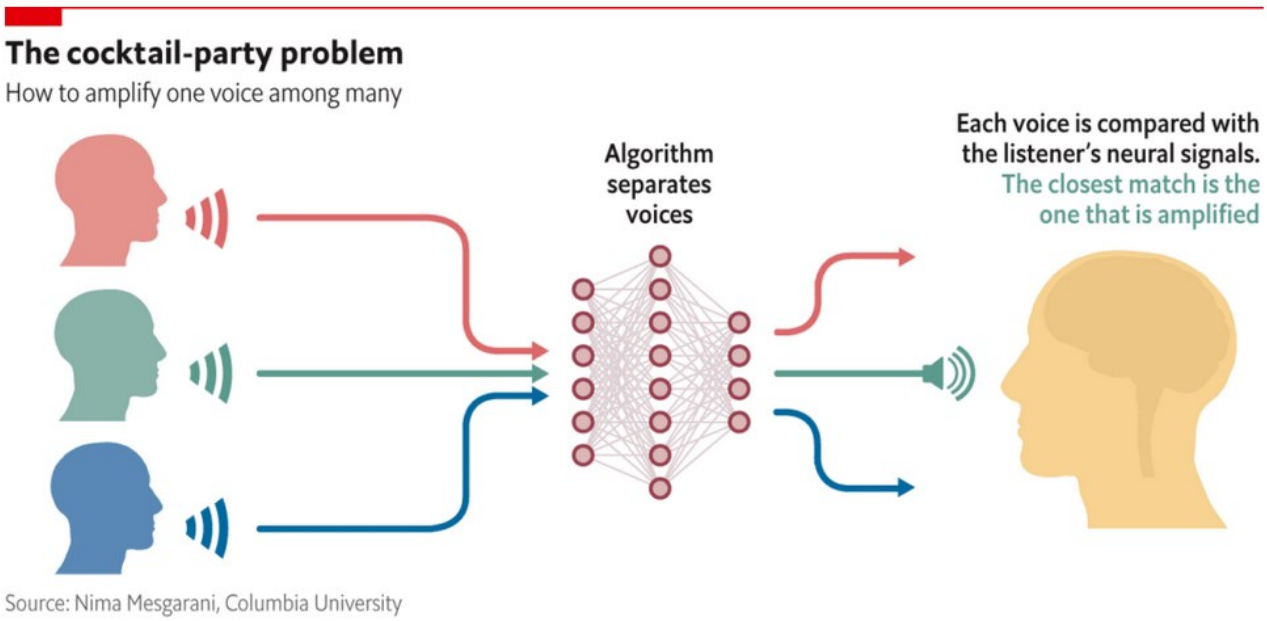


Рисунок 1 – Иллюстрация «Cocktail party problem» [13]

После фундаментальных работ Черри данной проблемой начали заниматься в различных областях науки. Специалисты в области физиологии, нейробиологии, психофизиологии, когнитивной психологии и биофизики пытались ответить на вопрос «как человеческий мозг выделяет нужный сигнал из смеси голосов?». Специалисты в области компьютерных наук и инженеры пытались понять «возможно ли разработать технику, способную решить поставленную задачу?» [4].

В 1983 году Джэнс Блоерт опубликовал работу по изучению пространственного слуха [14]. В его работе подробно освещен вопрос локализации источников звука человеком. Блоерт показывает, что сигналы, достигающие двух ушей человека различны по интенсивности и времени. Небольшие различия в этих сигналах достаточны для определения местоположения и направления входящих звуковых волн. Используя эту небольшую разницу, пространственный слух помогает мозгу осуществлять дальнейшую сложную обработку и выделять речевые сообщения в присутствии нескольких сторонних источников. Способность локализовать интересующий источник при регистрации информации двумя ушами получила название бинауральный слух.

В 1990 году Альберт Брегман вводит понятие анализа слуховой сцены (ASA) [15]. В психологии «анализ слуховой сцены» – особое направление в исследовании слухового восприятия, связанное с изучением принципов анализа человеком сложного звукового потока, возникающего в окружающей его среде [16]. Брегман утверждает, что есть много общего между слухом и зрением [15, 17, 18]. Когда мы рассматриваем визуальную сцену, края, текстуры и цвета анализируются и интерпретируются как перцептивные целостности. Точно так же звук, достигающий ушей, подвергается слуховому анализу сцены, состоящему из двух этапов: сегментация – акустическое разложение входного сигнала на набор частотно-временных областей (сегментов) и группировка – объединение сегментов одного источника в перцептивную структуру, называемую потоком.

В основе подхода Брегмана лежит закономерность того, что все частотные составляющие одного звука имеют тенденцию начинаться одновременно. Такое предположение позволяет группировать компоненты звука одного источника и отделять компоненты других источников в частотно-временном представлении.

Психофизические характеристики звука в основном включают три основные формы информации: пространственное местоположение, временную структуру и спектральную характеристику. Восприятие звука на фоне разговора нескольких человек однозначно определяется совокупностью данных трех форм. Определяющим для анализа слуховой сцены является то, что любое различие в любой из трех форм информации считается достаточным для разделения двух различных источников звука.

Дальнейшие работы по решению задачи выделения голоса из акустической смеси связаны с вычислительным анализом слуховой сцены (CASA) – вычислительный подход к анализу слуховой сцены [19]. CASA занимается автоматическим анализом акустической среды, интерпретацией дискретных звуковых событий в ней и моделированием звуковых компонентов.

Принципы вычислительного анализа слуховой сцены преследуют одну из двух целей:

1. Разработка системы, способной автоматически извлекать и отслеживать звуковой сигнал при активной голосовой смеси.

2. Разработка адаптивной слуховой системы, которая автоматически вычисляет процесс перцептивной группировки, отсутствующей в слуховой системе человека с нарушением слуха, тем самым позволяя этому человеку следить за звуковым сигналом в присутствии речеподобных помех.

Алгоритмы CASA направлены на разделение звуковых сигналов из смеси, основываясь на слуховой системе человека. Поэтому при записи помеховой обстановки используются не более двух микрофонов.

В литературе рассмотрены как моноуральные, так и бинауральные алгоритмы [20]. Алгоритмы моноуральных (один микрофон) систем CASA для разделения речи основаны на гармоничности, начале и окончании звука, амплитудно-частотной модуляции [21-23]. Бинауральные (два микрофона) системы CASA основаны на локализации звука и группировке на основе местоположения [14, 22, 24, 25].

Важнейшим бинауральным эффектом является эффект слуховой маскировки [26-27]. Данный эффект связан с процессом взаимодействия сигналов, что приводит к изменению слуховой чувствительности к маскируемому сигналу в присутствии маскирующего. Изменяется восприятие одного сигнала в присутствии другого: изменяется громкость, тембр либо второй сигнал может быть попросту не услышен. Другими словами более сильный сигнал маскирует более слабый.

Решения в области применения CASA основаны на эффекте слуховой маскировки или частотно-временной маскировки (T-F masking) [28-31]. Частотно-временная маскировка заключается в сокрытии сигналов помех в частотно-временном представлении. В своем исследовании Ванг [30] сформулировал цель алгоритмов CASA: найти идеальную бинарную маску. Значение идеальной маски в решении Ванга принимает либо 1, либо 0: 1 – в случае, если энергия сигнала полезного источника выше энергии помех и 0 – в противном случае. На Рисунке 2 [22] показано выделение речи одного диктора из «голосового коктейля» с помощью идеальной бинарной маски.

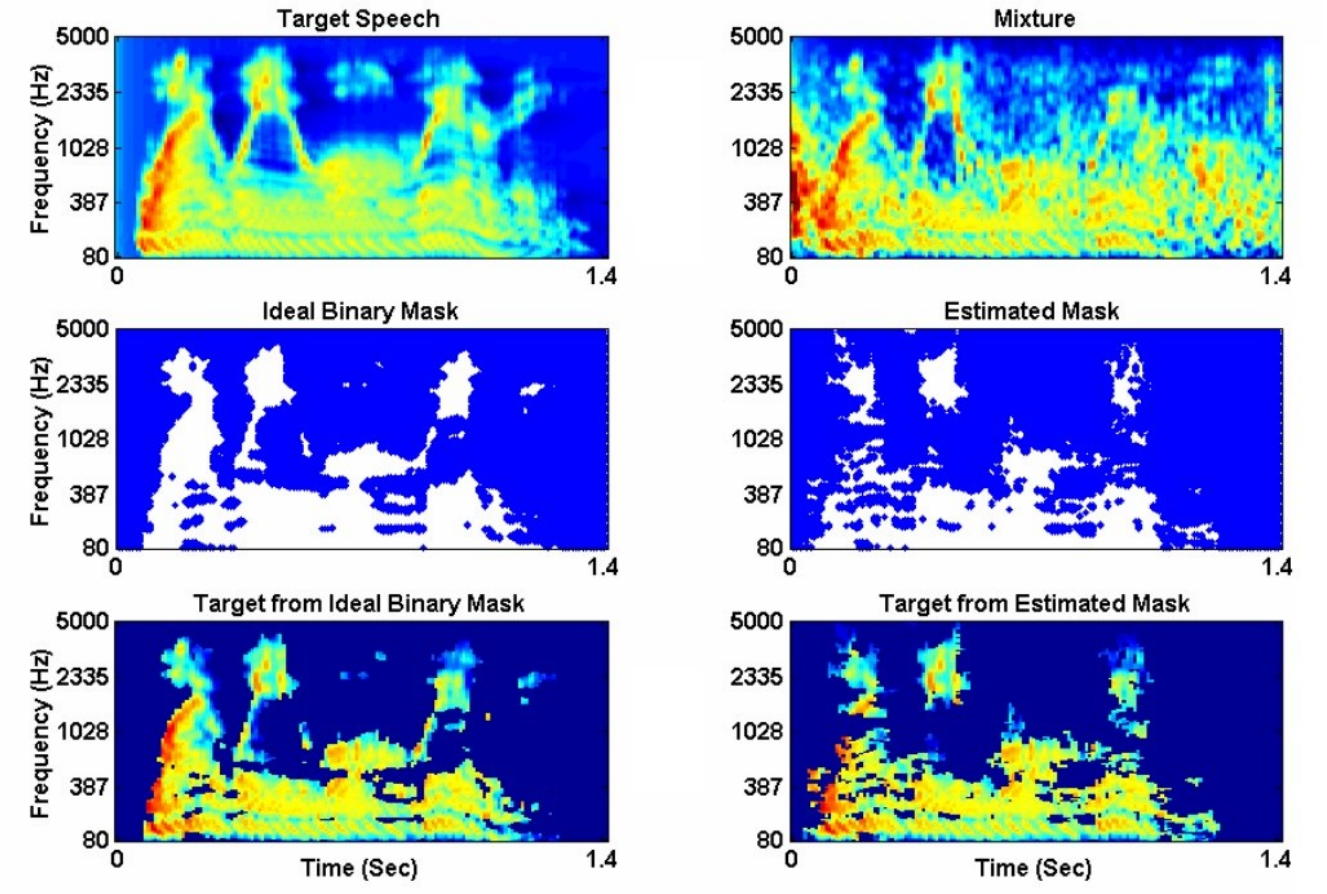


Рисунок 2 – Выделение голоса из смеси голосов методом идеальной бинарной маски [22]

Современные решения, основанные на применении частотно-временной маскировки, получили широкое распространение в задачах, когда количество одновременно действующих источников звука превышает число микрофонов.

Так, например, популярным методом оценки частотно-временной маски с использованием двух микрофонов является алгоритм DUET [32]. Этот алгоритм позволяет на основании данных, полученных от двух микрофонов, восстанавливать сигналы произвольного числа источников. Но, в силу своей специфики, чем большее число источников нужно восстановить, тем хуже этот алгоритм работает [33]. Базовое предположение и основной недостаток (для практической реализации) алгоритма: в каждый момент времени сигналы источников имеют уникальный частотный спектр – каждая частотная компонента сигнала смеси связана только с одним независимым источником.

Другим подходом к задачам разделения акустических сигналов является «слепое разделение сигналов» (BSS) [34, 35], набирающим популярность в середине 1990-х годов. Термин «слепое» используется для обозначения всех методов идентификации, основанных только на выходных наблюдениях. Отличительной чертой данного направления являлось наличие у системы нескольких входов и нескольких выходов (MIMO-системы).

В 1987 году был введен «анализ независимых компонент» (ICA) для линейной смеси, который соответствует общей структуре решения задач BSS на основе статистической независимости неизвестных источников и негауссовости сигналов [34, 36-38]. ICA стремится разложить сигналы на подкомпоненты для идентификации активности различных источников сигналов. На Рисунке 3 [39] показана схема разделения сигналов источников звука методом ICA.

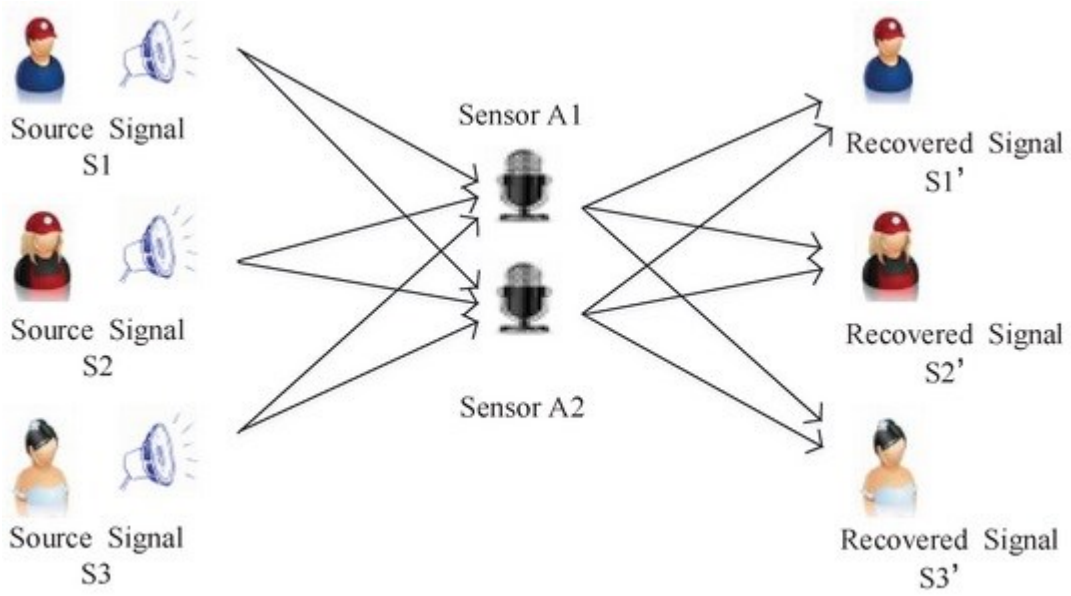


Рисунок 3 – Решение задачи разделения источников звука методом ICA [39]

Аналитически задача ICA выглядит следующим образом [36]: на вход системы микрофонов поступает вектор входных сигналов $\boldsymbol{X}$, которые представляют собой акустические смеси исходных сигналов $\boldsymbol{S}$. Матрица $\boldsymbol{A}$ – матрица смешивания сигналов, $\boldsymbol{W}$ – разделяющая матрица, причем:

$$\boldsymbol{W} = \boldsymbol{A}^{-1}. \quad (1)$$

На выходе системы формируется вектор выходных наблюдений $\boldsymbol{Y}$, представляющий собой оценки исходных сигналов. Вектор выходных наблюдений системы:

$$\boldsymbol{Y} = \boldsymbol{A}^{-1}\boldsymbol{X} = \boldsymbol{W}\boldsymbol{X}. \quad (2)$$

Алгоритмы ICA состоят в определении матрицы смешивания, используя два основных критерия:

1. Максимальная негауссовость сигналов.
2. Минимизация взаимной информации для независимости источников.

Кроме алгоритмов вычислительного анализа слуховой сцены CASA и алгоритмов анализа независимых компонент ICA для решения задачи выделения голоса из смеси в современной литературе широкое распространение получили алгоритмы пространственной фильтрации с использованием микрофонных решеток. Алгоритмы основаны на формировании диаграммы направленности в направлении на полезный источник [6-11]. Актуальные исследования связаны как с применением микрофонных решеток, в которых микрофоны расположены близко друг к другу, так и с массивами микрофонов, распределенных в пространстве случайным образом [40].

1.2. Основные сведения о микрофонных решетках

Микрофонные решетки представляют собой массив из нескольких микрофонов, объединенных совместной цифровой обработкой сигналов. Рисунок 4 [41] иллюстрирует цифровую обработку сигналов микрофонной решеткой.

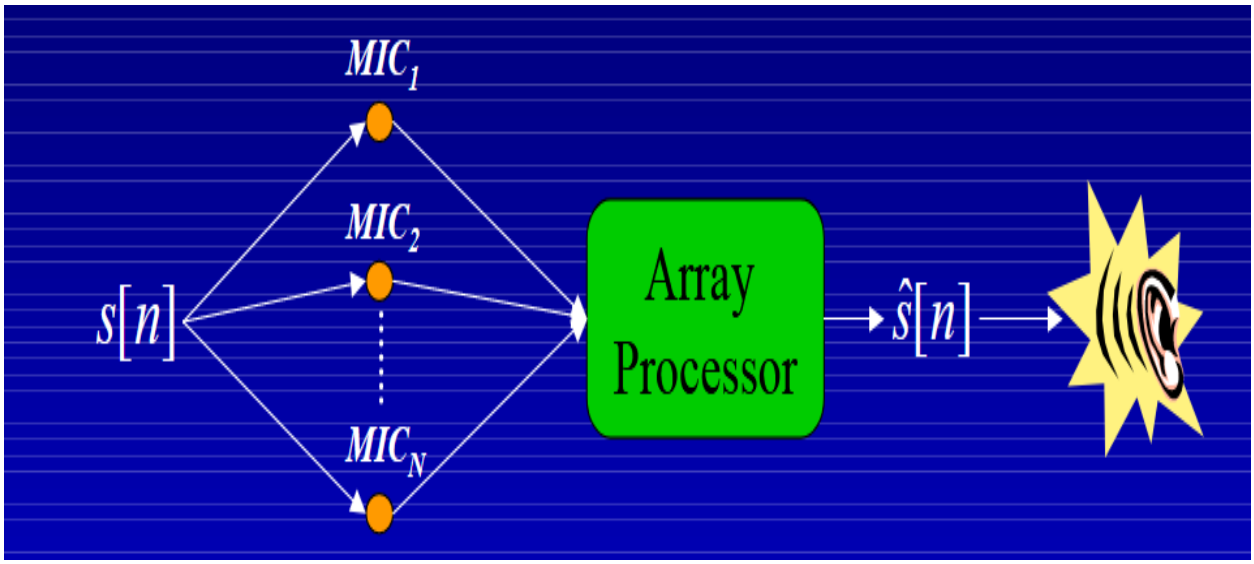


Рисунок 4 – Цифровая обработка сигналов микрофонной решеткой [41]

Направление разработки и применения микрофонных решеток активно развивается за рубежом, что подтверждается значительным числом научных трудов, например [8-10, 42, 43]. Публикаций в отечественной литературе крайне мало. В современной отечественной литературе в области применения микрофонных решеток выделяется работа к.т.н., доцента Столбова М.Б. [7], который описал не только принципы работы микрофонных решеток, классификацию основных алгоритмов обработки сигналов, но и осветил перспективы развития данной области.

Микрофонные решетки обеспечивают следующие преимущества по отношению к одноканальным системам [7]:

- направленность приема звука;
- подавление шумов точечных источников;
- подавление нестационарных шумов окружения;
- частичное ослабление реверберации;
- возможность пространственной локализации звука целевого диктора;
- возможность сопровождения двигающегося диктора и точечного источника.

Базовыми структурами микрофонных решеток являются так называемые Broadside и Endfire (Рисунок 5) [7, 44].

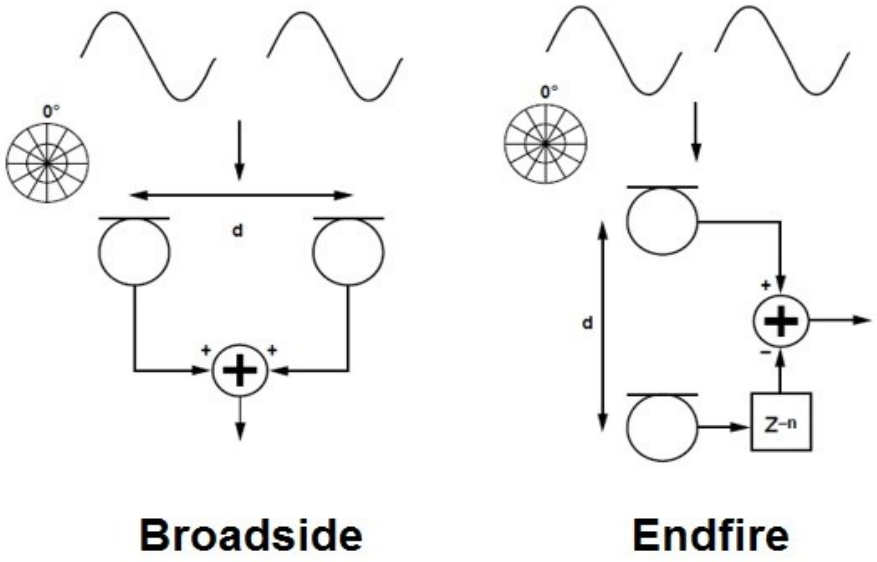


Рисунок 5 – Базовые структуры микрофонных решеток [44]

Данные структуры используют всенаправленные микрофоны (микрофоны, которые независимо от своей ориентации принимают сигнал с любых направлений). На Рисунке 6 показана зависимость приема сигнала от направления для различных частот одним всенаправленным микрофоном [44]. Для одного микрофона наблюдается инвариантность по частоте.

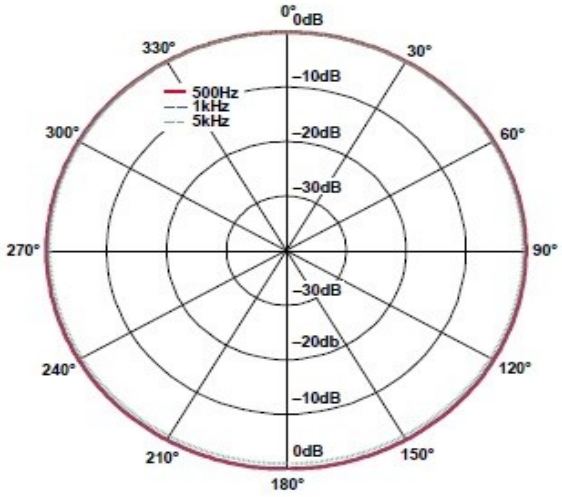


Рисунок 6 – Зависимость приема сигнала от направления одним всенаправленным микрофоном для частот 500 Гц, 1 и 5 кГц [44]

Структура Broadside представляет собой массив всенаправленных микрофонов, расположенный перпендикулярно направлению полезного сигнала.

Такие массивы обладают осью симметрии, относительно которой звук выделяется без ослабления как «спереди» массива, так и «сзади». Такие структуры получили широкое применение в приложениях, где волны звукового давления поступают на массив датчиков с одной стороны.

Рассмотрим структуру Broadside, состоящую из двух микрофонов, расположенных на расстоянии 7,5 см друг от друга. Минимальный отклик наблюдается при падении сигнала под углом 90° или 270° (за 0° в данном случае принимается угол между направлением полезного сигнала и нормали к линии элементов). Но данный отклик сильно зависит от частоты принимаемого сигнала. Теоретически у такой системы существует идеальный ноль на частоте 2,3 кГц. Выше данной частоты в зависимости от направления прихода имеются нули под другими углами (Рисунок 7).

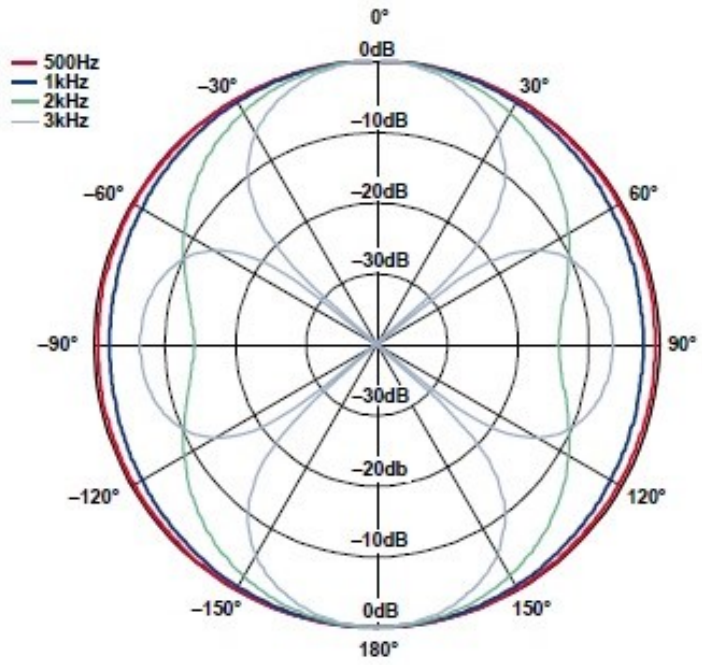


Рисунок 7 – Зависимость приема сигнала от направления структурой Broadside из двух всенаправленных микрофонов для частот 500 Гц, 1 кГц, 2 кГц и 3 кГц [44]

Структура Endfire состоит из нескольких микрофонов расположенных по направлению полезного акустического сигнала. Задержанный сигнал первого микрофона суммируется с сигналом следующего микрофона. Такие структуры используются для формирования кардиоидного, гиперкардиоидного или

суперкардиоидного отклика по направлению и теоретически полностью исключают звук, падающий на массив под углом 180°. Для формирования кардиоидного отклика по направлению сигнал от микрофонов должен задерживаться на время, равное распространению акустической волны между двумя элементами. У разработчиков таких систем есть две степени свободы для изменения выходного сигнала акустической системы: изменение расстояния между микрофонами и изменение времени задержки [44]. На Рисунке 8 показана зависимость приема сигнала от направления для различных частот структурой Endfire с двумя элементами и расстоянием между ними 2,1 см.

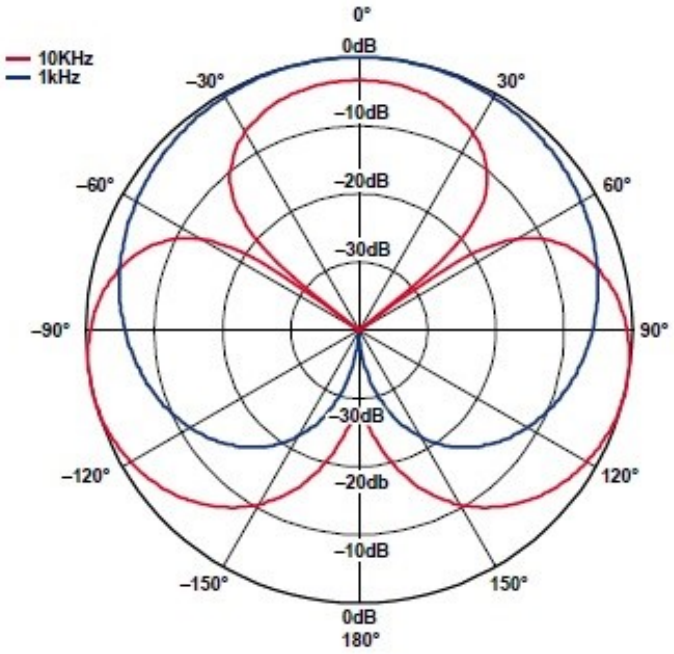


Рисунок 8 – Зависимость приема сигнала от направления структурой Endfire из двух всенаправленных микрофонов для частот 1 кГц и 10 кГц [44]

Согласно [7] рассмотренные структуры обладают следующими достоинствами и недостатками (Таблица 1):

Таблица 1 – Достоинства и недостатки Broadside и Endfire [7]

Структура МР	**Достоинства**	**Недостатки**
Broadside	Плоская геометрия Простая реализация обработки Возможность управления направлением луча	Меньшее подавление вне оси МР Малое расстояние между микрофонами и их большое число необходимо, чтобы предотвратить пространственную утечку
Endfire	Лучшее подавление вне оси Меньший общий размер	Неплоская (объемная) геометрия Более сложная обработка Подавление полезного сигнала в диапазоне низких частот Направление на источник полезного сигнала должно совпадать с осью МР Для двумерных решеток формирование луча возможно только в горизонтальном направлении (плоскости решетки)

1.3. Алгоритмы обработки сигналов микрофонных решеток

Микрофонные решетки применяются для решения широкого круга практических задач. Решаемые задачи различны, а, следовательно, различны и алгоритмы обработки сигналов. Выбор того или иного алгоритма обработки сигналов основан на сценариях акустической обстановки [7]:

- сценарии близкого/удаленного диктора
- сценарии стационарной/динамической акустической обстановки
- обработка в реальном времени/пост обработка

Обработка речевых сигналов микрофонной решеткой может происходить как во временной, так и в частотной области [28]. Во временной области сигналы на каждом микрофоне проходят через КИХ-фильтр и далее сигналы со всех микрофонов объединяются в один выходной сигнал системы. В частотной области широкополосный сигнал разбивается на узкополосные представления с помощью кратковременного преобразования Фурье, которые обрабатываются отдельно.

Рассмотрим алгоритмы обработки сигналов, получившие наибольшее распространение:

- алгоритм задержки и суммирования;
- алгоритмы фильтрации и суммирования;
- алгоритмы обработки сигналов в подрешетках;
- алгоритмы, минимизирующие мощность шума на выходе;
- алгоритмы, основанные на критерии минимума среднеквадратической ошибки;
- алгоритмы, основанные на критерии максимума отношения сигнал/шум.

Алгоритм задержки и суммирования

Простейшим алгоритмом формирования диаграммы направленности является алгоритм задержки и суммирования (Рисунок 9), известный как «delay-and-sum beamforming» [28, 43, 45].

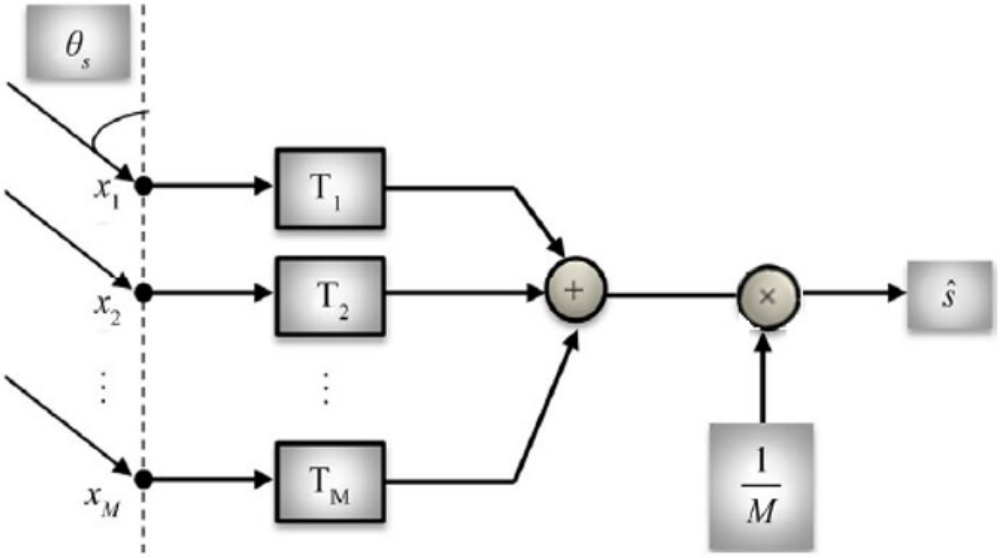


Рисунок 9 – Схема алгоритма задержки и суммирования [28]

На каждый микрофон (всего – M) вводится временная задержка T_m для дальнейшего когерентного суммирования сигналов. Такое суммирование позволяет усилить полезный сигнал и ослабить сигналы помех.

Речевой сигнал может быть разложен на узкополосные частотные составляющие, задержки могут быть аппроксимированы фазовыми сдвигами в каждой полосе частот.

Амплитудный множитель на выходе является элементом усреднения: он обратно пропорционален количеству микрофонов.

Во временной области выходной сигнал системы представляется как:

$$\hat{S}(t) = \frac{1}{M}\sum_{m=1}^{M} x_m(t - T_m). \quad (3)$$

Алгоритм фильтрации и суммирования

Наиболее общий класс алгоритмов. Применяется для реверберирующих сред. Алгоритм фильтрации и суммирования отличается тем, что и амплитуда и фаза являются частотно зависимыми параметрами [43, 45]. Сигнал на каждом микрофоне проходит через нерекурсивный фильтр и только потом подвергается суммированию с сигналами сторонних каналов. Общая схема алгоритма фильтрации и суммирования изображена на Рисунке 10, а выходной сигнал системы в частотной области представляет собой сумму частотно-зависимых компонент:

$$S_{out}(f) = \sum_{n=1}^{N} w_n(f)x_n(f). \quad (4)$$

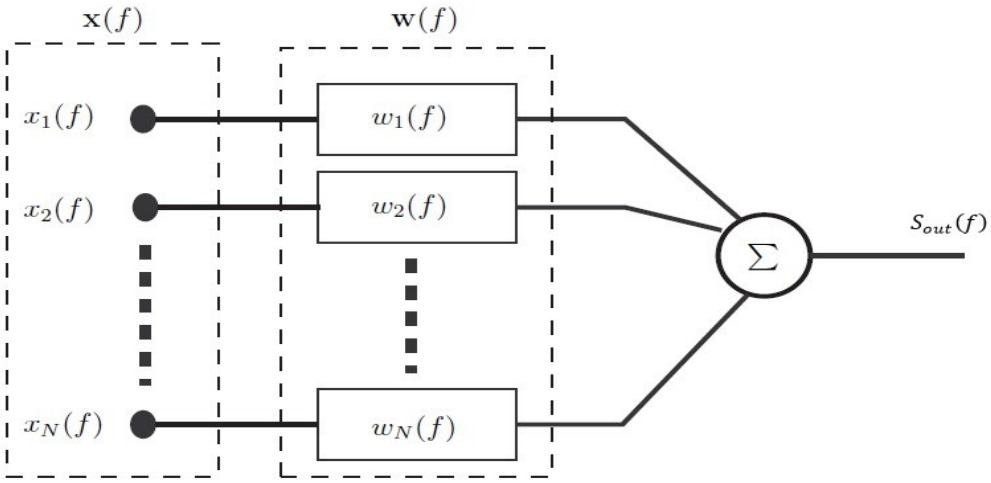


Рисунок 10 – Схема алгоритма фильтрации и суммирования [43]

Алгоритм обработки сигналов в подрешетках

Диаграмма направленности микрофонной решетки зависит от частоты принимаемого сигнала. При приеме широкополосных сигналов на разных частотах уровень боковых лепестков помех различен, также различна и ширина основного лепестка диаграммы направленности решетки. Чтобы обеспечить прием широкополосного сигнала создается инвариантность диаграммы направленности по частоте путем разбиения массива элементов на подрешетки. Каждая подрешетка представляет собой линейный эквидистантный массив приемников, принимающий сигнал в определенной полосе частот. Для того, чтобы уровень боковых лепестков оставался неизменным в каждой подрешетке используют фиксированное число элементов. Количество элементов микрофонной решетки может быть сокращено путем использования одного и того же приемника в разных подрешетках. Для извлечения выходного сигнала микрофонной решетки сначала формируются S выходных сигналов подрешеток различных диапазонов частот, которые подвергаются операции суммирования [43]:

$$S_{out}(f) = \sum_{s=1}^{S} \sum_{n=1}^{N} w_n(f) x_n(f). \quad (5)$$

В обзоре [43] приведен пример микрофонной решетки из девяти элементов, разделенных на четыре подрешетки и обрабатывающей широкополосный сигнал соответственно в четырех различных диапазонах частот (Рисунок 11).

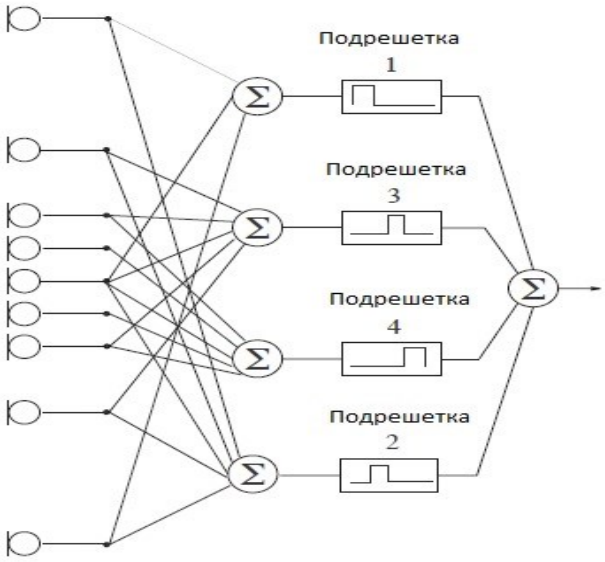


Рисунок 11 – Схема обработки широкополосного сигнала микрофонной решеткой из девяти элементов в четырех диапазонах частот [43]

В литературе [10] микрофонная решетка из двадцати девяти элементов также разделена на четыре подрешетки для работы в диапазоне частот от 500 Гц до 8 кГц (Рисунок 12). Такая техническая реализация позволяет поддерживать постоянной ширину основного лепестка диаграммы направленности на всем рассматриваемом частотном диапазоне.

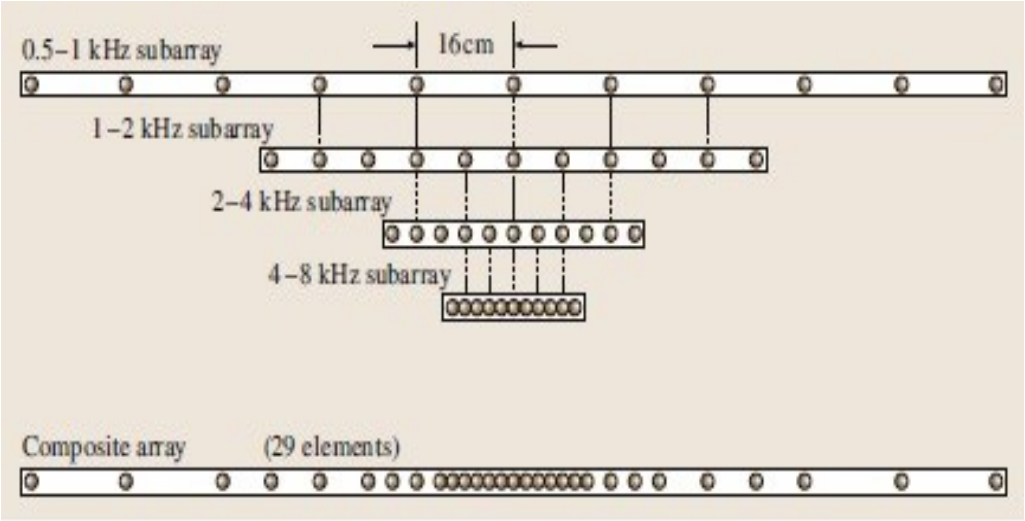


Рисунок 12 – Расположение микрофонов в двадцати девяти элементной решетке для поддержания постоянной ширины основного лепестка на всем частотном диапазоне [10]

Рассмотренные алгоритмы направлены на реализацию желаемого отклика системы в заданном направлении. Например, необходимо получить сигнал, поступающий с определенного направления, и в этом случае желаемый отклик будет принят за единицу в этом направлении. Или, например, доступна информация о действующей на определенной частоте помехе и о направлении с которого она приходит, тогда желаемый отклик на этой частоте и направлении равен нулю.

Дальнейшие алгоритмы, которые будут рассмотрены, основаны на статистических свойствах искомых и интерференционных сигналов. Данные алгоритмы оптимизируют некоторую функцию, за счет адаптивной фильтрации входящих сигналов, выделяя полезный сигнал и отклоняя помехи, приходящие с других направлений. Оптимизация заключается в применении различных критериев, таких как максимальное отношение сигнал/шум (MSNR), минимальная среднеквадратичная ошибка (MMSE), минимальная дисперсия шума (MVDR) и др.

Алгоритмы, минимизирующие мощность шума на выходе

Большинство адаптивных методов полагаются на минимизацию среднеквадратической ошибки между опорным сигналом и выходным сигналом. К сожалению, алгоритм наименьшего среднеквадратического отклонения (СКО) может ухудшить желаемый сигнал, поскольку он направлен на минимизацию среднеквадратической ошибки и не предъявляет требований к порогу искажений желаемого сигнала. Адаптивный алгоритм, который учитывает этот недостаток, называется алгоритмом Фроста [28] (Рисунок 13).

В данной схеме коэффициенты фильтра адаптируются через использование алгоритма наименьшего СКО. Алгоритм используется для минимизации мощности шума на выходе при сохранении постоянного коэффициента усиления и линейной фазы на отклик системы в направлении на полезный источник.

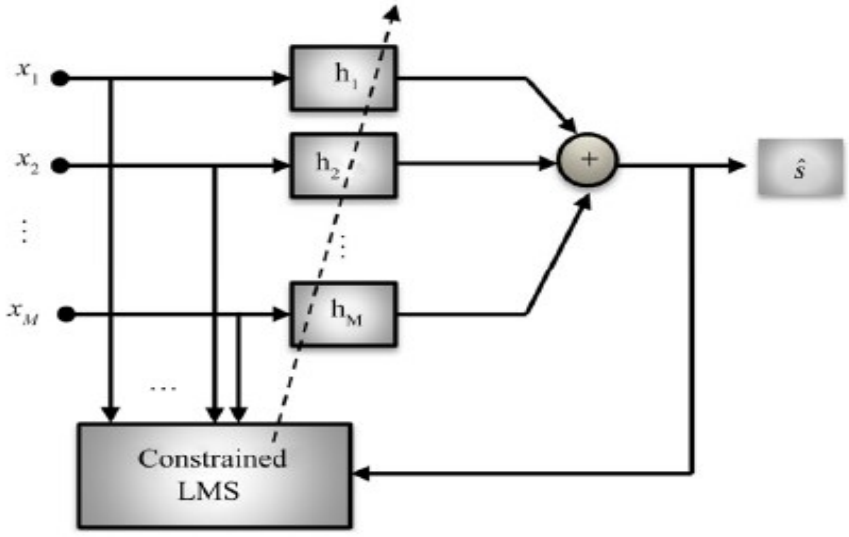


Рисунок 13 – Схема алгоритма Фроста [28]

К данному классу также можно отнести алгоритм подавления боковых лепестков (Рисунок 14).

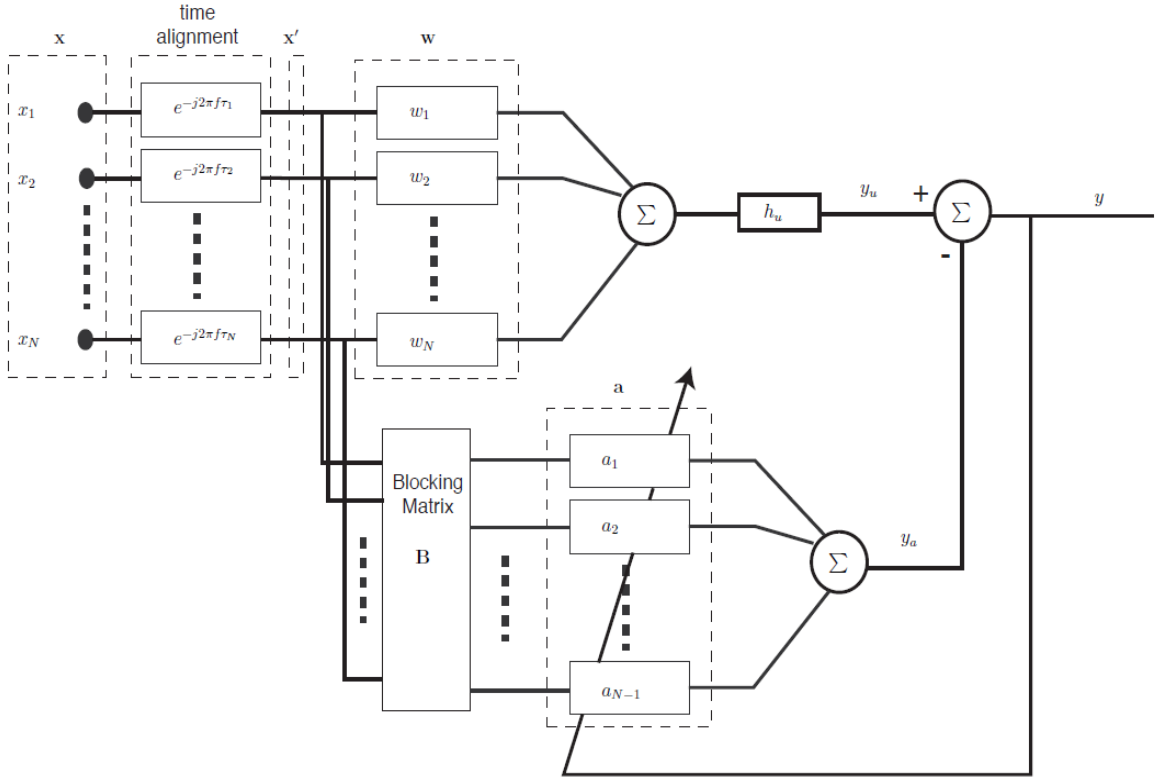


Рисунок 14 – Схема алгоритма подавления боковых лепестков [43]

Схему алгоритма можно разделить на верхний и нижний тракт. Верхний предназначен для формирования оценки полезного сигнала. Сигналы выравниваются по времени и с помощью формирования диаграммы направленности тракт настраивается на полезный сигнал. Нижний тракт обеспечивает исключение полезного сигнала за счет наличия блокирующей матрицы $\boldsymbol{B}$ [46]. После блокирующей матрицы сигналы нижнего тракта проходят фильтр **a,** адаптивно минимизирующий мощность шума на выходе. Выходной сигнал системы представляет собой разностный сигнал верхнего и нижнего тракта: из оценки полезного сигнала вычитается общий сигнал интерференции.

В работе [47] описан алгоритм, минимизирующий мощность шума на выходе для двухэлементной дифференциальной микрофонной решетки, адаптивно формирующий нули в направлении источников когерентного шума.

Алгоритмы, основанные на критерии минимума среднеквадратической ошибки

Данные алгоритмы используются при наличии достаточной информации о полезном сигнале. Если данное условие выполняется, то имеется возможность сформировать опорный сигнал q(t). Сигнал ошибки f(t) отражает различие между желательной реакцией микрофонной решетки и выходным сигналом решетки [48]. Алгоритмы, основанные на минимизации среднеквадратической ошибки стремятся исключить данное различие.

$$f(t) = q(t) - \boldsymbol{W^H}\big(\boldsymbol{S} + \boldsymbol{X}(t)\big), \tag{6}$$

где $\boldsymbol{W}$ – весовой вектор, $\boldsymbol{S}$ – вектор полезного сигнала**,** $\boldsymbol{X}$ – вектор интерференции.

На Рисунке 15 показана схема реализации алгоритма минимизации СКО.

Минимальное среднеквадратическое отклонение при такой постановке задачи:

$$\langle|f|^2\rangle = \langle|q|^2\rangle - \boldsymbol{R}^H\boldsymbol{M}^{-1}\boldsymbol{R}, \tag{7}$$

а выражение для оптимальных весовых коэффициентов по критерию минимизации среднеквадратической ошибки можно найти из условия:

$$\boldsymbol{W} = \boldsymbol{M}^{-1}\boldsymbol{R}, \tag{8}$$

где $\boldsymbol{R}$ – корреляционный вектор:

$$\boldsymbol{R} = \langle(\boldsymbol{S} + \boldsymbol{X}(t))q(t)\rangle. \tag{9}$$

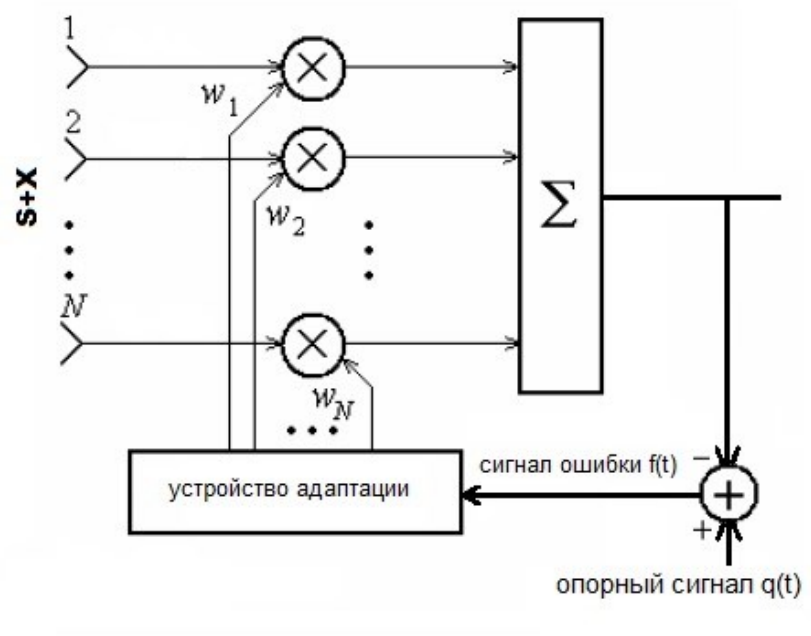


Рисунок 15 – Схема алгоритма минимизации СКО

По данному критерию работают алгоритмы постфильтрации [28, 43]: алгоритм фильтрации и суммирования с добавлением фильтра к выходу системы (Рисунок 16).

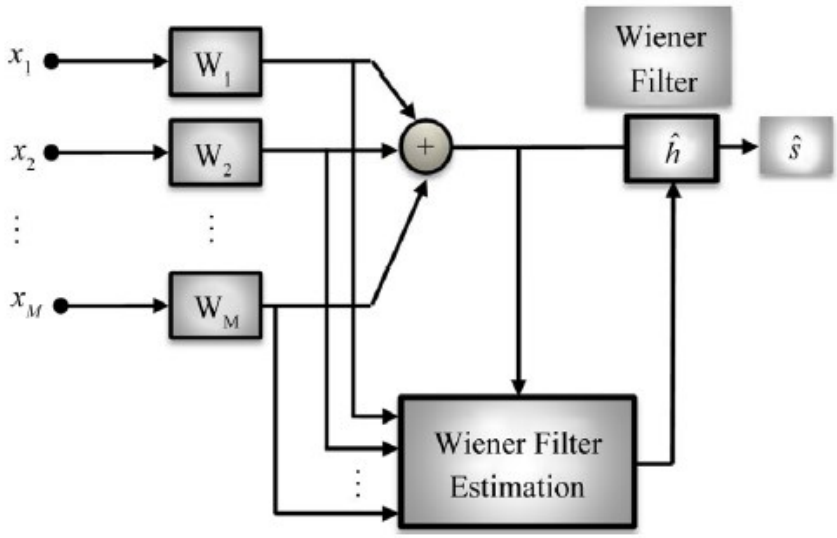


Рисунок 16 – Схема алгоритма постфильтрации Винера [28]

Алгоритмы, основанные на критерии максимума отношения сигнал/шум

За отношение сигнал/шум на выходе решетки принимается отношение средней мощности полезного сигнала к средней мощности помехи (собственного шума приемников и внешних источников помех).

Отношение сигнал/шум на выходе микрофонной решетки можно записать в матричной форме [48]:

$$\Lambda = \frac{\boldsymbol{W}^H \boldsymbol{M}_s \boldsymbol{W}}{\boldsymbol{W}^H \boldsymbol{M} \boldsymbol{W}}, \quad (10)$$

где $\boldsymbol{M}$ – корреляционная матрица помехи, $\boldsymbol{M_s}$ – корреляционная матрица полезного сигнала.

Оптимальный весовой вектор, обеспечивающий максимизацию данного отношения можно найти из условия:

$$\boldsymbol{W} = \boldsymbol{M}^{-1}\boldsymbol{S}. \quad (11)$$

При этом максимальное значение отношения сигнал/шум равно:

$$\Lambda_{max} = \varsigma \boldsymbol{S}^H \boldsymbol{M}^{-1} \boldsymbol{S}, \quad (12)$$

где ς – отношение мощности полезного сигнала к мощности собственного шума в отдельном элементе решетки.

1.4. Геометрия современных микрофонных решеток

Для решения конкретных задач с помощью микрофонной решетки важно учитывать ее геометрию [49-51]. При этом учитываются различные критерии оптимизации: постоянство ширины основного лепестка диаграммы направленности в широком диапазоне частот, сокращение общего количества микрофонов в решетке, уменьшение уровня боковых лепестков, максимизация коэффициента направленного действия и др. Например, для решения задач по локализации акустического источника необходимо обладать априорной информацией о геометрии решетки [52], и, напротив, при использовании алгоритмов минимизации дисперсии шума, геометрия решетки практически не важна [43]. Упрощают задачу оценки параметров сигналов микрофонные решетки

с эквидистантным размещением микрофонов: линейные и круговые микрофонные решетки [53].

В работе [54] рассматриваются конфигурации микрофонных решеток из 128 микрофонов, которые могут обеспечить преимущество в извлечении картины акустического поля по сравнению с обычными круговыми решетками (Рисунок 17).

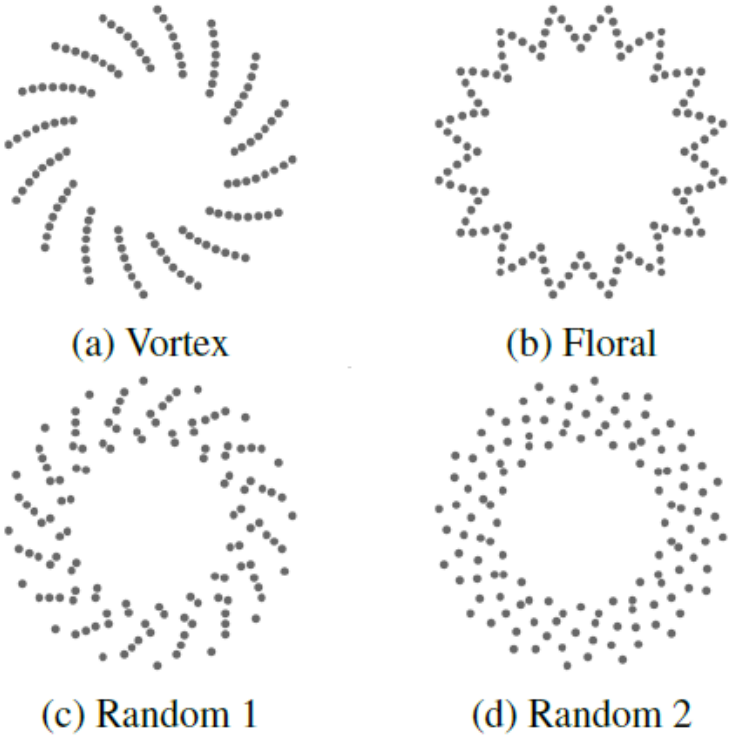


Рисунок 17 – конфигурации микрофонных массивов из 128 микрофонов: a) вихревая, b) «цветок», c) произвольная 1, d) произвольная 2 [54]

В обзоре [7] приведены примеры решеток с неравномерных расположением микрофонов – решетки компаний SquareHead и Norsonic. Микрофоны размещены на концентрических окружностях. Чем больше апертура системы, тем большее число микрофонов размещено в решетке.

Широкое распространение получили не только линейные и круговые массивы, но и трехмерные сферические решетки микрофонов, которые способны более детально извлекать информацию об акустическом поле [55-57].

По мере удаления источника звука от микрофонного массива точность его локализации будет спадать за счет затухания звуковой волны. Эффективность

локализации источника, а также эффективность выделения речи целевого диктора, может быть повышена за счет использования одновременно нескольких микрофонных решеток, распределенных в пространстве. В литературе такие распределенные системы получили название «сеть акустических датчиков (acoustic sensor network)» [58-59]. Эффективность использования таких систем связана с тем, что значительно увеличивается вероятность того, что один из микрофонов (массивов микрофонов) будет находиться ближе к полезному источнику. Наибольшее распространение такие системы получили при использовании на больших пространствах, когда требуется извлечь информацию из любой точки пространства наблюдения. Решаемые такими системами задачи различны: локализация источника, выделение голоса определенного человека, разделение акустических источников, определение траектории перемещений дикторов, определение направленности акустического источника и др.

Так, в исследовании [52] показано, что можно получить надежную систему локализации звука, использующую несколько микрофонных массивов. Экспериментальные результаты с использованием этого подхода показали уменьшение средней ошибки локализации источника до 8 см. Результаты получены для десяти двух-элементных решеток при 0 дБ. Геометрия массива приведена на Рисунке 18.

В работе [59] проведен анализ производительности двухступенчатого процесса локализации в двумерном пространстве на основе измерений разницы времени прихода сигналов на микрофон. Оптимальное решение получается при обработке данных, полученных с распределенных в пространстве микрофонов, центральным узлом. Исследования проведены для массивов из 24 микрофонов, разделенных по 6 микрофонов на 4 микрофонные решетки, располагаемых в одной плоскости на окружности в 4 разных секторах. Источник находился в центре окружности. Показано, что оптимальная точность локализации достигается, когда микрофоны равномерно распределены вокруг источника.

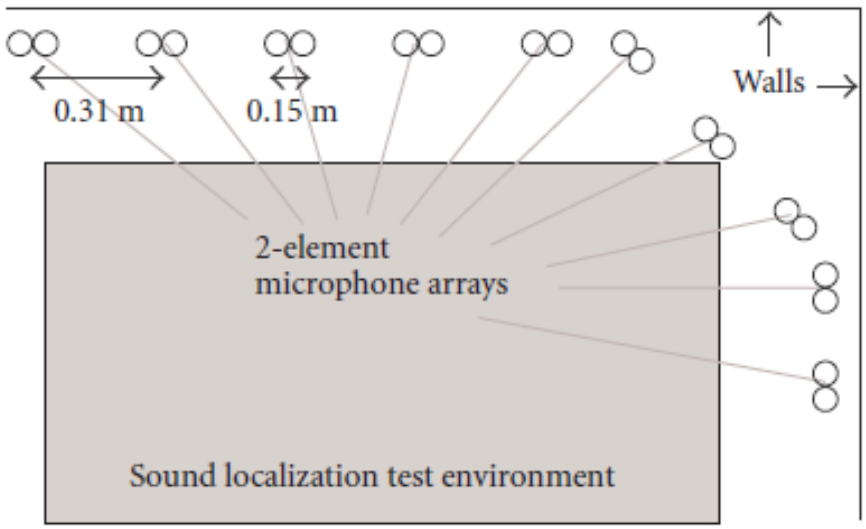


Рисунок 18 – Размещение десяти двух-элементных массивов для локализации источника [52]

В работе [60] представлена оригинальная методология локализации источника звука с использованием микрофонных массивов, случайно распределенных в пространстве. Для построения картины звукового поля используется пространственное преобразование лучей, в котором акустические источники выглядят как линейные структуры, которые возможно обнаружить с помощью методов анализа изображений. Эффективность предложенного решения подтверждается исследованиями в реверберирующих средах.

Многоканальный фильтр Винера [61-65] используется в акустических приложениях для улучшения качества речевого сообщения за счет уменьшения энергии интерференции. Шумоподавление основано на оценке требуемой составляющей сигнала в одном из микрофонов, определяемого эталонным. В работе [64] в помещении объёмом 7,5x5x3,5 $м^3$ и временем реверберации в 400 мс было исследовано влияние выбора эталонного микрофона на производительность алгоритма для разных положений одного источника речи относительно распределенной решетки из 6 микрофонов (по 3 микрофона на двух смежных стенах). В исследовании [65] для той же конфигурации микрофонного массива исследован метод взвешенного выбора эталонного микрофона. В отличие от

произвольного выбора, новый метод показывает выигрыш в выходном отношении сигнал/шум.

Для одиночной микрофонной решетки предполагается, что речевая активность полезного сигнала и сигналов помех является глобальной, то есть общей для всех микрофонов близко расположенных друг к другу [66]. Однако это предположение может быть нарушено для распределенных микрофонных решеток, когда речевая активность на одном микрофонном массиве может значительно отличаться от активности на других массивах, за счет пространственного положения. Авторы работы [66] обращают внимание на то, что распределенные микрофонные решетки для определенных сценариев могут быть заменены на одиночные микрофоны, распределенные в пространстве. В исследовании имитируется разделение речи каждого человека в двух отдельных группах с участием трех и двух человек соответственно (Рисунок 19). Микрофоны располагались на окружностях радиуса 10 см и затем были удалены от центра на 50 см.

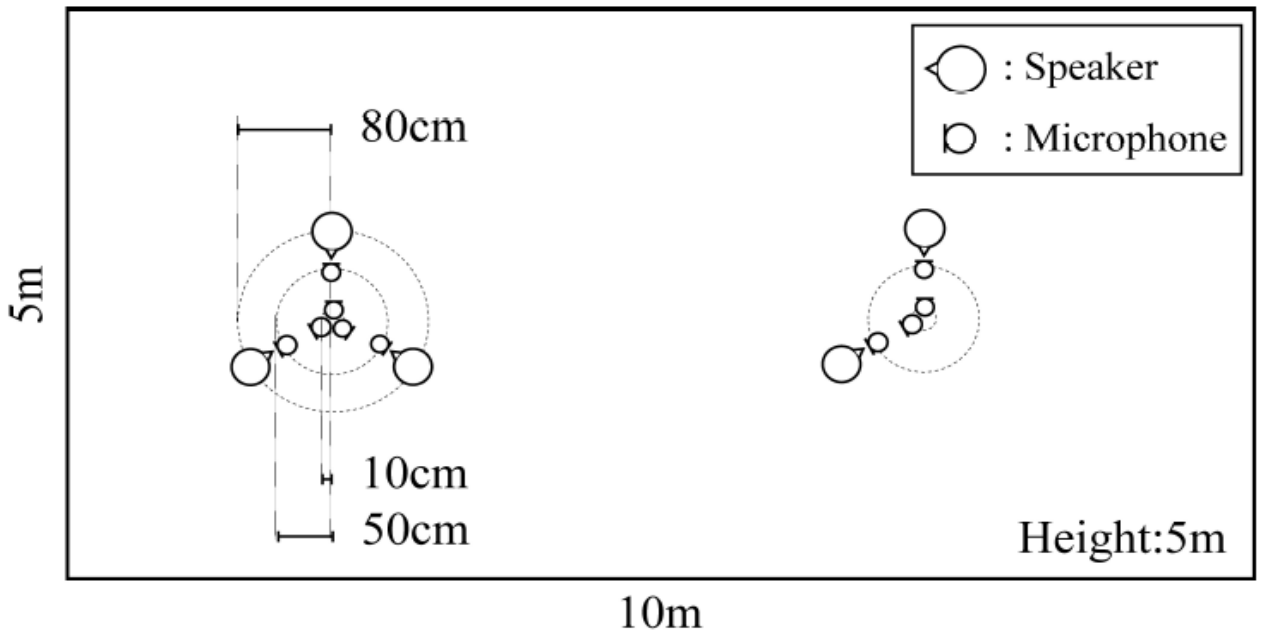


Рисунок 19 – Взаимное расположение одновременно говорящих людей и сети из 5 распределенных акустических датчиков [66]

С помощью нескольких массивов микрофонов предложено решение задачи для отслеживания перемещений нескольких дикторов [67]. Описан подход для одновременного определения координат движения нескольких дикторов. Апробация результатов моделирования движущихся источников проводилась в реверберирующем конференц-зале объемом 3,7x6,8x2,6 м^3. Сигналы от круглых микрофонных решеток с 5 микрофонами и радиусом 5 см, регистрировались с частотой 48 кГц. Доказана способность точно отслеживать статические и движущиеся источники. При использовании не менее трех массивов микрофонов, распределенных в пространстве, получена точность, допустимая для практического применения системы.

Современные исследования связаны с применением распределенных микрофонных систем с большой апертурой. В работе [68] описывается реализация трехмерного массива из 256 микрофонов, которые расположены на стенах и потолке исследуемого пространства для оценки диаграммы направленности источников звука в реверберационных средах (Рисунок 20).

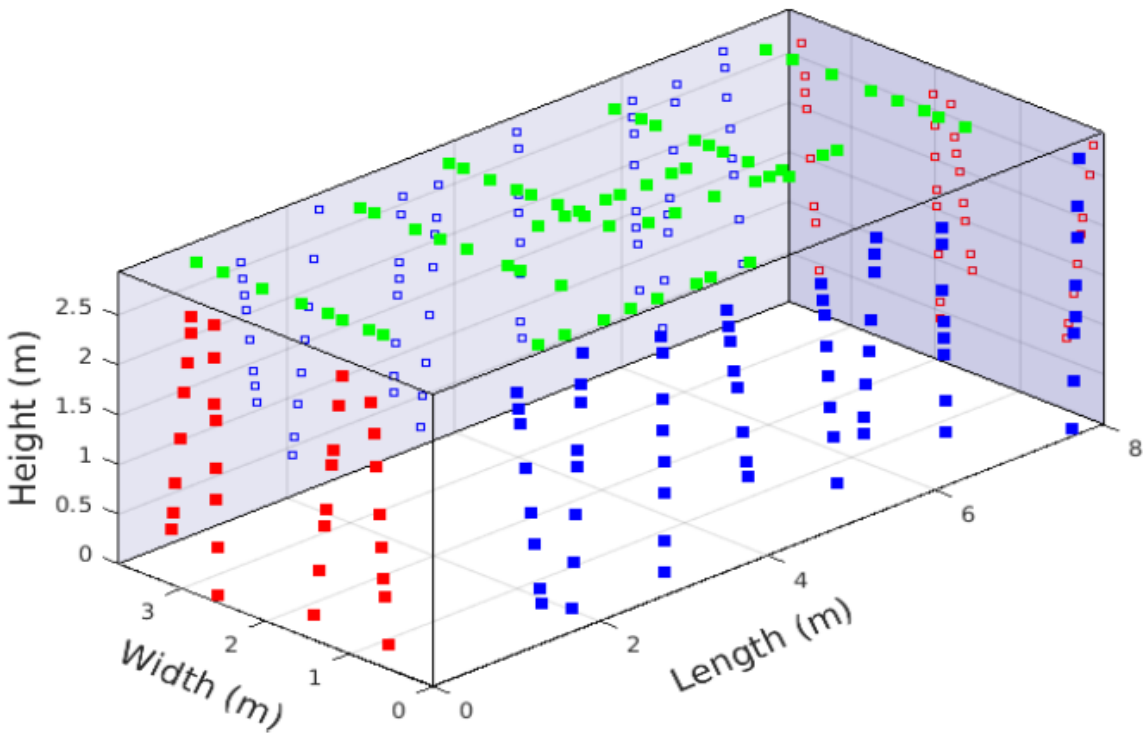


Рисунок 20 – Расположение массива из 252 микрофонов в исследуемом пространстве наблюдения [68]

В данном параграфе показаны различные конфигурации микрофонных решеток, используемые в настоящее время. Геометрия микрофонных решеток крайне разнообразна из-за различных задач, стоящих перед исследователями. Поэтому в рамках данной работы нет возможности перечислить все возможные конфигурации, используемые в настоящих исследованиях.

Несмотря на достоинства каждого из рассмотренных подходов (T-F masking, ICA, Beamforming, acoustic sensor network), применение только одного из них не дает безупречных результатов, что объясняется накладываемыми ограничениями. Методы пространственной фильтрации не всегда дают хорошие результаты в недоопределенных случаях (количество помех больше количества микрофонов). Методы анализа независимых компонент выполняются в предположении, что исходные сигналы статистически независимы и не требуют наличия информации о конфигурации микрофонной решетки или направления прихода исходных сигналов [28]. Но применение таких методов достаточно ограничено в реальных реверберирующих средах. Когда количество источников больше количества микрофонов линейное разделение источников методом обращения матрицы смешивания невозможно. Но в недоопределенных случаях хорошие результаты демонстрируют методы частотно-временной маскировки. Данный класс основан на предположении об уникальности спектра – каждая точка частотно-временного представления соответствует одному сигналу. Но сигналы, разделенные этим методом, как правило, имеют нелинейные искажения.

В современной литературе появляются публикации, основанные на комбинированном сочетании данных подходов. Так в работе [69] описан алгоритм, совмещающий методы частотно-временной маскировки и алгоритм многоканальной фильтрации Винера. Или в [70] показан способ выделения речи из голосовой смеси способом, объединяющим пространственную фильтрацию и слепое разделение сигналов.

ВЫВОДЫ ПО ПЕРВОЙ ГЛАВЕ

В данной главе рассмотрены основные подходы для разделения речи:

-методы частотно-временной маскировки или T-F masking

-анализ независимых компонент или ICA

-методы формирования диаграммы направленности или Beamforming

-методы разделения речи с использованием распределенных в пространстве микрофонных массивов.

В настоящее время для разделения сигналов речевых источников разработано большое число алгоритмов, что обусловлено высокой степенью сложности данной задачи. Сложность разделения зависит от количества источников, количества микрофонов и их расположения, уровня шума, способа смешивания сигналов, а также от предварительной информации об источниках, микрофонах и параметрах смешивания.

В основном, получили распространение алгоритмы, реализованные в частотной области. Они активно развиваются в направлении комплексирования различных методов обработки, но, по-прежнему, не дают безупречных результатов во всем многообразии внешних условий и ограничений.

Таким образом, актуальна реализация и исследование свойств алгоритма оптимальной пространственной фильтрации речевых сообщений на фоне пространственно-распределенных источников помех во временной области, с использованием полной полосы частот сообщения без разбиения, в реальном масштабе времени, позволяющего выделять речевые сообщения из любой точки пространства наблюдения с максимальным отношением сигнал/помеха независимо от взаимного расположения целевого диктора и других дикторов, являющихся источниками речевых помех.

ГЛАВА 2. ПРОСТРАНСТВЕННАЯ ОБРАБОТКА РЕЧЕВЫХ СИГНАЛОВ ВО ВРЕМЕННОЙ ОБЛАСТИ

В данной главе предложен алгоритм обработки речевых сигналов микрофонным массивом во временной области, основанный на первоначальном введении временных задержек, зависящих от пространственных координат, и дальнейшем увеличении отношения сигнал/помеха на выходе массива за счет расчета оптимальных на интервалах стационарности весовых коэффициентов микрофонов. Основные результаты второй главы опубликованы в работах автора [A1-A2, A5-A8, A11-A12].

2.1. Метод пространственной фильтрации речевых сигналов на фоне распределенных помех

Измерительная система представляет собой N разнесенных в пространстве микрофонов. Каждый микрофон регистрирует сумму всех k акустических сигналов, один из которых определен как полезный речевой сигнал определенного диктора $S(t)$, остальные – помехи – речевые сигналы других дикторов $G_f(t)$, f=1,…,k-1. При разных расстояниях r_i от источника до приемного устройства с номером i, i=1,...,N сигналу требуется разное время, чтобы достичь микрофона. Микрофонные решетки работают с цифровым сигналом, поэтому v-й отсчет сигнала, регистрируемого i-м микрофонов, можно представить следующим образом (V_s – скорость звука, F_s – частота дискретизации):

$$Q_i(t_v) = \frac{A}{r_i} S\left(t_v - \frac{r_i}{V_s} F_s\right) + \sum_{f=1}^{k-1} \frac{A_{fi}}{r_{fi}} G_f(t_v - \frac{r_{fi}}{V_s} F_s), \quad (13)$$

где A и A_{fi} – масштабные множители, r_i – расстояние от источника полезного сигнала до i-ого приемника, r_{fi} – расстояние от f-ого источника помехи до i-ого приемника.

Для выделения полезного речевого сигнала из смеси речевых сигналов сторонних дикторов в каждый регистрируемый *i*-м микрофоном сигнал вводятся временные задержки, зависящие от пространственных координат:

$$\tau_{ji} = \tau_{opt} - \frac{r_{ji}}{V_s}, \tag{14}$$

где r_{ji} – расстояние от произвольно-выбранной точки пространства наблюдения j=1,...,p (точки фокусировки системы N микрофонов) до i-ого микрофона.

Временная оптимальная задержка τ_{opt} для всех микрофонов постоянна и равна отношению наибольшего пути распространения звука в исследуемом пространстве без учета отражений (для прямоугольного периметра акустической сцены – диагональ) к скорости звука:

$$\tau_{opt} = \frac{\sqrt{a^2 + b^2}}{V_s}, \tag{15}$$

где a и b соответственно – линейные размеры прямоугольного периметра.

Третьим линейным размером можно пренебречь только в том случае, если источники сигналов и микрофоны расположены в одной плоскости.

Временная задержка τ_{ji} для каждой точки пространства наблюдения и для каждого микрофона индивидуальна. Знание координат размещения N микрофонов позволяет вычислить матрицу временных задержек T для каждой точки исследуемого пространства наблюдения:

$$T = \begin{pmatrix} \tau_{11}\,\tau_{21}\,\tau_{31}\ldots\tau_{p1} \\ \tau_{12}\,\tau_{22}\,\tau_{32}\ldots\tau_{p2} \\ \tau_{13}\,\tau_{23}\,\tau_{33}\ldots\tau_{p3} \\ \ldots\ \ \ldots\ \ \ldots\ \ \ldots\ \ \ldots \\ \tau_{1N}\tau_{2N}\tau_{3N}\ldots\tau_{pN} \end{pmatrix}, \tag{16}$$

где p – общее количество точек пространства наблюдения.

После введения задержек сигнал i-ого микрофона можно записать:

$$\begin{aligned} Q_i(t_v) = {} & \frac{A}{r_i} S\left(t_v - \left(\frac{r_i}{V_s} + \tau_{opt} - \frac{r_{ji}}{V_s}\right) F_s\right) + \\ & + \sum_{f=1}^{k-1} \frac{A_{fi}}{r_{fi}} G_f\left(t_v - \left(\frac{r_{fi}}{V_s} + \tau_{opt} - \frac{r_{ji}}{V_s}\right) F_s\right). \end{aligned} \tag{17}$$

Введение временных задержек τ_{ji} приводит к синхронному приему речевого сигнала всеми микрофонами. При совпадении координат источника с координатами точки фокусировки системы:

$$\frac{r_i}{V_s} = \frac{r_{ji}}{V_s}, \tag{18}$$

а сигнал i-ого микрофона можно переписать в виде:

$$Q_i(t_v) = \frac{A}{r_i} S\left(t_v - \tau_{opt} F_s\right) + \sum_{f=1}^{k-1} \frac{A_{fi}}{r_{fi}} G_f(t_v - (\tau_{opt} + \frac{r_{fi}}{V_s} - \frac{r_{ji}}{V_s}) F_s). \tag{19}$$

Выходной сигнал многопозиционной системы из N микрофонов для определенной точки пространства наблюдения представляет собой суперпозицию сигналов всех приемников:

$$\begin{aligned} Q_{ex_j}(t_v) = \sum_{i=1}^{N} Q_i(t_v) = S\left(t_v - \tau_{opt} F_s\right) \sum_{i=1}^{N} \frac{A}{r_i} + \\ + \sum_{i=1}^{N} \sum_{f=1}^{k-1} \frac{A_{fi}}{r_{fi}} G_f \left(t_v - (\tau_{opt} + \frac{r_{fi}}{V_s} - \frac{r_{ji}}{V_s}) F_s\right). \end{aligned} \tag{20}$$

При такой пространственной фильтрации мощность полезного речевого сигнала возрастает в N^2 раз за счет синхронного приема, а мощность речевых помех за счет взаимного наложения – в N раз. Таким образом, отношение сигнал/помеха увеличивается в N раз.

Такой подход можно использовать для параллельной схемы акустического наблюдения за большим числом сторонних источников, расположенных на множестве p точек пространства наблюдения.

Количество точек пространства наблюдения p целесообразно выбирать из условия:

$$p = \frac{D}{\Delta x^2}, \tag{21}$$

где D – площадь исследуемого пространства наблюдения, Δx – разрешающая способность многопозиционной системы.

Разрешающую способность Δx многопозиционной акустической системы можно определить как диаметр поперечного сечения на уровне половины максимального значения пространственной автокорреляционной функции $R_j(x', y')$ системы сигналов от N микрофонов для каждой точки пространства наблюдения:

$$R_j(x', y') = \sum_{r=1}^{M} Q_{ex_j}^2(x', y', x_0, y_0), \tag{22}$$

где (x', y') – координаты точки фокусировки системы, (x_0, y_0) – фиксированные координаты точечного источника речевого сигнала, M – количество отсчетов дискретизированного по времени речевого сигнала.

За счет когерентного суммирования мощность полезного сигнала возрастает, следовательно, в точках пространства наблюдения с координатами акустического источника будут находиться максимумы пространственной автокорреляционной функции. Таким образом, расчет пространственной автокорреляционной функции позволяет определить координаты источников акустических сигналов.

2.2. Применение алгоритмов пространственной обработки сигналов для увеличения отношения сигнал/помеха

Метод введения пространственно-зависимых временных задержек дает хорошие результаты по выделению сигнала из помех, но при большом количестве сторонних источников отношение сигнал/помеха выделяемого речевого сообщения резко снижается. Повышение качества выделяемого речевого сообщения становится возможным, если использовать алгоритмы пространственной обработки сигналов, основанные на использовании оценки корреляционной матрицы помехи [71, 72]. Задача сводится к нахождению вектора весовых коэффициентов микрофонов, максимизирующего отношение сигнал/помеха речевого сообщения после выделения сигнала с помощью алгоритма введения задержек.

Нахождение весового вектора при оптимальной обработке речевого сигнала

Алгоритмы оптимальной обработки применяются, если [71]:

- точно известна корреляционная матрица помехи (известна форма сигналов помех и известны пространственные координаты акустических источников);
- имеется априорная информация о полезном источнике.

Используя предположение о некоррелированности голосов k-1 дикторов корреляционную матрицу помехи можно записать как:

$$\boldsymbol{M} = \boldsymbol{M}_n + \sum_{f=1}^{k-1} \boldsymbol{M}_f, \tag{23}$$

где $\boldsymbol{M}_n$ – корреляционная матрица собственного шума, $\boldsymbol{M}_f$ – корреляционная матрица f-й помехи:

$$\boldsymbol{M_f} = \begin{pmatrix} \left(\sum \frac{A_{f1}}{r_{f1}} G_f\,(t_v - \Delta_1)\right)^2 & \cdots & \sum \frac{A_{fN}}{r_{fN}} G_f(t_v - \Delta_N) \sum \frac{A_{f1}}{r_{f1}} G_f(t_v - \Delta_1) \\ \sum \frac{A_{f1}}{r_{f1}} G_f(t_v - \Delta_1) \sum \frac{A_{f2}}{r_{f2}} G_f(t_v - \Delta_2) & \cdots & \sum \frac{A_{fN}}{r_{fN}} G_f(t_v - \Delta_N) \sum \frac{A_{f2}}{r_{f2}} G_f(t_v - \Delta_2) \\ \cdots & \cdots & \cdots \\ \sum \frac{A_{f1}}{r_{f1}} G_f(t_v - \Delta_1) \sum \frac{A_{fN}}{r_{fN}} G_f(t_v - \Delta_N) & \cdots & \left(\sum \frac{A_{fN}}{r_{fN}} G_f\,(t_v - \Delta_N)\right)^2 \end{pmatrix}, \tag{24}$$

где суммирование ведется по временным отсчетам интервала наблюдения $v = 1,..,L$, Δ_i – задержка:

$$\Delta_i = (\tau_{opt} + \frac{r_{fi}}{V_s} - \frac{r_{ji}}{V_s}) F_s. \tag{25}$$

Алгоритм введения задержек за счет синхронного приема полезного сигнала формирует вектор-столбец $\boldsymbol{S}$:

$$\boldsymbol{S} = \begin{pmatrix} \frac{A}{r_1} \sum S\left(t_v - \tau_{opt} F_s\right) \\ \frac{A}{r_2} \sum S\left(t_v - \tau_{opt} F_s\right) \\ \cdots \\ \frac{A}{r_N} \sum S\left(t_v - \tau_{opt} F_s\right) \end{pmatrix}. \tag{26}$$

Оптимальный весовой вектор, максимизирующий отношение сигнал/помеха, с точностью до постоянного множителя можно найти из условия [71]:

$$\boldsymbol{W} = \alpha \boldsymbol{M}^{-1} \boldsymbol{S}, \tag{27}$$

где α – произвольный множитель, $\boldsymbol{M}^{-1}$ – обратная к корреляционной матрица помехи, $\boldsymbol{S}$ – вектор-столбец полезного сигнала.

Нахождение весового вектора при адаптивной обработке

Алгоритмы адаптивной обработки применяются при отсутствии априорной информации о входных сигналах [71]. В таком случае для нахождения весового вектора необходимо использовать оценку корреляционной матрицы помехи. Оценочная корреляционная матрица отличается тем, что в ней содержится информация не только о помехах, но и о полезном сигнале. Оценка корреляционной матрицы по Z статистически независимым выборкам входного процесса R можно выразить как:

$$\widehat{\boldsymbol{M}} = \frac{1}{Z}\sum_{z=1}^{Z} \boldsymbol{R}(\boldsymbol{z})\boldsymbol{R}^{\boldsymbol{H}}(\boldsymbol{z}). \tag{28}$$

При большом числе статистически независимых выборок по отношению к числу микрофонов N:

$$\widehat{\boldsymbol{M}} \overset{Z \gg N}{\longrightarrow} \boldsymbol{M}. \tag{29}$$

Алгоритм введения задержек позволяет выделить речевые сообщения из любой точки пространства наблюдения. Расчет автокорреляционной функции выходных сигналов акустической системы позволяет определить координаты всех акустических источников. Поэтому для нахождения оценки корреляционной матрицы помехи могут быть использованы сигналы мнимых источников – выделенные с помощью алгоритма введения задержек речевые сообщения G'_f – содержащие смесь полезного сигнала и самих помех.

Оценку корреляционной матрицы помехи найдем из соотношения:

$$\boldsymbol{M} = \boldsymbol{M}_n + \sum_{f=1}^{k-1} \boldsymbol{M}_f, \tag{30}$$

где $\boldsymbol{M_f}$ – оценка корреляционной матрицы f-ой помехи:

$$\boldsymbol{M_f} = \begin{pmatrix} \left(\sum \frac{A_{f1}}{r_{f1}} {G_f}'(t_v - \Delta_1)\right)^2 & \cdots & \sum \frac{A_{fN}}{r_{fN}} {G_f}'(t_v - \Delta_N) \sum \frac{A_{f1}}{r_{f1}} {G_f}'(t_v - \Delta_1) \\ \cdots & \cdots & \cdots \\ \sum \frac{A_{f1}}{r_{f1}} {G_f}'(t_v - \Delta_1) \sum \frac{A_{fN}}{r_{fN}} {G_f}'(t_v - \Delta_N) & \cdots & \left(\sum \frac{A_{fN}}{r_{fN}} {G_f}'(t_v - \Delta_N)\right)^2 \end{pmatrix}. \quad (31)$$

В задачах по выделению голоса из смеси голосов за полезный сигнал может быть условно назначен любой выделенный с помощью алгоритма задержек и суммирования голос G'_s. Это связано с тем, что сигнал полезного источника за счет синхронного приема преобладает над помехами.

$$\boldsymbol{S} = \begin{pmatrix} \frac{A_s}{r_{s1}} \sum {G'}_s(t_v - \tau_{opt} F_s) \\ \frac{A_s}{r_{s2}} \sum {G'}_s(t_v - \tau_{opt} F_s) \\ \cdots \\ \frac{A_s}{r_{sN}} \sum {G'}_s(t_v - \tau_{opt} F_s) \end{pmatrix}. \quad (32)$$

Весовой вектор, рассчитываемый по формуле (27) представляет собой набор весовых коэффициентов каждого из N приемных каналов, зависящих от времени $w_i=w_i(t)$. Данный набор индивидуален для каждой j-й точки пространства наблюдения:

$$\boldsymbol{W}_j = (w_1, w_2, \ldots, w_N)^T. \quad (33)$$

Для весового вектора применяется условие нормировки:

$$\sum_i w_i^2 = 1. \quad (34)$$

В каждый момент времени помеховая обстановка изменяется, поэтому при решении задач по выделению голоса предлагается разбивать речевые сообщения на интервалы анализа, в рамках которых будут сделаны оптимальные оценки параметров алгоритма обработки. В соответствии с ГОСТ Р 51061-97 [73] средняя длительность контрольной фразы равна 2,4 с. Выходной сигнал акустической системы на таких стационарных интервалах можно представить как:

$$Q_{ex_j}(t_v) = \sum_{i=1}^{N} w_i Q_i(t_v) = S\left(t_v - \tau_{opt} F_s\right) \sum_{i=1}^{N} \frac{A}{r_i} w_i + \\ + \sum_{i=1}^{N} \sum_{f=1}^{k-1} w_i \frac{A_{fi}}{r_{fi}} G_f \left(t_v - (\tau_{opt} + \frac{r_{fi}}{V_s} - \frac{r_{ji}}{V_s}) F_s\right). \quad (35)$$

Каждые 2,4 секунды происходит перерасчет автокорреляционной функции: осуществляется проверка координат действующих источников речи для индикации перемещений и переопределяется оптимальный весовой вектор.

2.3. Структурная схема алгоритма обработки речевых сигналов во временной области

В данной работе для выделения речевых сообщений из смеси голосов предложен оригинальный алгоритм обработки речевых сообщений микрофонной решеткой во временной области, структурная схема которого изображена на Рисунке 21. Уникальность алгоритма обеспечивается введением точных временных задержек, зависящих от пространственных координат нахождения источника речи, а также адаптивным формированием оптимальных на интервалах стационарности весовых коэффициентов микрофонов.

Алгоритм основан на знании координат размещения микрофонов. Знание данных координат позволяет построить матрицу временных задержек для каждой точки пространства наблюдения до начала проведения вычислений.

В блоке введения задержек в каждый сигнал микрофона для каждой точки пространства наблюдения накладывается ранее вычисленная определенная временная задержка и формируются *J* выходных сигналов для всех точек исследуемого пространства. Блок расчета корреляционной функции из *J* сигналов отбирает *P* сигналов наибольшей мощности. По координатам максимумов корреляционной функции происходит определение координат всех одновременно говорящих дикторов. В блоке формирования корреляционной матрицы помехи один из *P* сигналов назначается полезным и дальнейшая обработка направлена на

повышение отношения сигнал/помеха выбранного речевого сообщения. Для нахождения корреляционной матрицы помехи используются сигналы мнимых источников – выделенные с помощью введения задержек речевые сообщения мешающих дикторов. За счет прямого обращения сформированной матрицы помехи производится расчет оптимального весового вектора и рассчитанные весовые коэффициенты передаются на вход блока формирования выходного сигнала с наибольшим отношением сигнал/помеха.

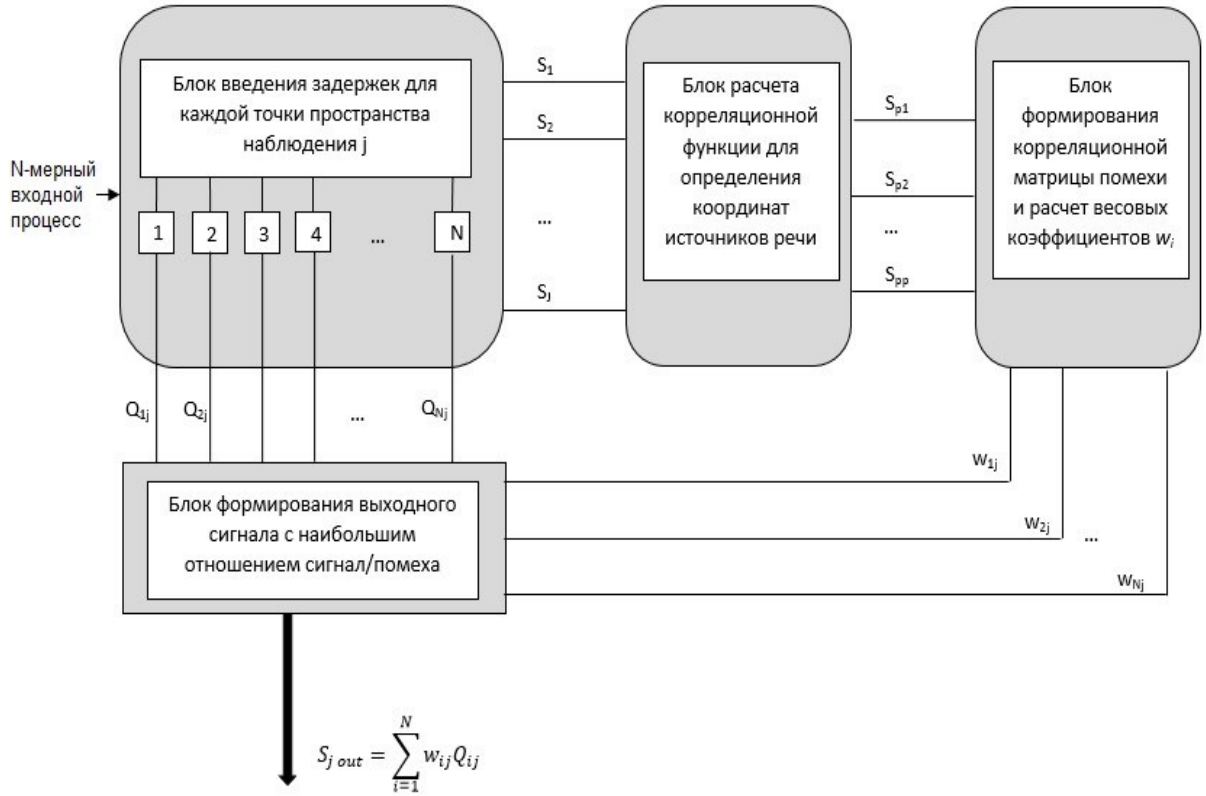


Рисунок 21 – Структурная схема алгоритма обработки речевых сигналов микрофонной решетки во временной области

Наибольшее отношение сигнал/помеха выделенного речевого сообщения достигается совместным применением алгоритма введения задержек и адаптивного расчета оптимальных на интервалах стационарности весовых коэффициентов w_i каждого приемного канала. Каждые 2,4 секунды расчеты обновляются.

2.4. Критерии контроля качества выделяемого речевого сообщения

Для оценки эффективности выделения речевого сообщения из помех от сторонних источников в данной работе использованы два объективных критерия: отношение сигнал/помеха и разборчивость выделенного речевого сообщения.

Отношение сигнал/помеха

Применительно к постановке задаче отношением сигнал/помеха λ в данной работе называется отношение энергии выделенного сигнала к энергии помех от сторонних источников:

$$\lambda = \frac{\sum_v (S\left(t_v - \tau_{opt} F_s\right) \sum_{i=1}^{N} \frac{A}{r_i} w_i)^2}{\sum_v (\sum_{i=1}^{N} \sum_{f=1}^{k-1} \frac{A_{fi}}{r_{fi}} w_i G_f \left(t_v - (\tau_{opt} + \frac{r_{fi}}{V_s} - \frac{r_{ji}}{V_s}) F_s\right))^2}. \quad (36)$$

Для контроля отношения сигнал/помеха выделенного речевого сообщения только с помощью алгоритма введения задержек весовые коэффициенты принимаются равными $w_i = 1$.

Разборчивость речевого сообщения

В настоящее время известно множество методов расчета и измерения разборчивости речи, применяющихся для оценки качества акустики помещений, линий связи, а также защищенности речевой информации. Наибольшее распространение получили методы определения разборчивости, позволяющие получить автоматизированные расчеты, хорошо согласующиеся с субъективными оценками: AI (Articulation Index) – индекс артикуляции; %ALcons (Percentage Articulation Loss of Consonants) – процент артикуляционных потерь согласных; STI (Speech Transmission Index) – индекс передачи речи; RASTI (Rapid Speech Transmission Index) – быстрый индекс передачи речи; SII (Speech Intelligibility Index) — индекс разборчивости речи и многие другие [27, 74-76]. На Рисунке 22 приведена классификация объективных методов оценки разборчивости речи.

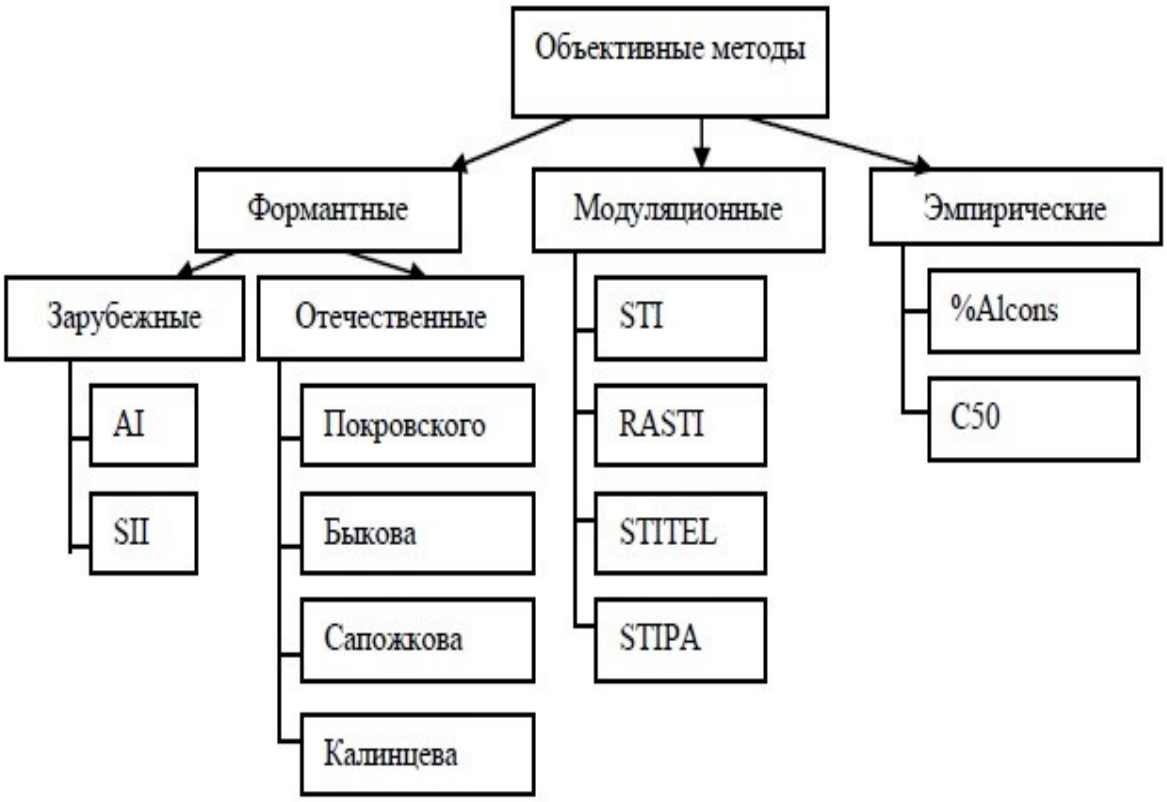


Рисунок 22 – Объективные методы оценки разборчивости речи [74]

Для оценки разборчивости выделенного речевого сообщения в данной работе используется методика, предложенная Н.Б. Покровским [76]. Методика состоит в расчете суммарной разборчивости формант *A:*

$$A = \sum_{k=1}^{n} P_k \Delta A_{mk}, \qquad (37)$$

где P_k – коэффициент восприятия формант в k-й полосе, n – число расчетных полос в спектре сигнала, ΔA_{mk} – максимальная вероятность появления формант в k-й полосе частот. На Рисунке 23 показано распределение формант по частотам.

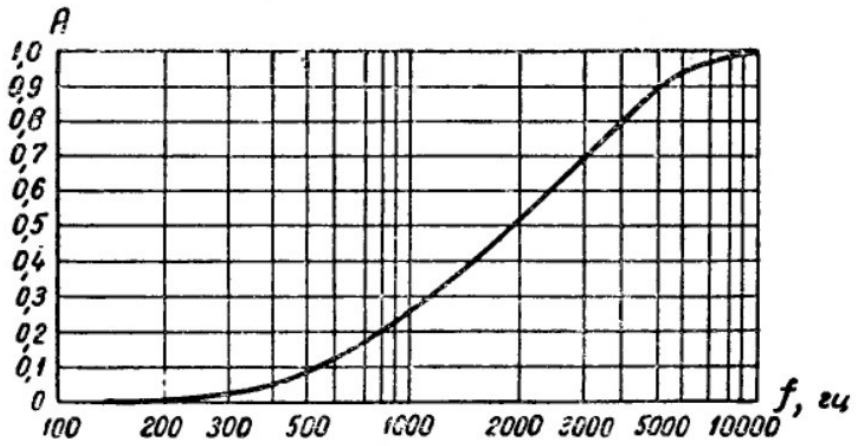


Рисунок 23 – Формантное распределение [76]

В Таблице 2 отражена информация о частотных границах, среднегеометрических частотах и ширине октавных полос, на которые разделяется частотный диапазон речевого сигнала для определения разборчивости.

Таблица 2 – Частотные границы, среднегеометрические частоты и ширина октавных полос

Номер полосы	Частотные границы полос, Гц	Среднегеометрическая частота полосы $f_{ср.i}$, Гц	Ширина полосы Δf_i, Гц
1	90-180	125	90
2	180-355	250	175
3	355-710	500	355
4	710-1400	1000	690
5	1400-2800	2000	1400
6	2800-5600	4000	2800
7	5600-11200	8000	5600

По Рисунку 23 для каждой полосы частот были определены значения ΔA_{mk} как разница в значениях функции на концах исследуемой полосы (Таблица 3).

Таблица 3 – Расчет значений ΔA_{mk}

Номер полосы	Среднегеометрическая частота полосы $f_{ср.i}$, Гц	ΔA_{mk}
1	125	0,01
2	250	0,03
3	500	0,12
4	1000	0,20
5	2000	0,30
6	4000	0,26
7	8000	0,08

Для расчета коэффициента восприятия формант P необходимо использовать Рисунок 24 [77].

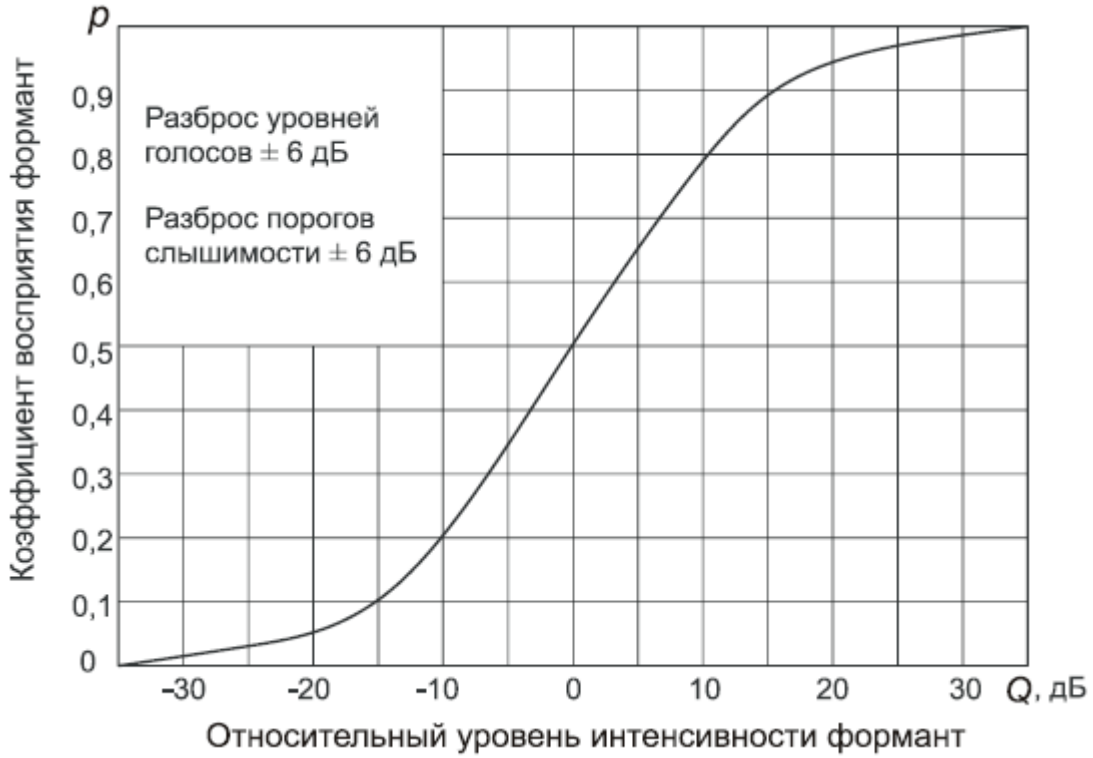


Рисунок 24 – Зависимость коэффициента восприятия P от относительного уровня интенсивности формант Q [77]

Уровень интенсивности формант в каждой полосе определяется по формуле:

$$Q_k = \lambda_k - \Delta B_k, \quad (38)$$

где λ_k – отношение сигнал/помеха для k-й полосы в дБ, ΔB_k – значение формантного параметра, выражающего разницу между спектрами речи и формант в k-й октавной полосе, определяемого по Рисунку 25.

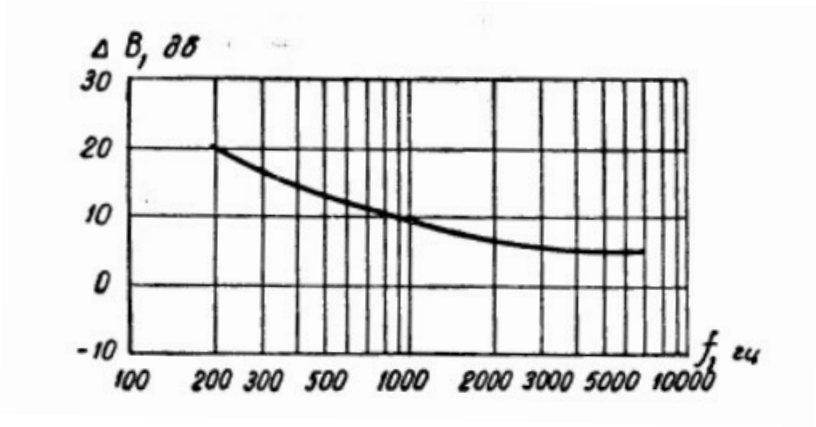
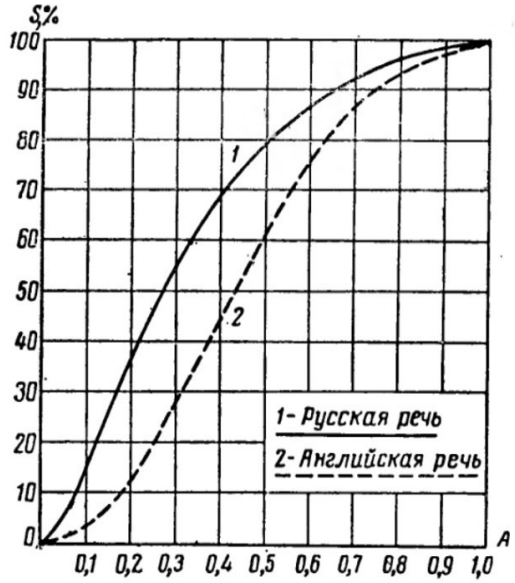
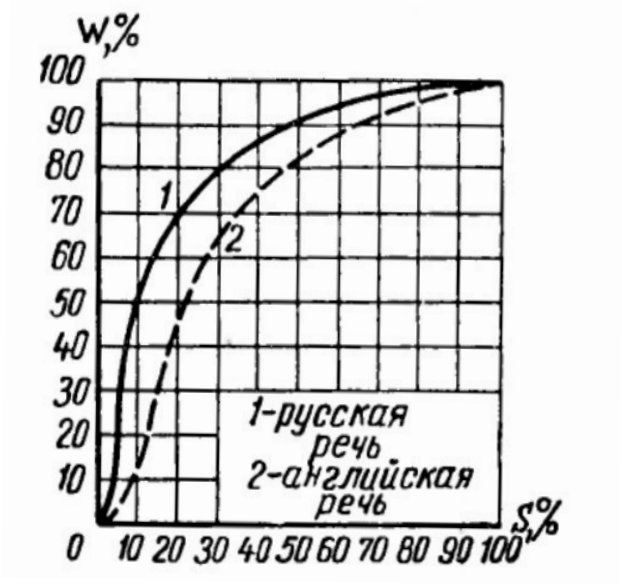


Рисунок 25 – Разность между спектрами речи и формант [76]

После расчета суммарной разборчивости формант A необходимо перейти к слоговой разборчивости S (Рисунок 26 а) и далее к словесной W (Рисунок 26 б).



а) Зависимость слоговой разборчивости от разборчивости формант



б) Зависимость словесной разборчивости от слоговой

Рисунок 26 – Зависимость словесной и слоговой разборчивости от разборчивости формант [76]

В Таблице 4 (ГОСТ 16600-72 [78]) приведены классы качества речевого сообщения по результатам расчета разборчивости звуков, слов, многосложных слов для трактов радиотелефонной связи.

Таблица 4 – Классы качества в соответствии с нормой разборчивости

Класс качества	Характеристика класса качества	Нормы разборчивости, %		
		Звуков	Слов	Многосложных слов
I	Понимание передаваемой речи без малейшего напряжения внимания	Свыше 90	Свыше 95	Свыше 98
II	Понимание передаваемой речи без затруднений	Свыше 85 до 90	Свыше 92 до 95	Свыше 94 до 98
III	Понимание передаваемой речи с напряжением внимания без переспросов и повторений	Свыше 78 до 85	Свыше 87 до 92	Свыше 89 до 94
IV	Понимание передаваемой речи с большим напряжением внимания, переспросами и повторениями	Свыше 60 до 78	Свыше 62 до 87	Свыше 70 до 89
V	Полная неразборчивость связного текста (срыв связи)	До 60	До 62	До 70

Таблица 5 (ГОСТ 50840-95 [79]) содержит информацию о соответствии классов качества речевого сообщения нормам слоговой разборчивости в трактах с параметрическим компандированием и с кодированием волны речевого сигнала, или для синтезированной и естественной речи [80].

Таблица 5 – Классы качества в соответствии с нормой слоговой разборчивости

Класс качества	Характеристика класса качества	Норма слоговой разборчивости для синтезированной речи, %	Норма слоговой разборчивости для естественной речи, %
Высший	Понимание передаваемой речи без малейшего напряжения внимания	>93	>80
I	Понимание передаваемой речи без затруднений	86-93	56-80
II	Понимание передаваемой речи с напряжением внимания без переспросов и повторений	76-85	41-55
III	Понимание передаваемой речи с некоторым напряжением внимания, редкими переспросами и повторениями	61-75	25-40
IV	Понимание передаваемой речи с большим напряжением, частыми переспросами и повторениями	45-60	<25

ВЫВОДЫ ПО ВТОРОЙ ГЛАВЕ

Во второй главе предложен метод пространственной обработки речевых сигналов микрофонной решеткой из N микрофонов, основанный на введении пространственно-зависимых временных задержек в каждый сигнал микрофона. Такой подход позволяет усилить отношение сигнал/помеха в N раз.

Разработан алгоритм обработки сигналов во временной области, максимизирующий отношение сигнал/помеха на выходе решетки для любой точки пространства наблюдения. Увеличение отношения сигнал/помеха достигается применением оптимальных на интервалах стационарности весовых коэффициентов микрофонов решетки.

Описаны два объективных критерия, применяемых для оценки эффективности фильтрации речевых сообщений: отношения сигнал/помеха и разборчивость.

ГЛАВА 3. КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ СИСТЕМЫ ПРОСТРАНСТВЕННОЙ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ

Третья глава посвящена компьютерному моделированию системы пространственной обработки речевых сигналов. Основные результаты третьей главы опубликованы в работах автора [А1-А2, А4, А9-А10, А13-А16].

Компьютерное моделирование системы пространственной обработки речевых сообщений было реализовано при следующих условиях и ограничениях:

1) Для моделирования была выбрана акустическая сцена площадью 24 м^2 (линейные размеры 6 и 4 м соответственно). Измерительная система из ненаправленных микрофонов была размещена на высоте 170 см от пола – средний рост человека. Такое размещение позволяет при моделировании пренебрегать третьим линейным размером исследуемого пространства.

2) Источники речевых сообщений считаются точечными источниками, всенаправленными, звуковое давление уменьшается обратно пропорционально расстоянию от источника звука. В качестве речевых сообщений используются аудиозаписи голосов разных людей. Для точного воспроизведения всех звуков речи требуется частота дискретизации около 20 кГц [81]. Для настоящего исследования частота дискретизации аудиосигналов F_s выбрана равной 22050 Гц. Длительность каждой фразы составляет 4 с.

3) При моделировании рассматривается принципиальная возможность выделения речевого сообщения на фоне помех, поэтому считаем, что звуковой сигнал достигает микрофонов кратчайшим путем – по прямой, реверберация не учитывается.

4) Энергия каждого записанного фрагмента одинакова для каждого человека, за счет нормирования.

5) Источники речевых сообщений стационарны в пространстве. Во время разговора пространственных перемещений нет.

3.1. Нахождение оптимальной конфигурации микрофонной решетки для выделения речевых сообщений из помех

При такой постановке задачи необходимо установить оптимальную для предложенного алгоритма конфигурацию микрофонной решетки для выделения речевых сообщений. Поскольку высота размещения микрофонов зафиксирована, сравнению подлежали три конфигурации микрофонного массива из десяти ненаправленных микрофонов: размещение по одной стене помещения, угловое размещение, размещение микрофонов по периметру помещения (Рисунок 27). При проведении моделирования по нахождению оптимальной конфигурации полагалось, что на акустической сцене одновременно разговаривает четыре человека, один источник условно обозначен за полезный, остальные источники – помехи.

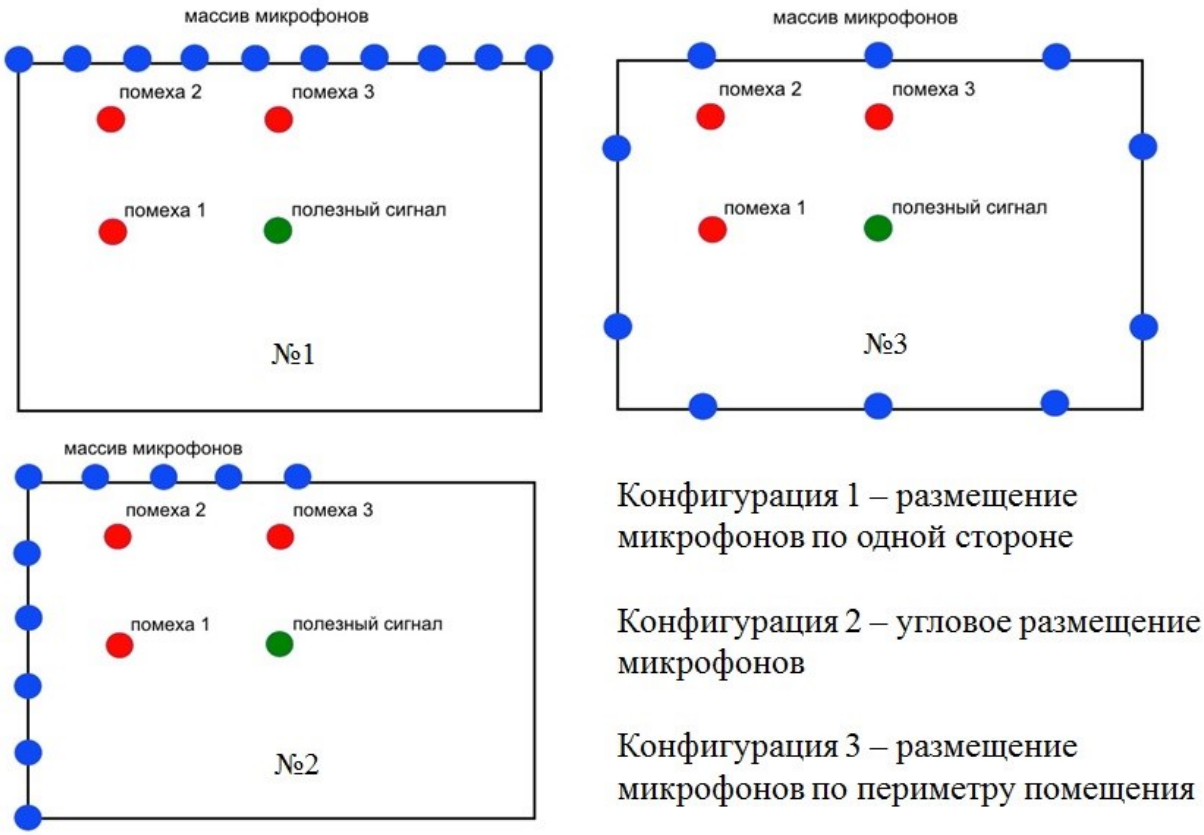


Рисунок 27 – Сравниваемые конфигурации микрофонных решеток. Полезный сигнал в центре акустической сцены

Для конфигурации №1 расстояние между микрофонами равно 67 см; для конфигурации №2 – 80 см и для конфигурации №3 – 200 см. Пространственные координаты источников известны, координаты размещения микрофонов также известны.

Для определения оптимальной конфигурации был произведен расчет отношения сигнал/помеха, а также расчет разборчивости выделенного речевого сообщения. Выделение речевого сообщения было произведено с помощью алгоритма введения задержек. Весовые коэффициенты микрофонов равны единице.

Расчет выполнен для двух разных взаимных расположений источников полезного сигнала и источников помехи: полезный источник находится в центре помещения (Рисунок 27) и полезный сигнал смещен относительно центра акустической сцены (Рисунок 28).

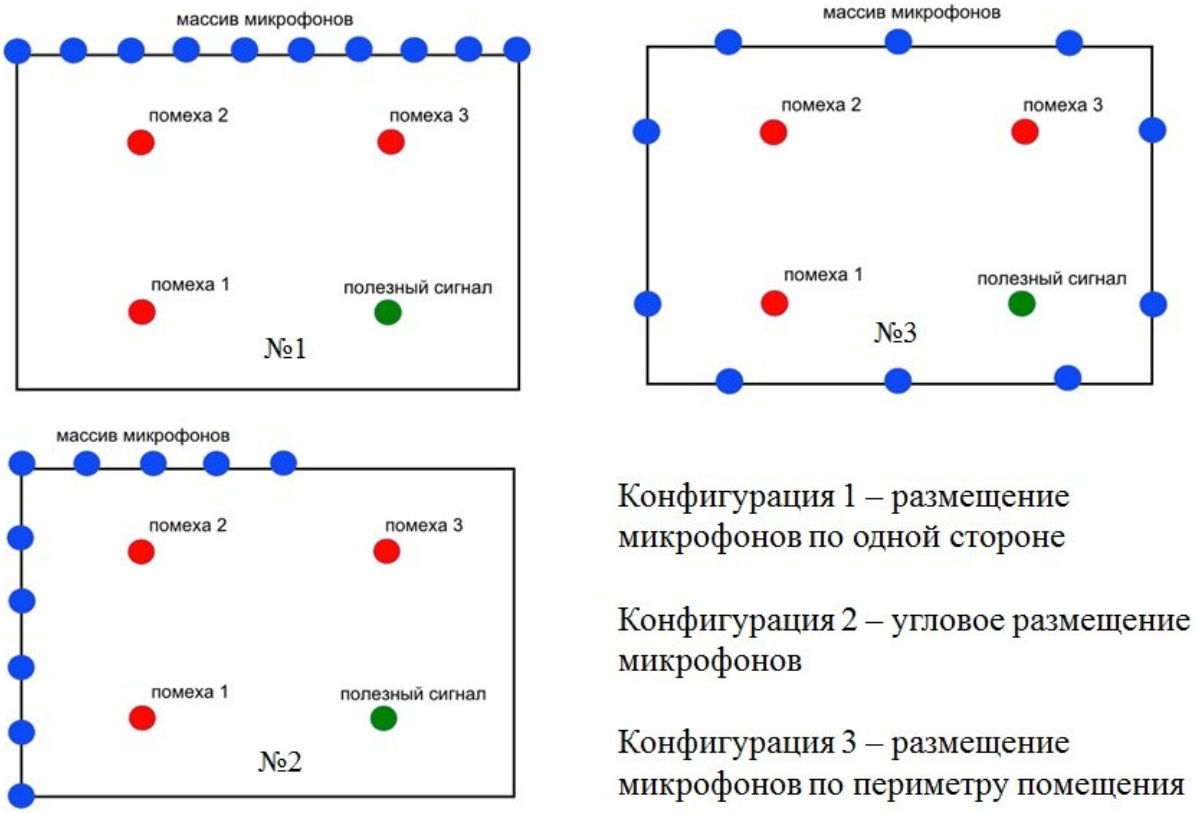


Рисунок 28 – Сравниваемые конфигурации микрофонных решеток. Полезный сигнал смещен от центра акустической сцены

В Таблице 6 представлены результаты расчета отношения сигнал/помеха и показателя разборчивости речевых сообщений, выделенных с помощью временных задержек для трех рассматриваемых конфигураций и для двух различных координат полезного источника.

Таблица 6 – Расчет отношения сигнал/помеха и разборчивости речевых сообщений выделенных с помощью временных задержек для трех рассматриваемых конфигураций для двух различных координат полезного источника

№	Конфигурация многопозиционной системы	Отношение сигнал/помеха выделенного речевого сообщения *λ, отн.ед.*		Разборчивость выделенного речевого сообщения			
				S, %		*W, %*	
		Рис. 27.	Рис. 28.	Рис. 27.	Рис. 28.	Рис. 27.	Рис. 28.
1	размещение микрофонов по одной стене помещения	0,7295	0,6279	52,6	53,7	92,3	92,7
2	угловая конфигурация размещения микрофонов	0,6999	0,8098	53,4	58,2	92,6	94,1
3	размещение микрофонов по периметру помещения	1,1649	2,2123	61,7	70,9	95,1	97,0

На Рисунке 29 показана визуализация расчетных соотношений для трех конфигураций микрофонного массива, указанных в Таблице 6.

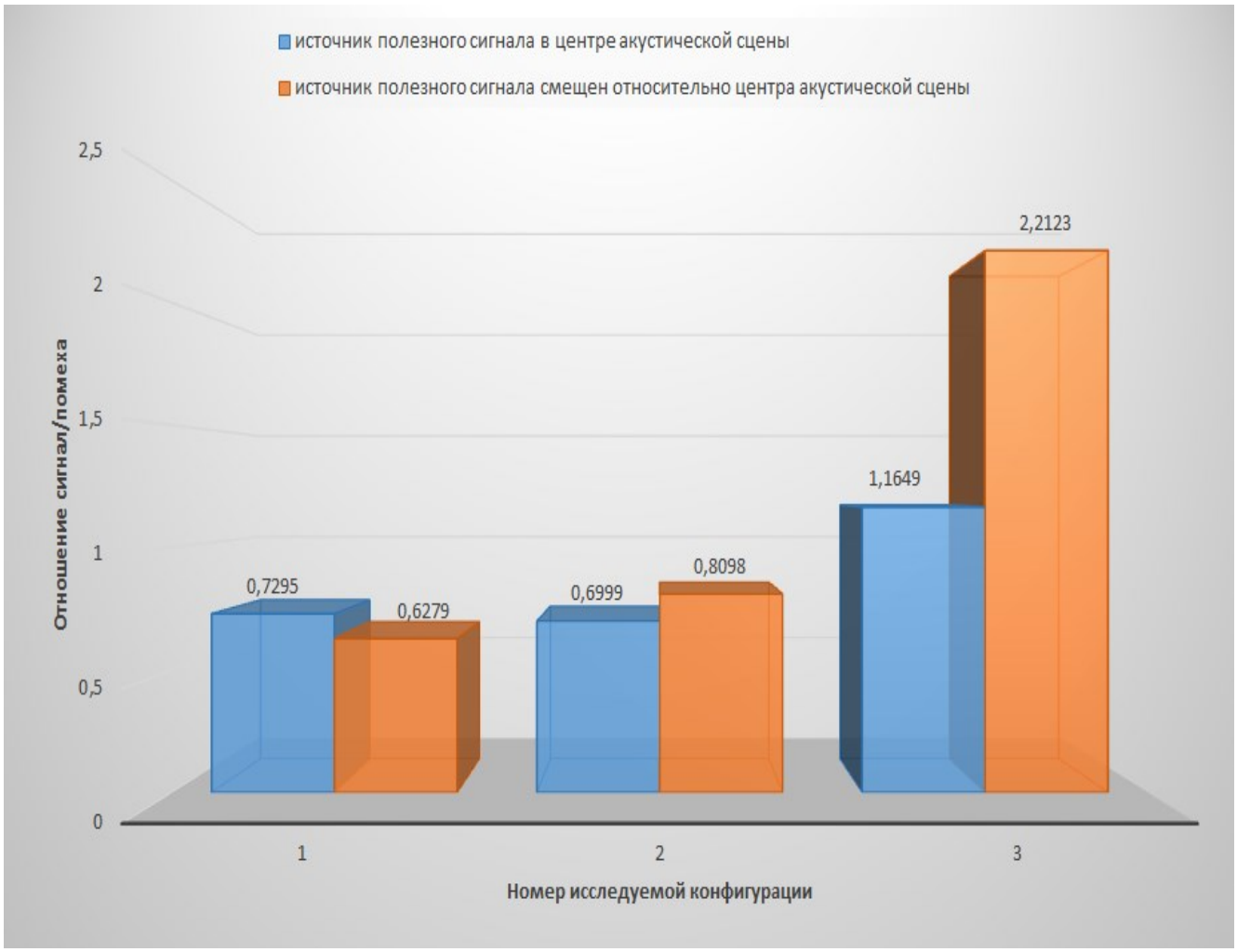


Рисунок 29 – Визуализация расчетных соотношений для трех конфигураций микрофонного массива, указанных в Таблице 6

При выделении речевого сообщения конфигурацией №3 для двух разных положений источника полезного сигнала достигается уровень разборчивости, соответствующий пониманию передаваемой речи без затруднений.

На Рисунке 30 показаны реализации полезного речевого сообщения: исходного неискаженного сообщения (верхний сигнал) и выделенного из помех с наибольшим отношением сигнал/помеха (нижний).

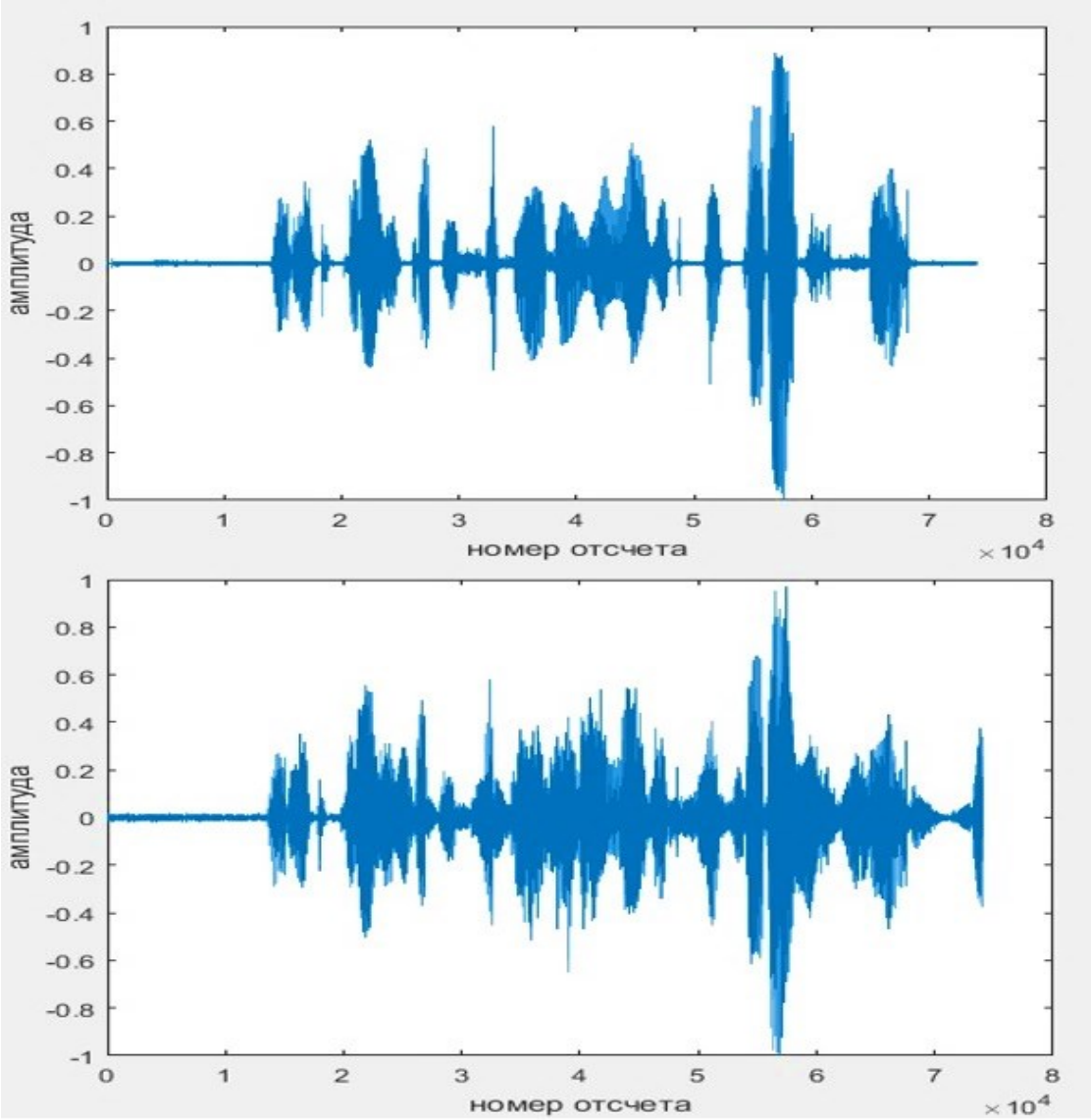


Рисунок 30 – Реализация исходного полезного сигнала (верхний) и сигнала, выделенного из помех конфигурацией № 3 с наилучшим отношением сигнал/помеха (нижний)

На Рисунке 31 показан расчет коэффициента взаимной корреляции для исходного полезного и выделенного из помех речевого сообщения (Рисунок 30), который доказывает эффективность фильтрации полезного сигнала из помех.

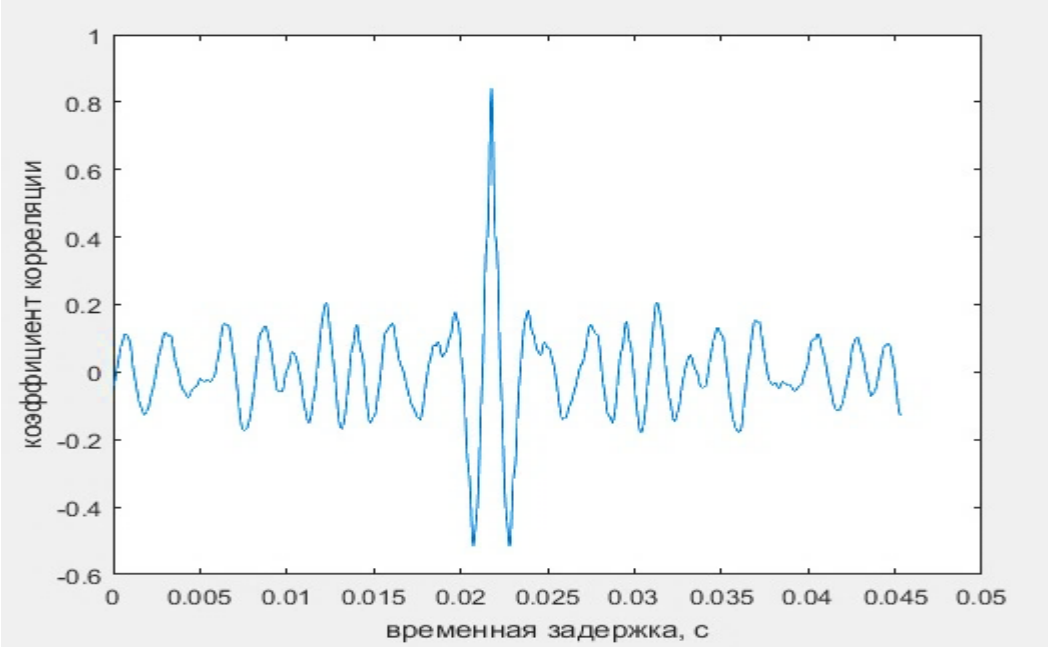


Рисунок 31 – Коэффициент корреляции исходного полезного сигнала и сигнала, выделенного из помех конфигурацией № 3 с наилучшим отношением сигнал/помеха

При различном взаимном расположении полезного сигнала и помех конфигурация №3 обеспечивает наилучшие показатели отношения сигнал/помеха и разборчивости выделенного речевого сообщения, а, следовательно, конфигурация с размещением микрофонов по периметру для ограниченного пространства является оптимальной для предложенного алгоритма.

3.2. Исследование пространственной разрешающей способности многопозиционной акустической системы

Для оценки пространственной разрешающей способности акустической системы на той же высоте (170 см) по периметру помещения были распределены двадцать микрофонов. В центре акустической сцены находится один источник полезного сигнала, сторонние источники отсутствуют.

Для каждой точки пространства вводятся определенные временные задержки, пропорциональные расстоянию от данной точки до определенного

микрофона и производится расчет пространственной автокорреляционной функции. Результаты расчета автокорреляционной функции по пространственным координатам отображены на Рисунке 32.

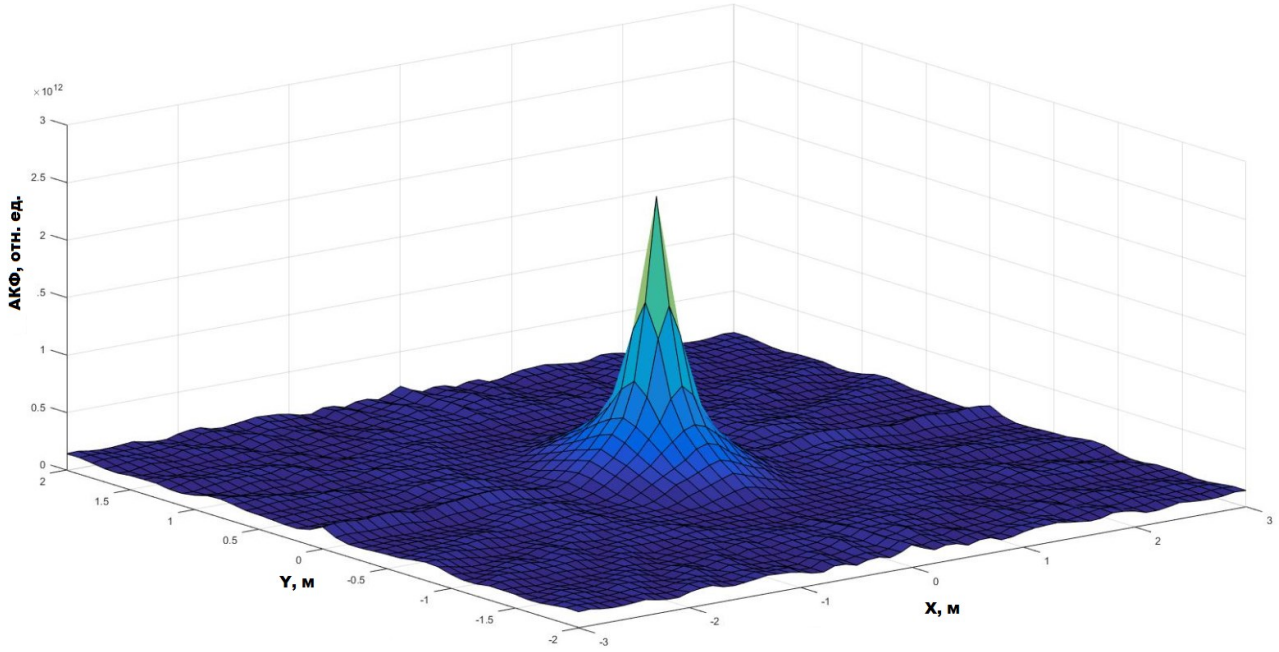


Рисунок 32 – Расчет пространственной автокорреляционной функции для одного акустического источника речи (70-7000 Гц)

Для определения разрешающей способности было получено сечение корреляционной функции по уровню -3 дБ (Рисунок 33).

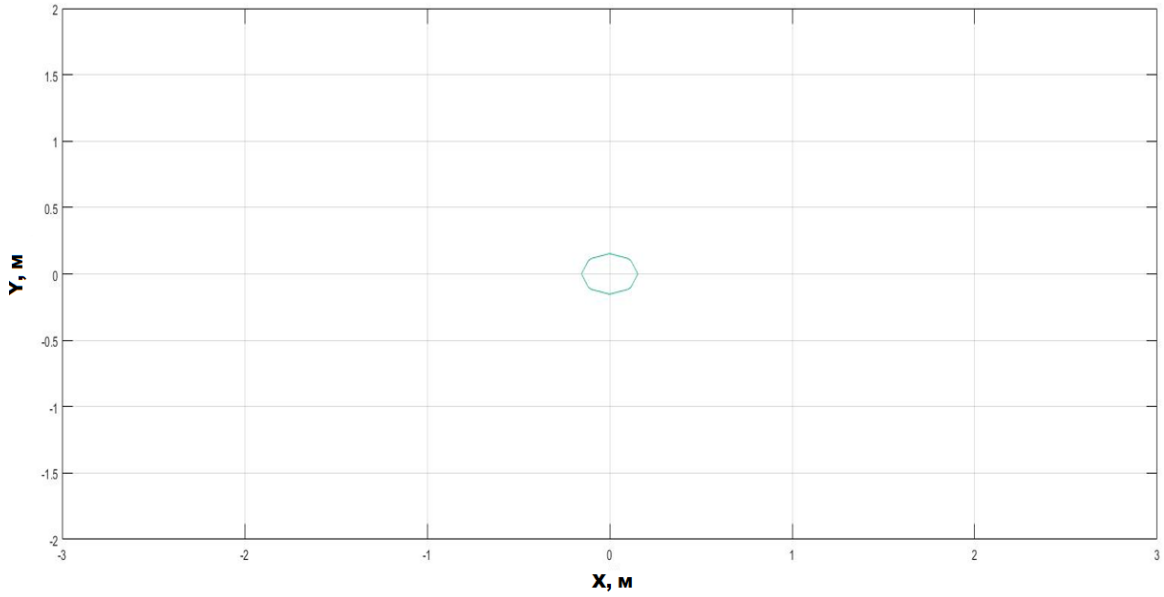


Рисунок 33 – Сечение корреляционной функции по уровню -3 дБ

Пространственная разрешающая способность акустической системы на частоте 70-7000 Гц составила 28 см.

Для выделения сигнала линии связи (частотный диапазон 300-3400 Гц) разрешающая способность составляет 24 см (Рисунок 34).

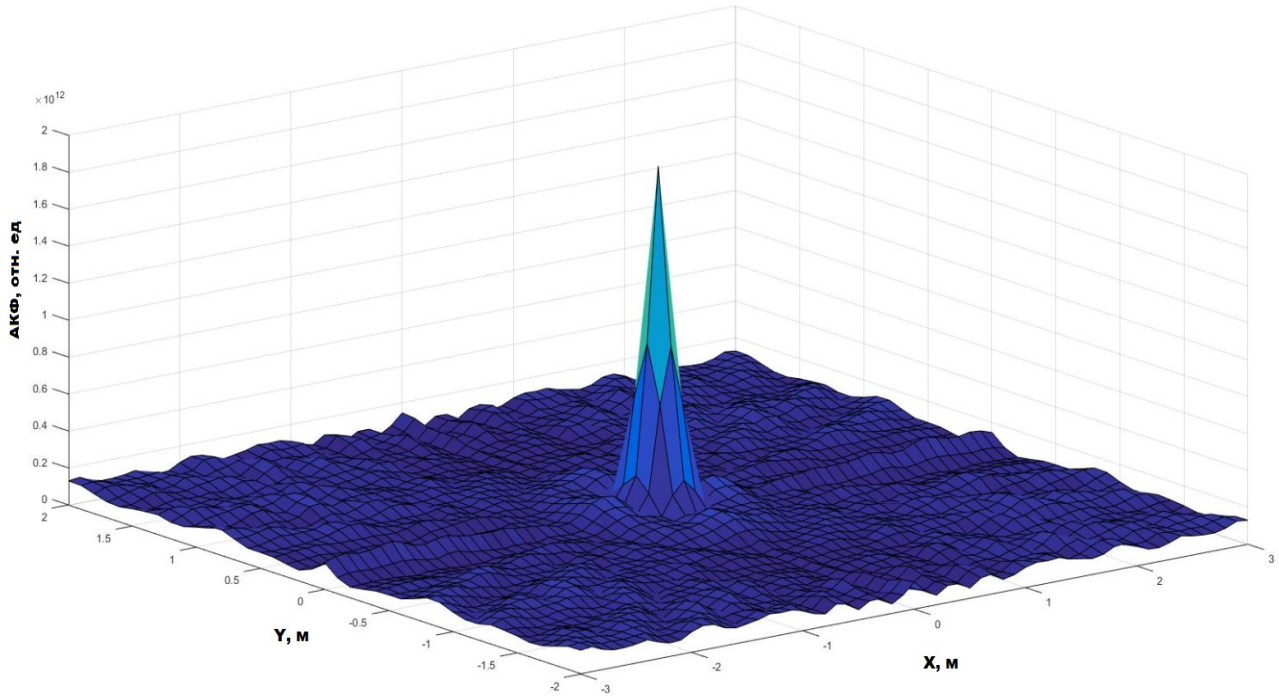


Рисунок 34 – Расчет пространственной автокорреляционной функции для одного акустического источника речи (300-3400 Гц)

При увеличении диапазона частот принимаемых сигналов пространственная разрешающая способность снижается: так при расширении спектра сигнала до 10,5 кГц разрешающая способность составляет 39 см (Рисунок 35).

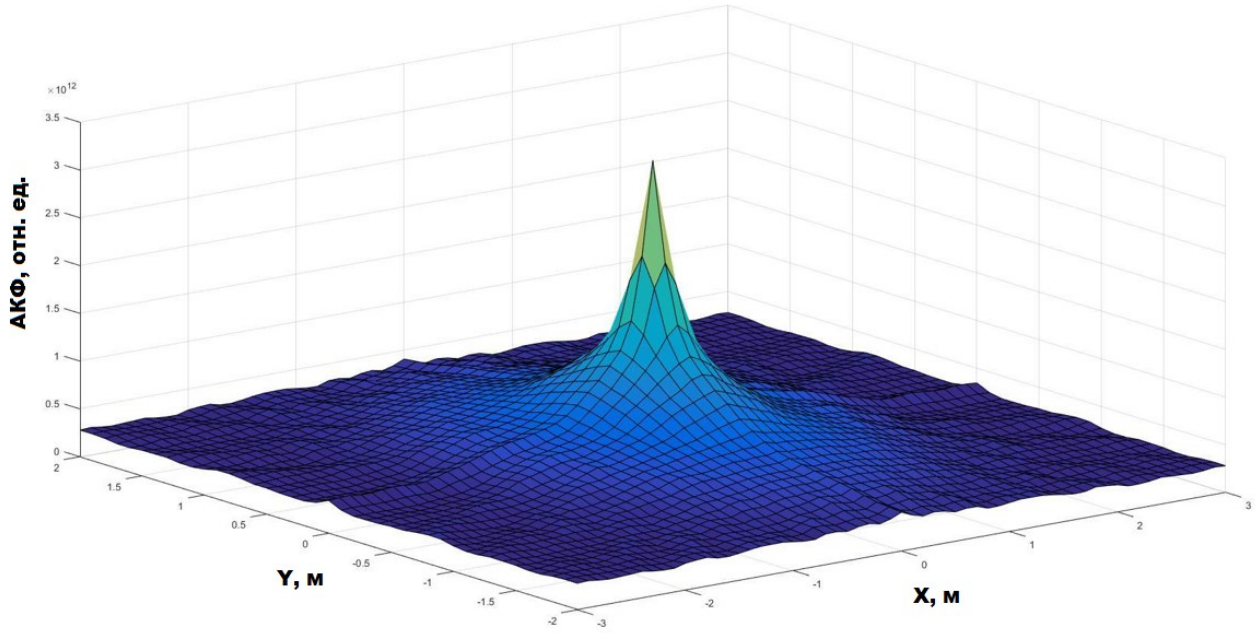


Рисунок 35 – Расчет пространственной автокорреляционной функции для одного акустического источника речи (100-10600 Гц)

Расчет пространственной автокорреляционной функции позволяет определять координаты акустических источников. Это подтверждается численным экспериментом по определению координат акустических источников восьми одновременно говорящих людей. Частотный диапазон работы акустической системы 70-7000 Гц. Координаты источников речи не известны, известны координаты размещения микрофонов. После введения временных задержек, зависящих от пространственных координат, производится расчет пространственной автокорреляционной функции в каждой точке пространства. Результаты расчета приведены на Рисунке 36.

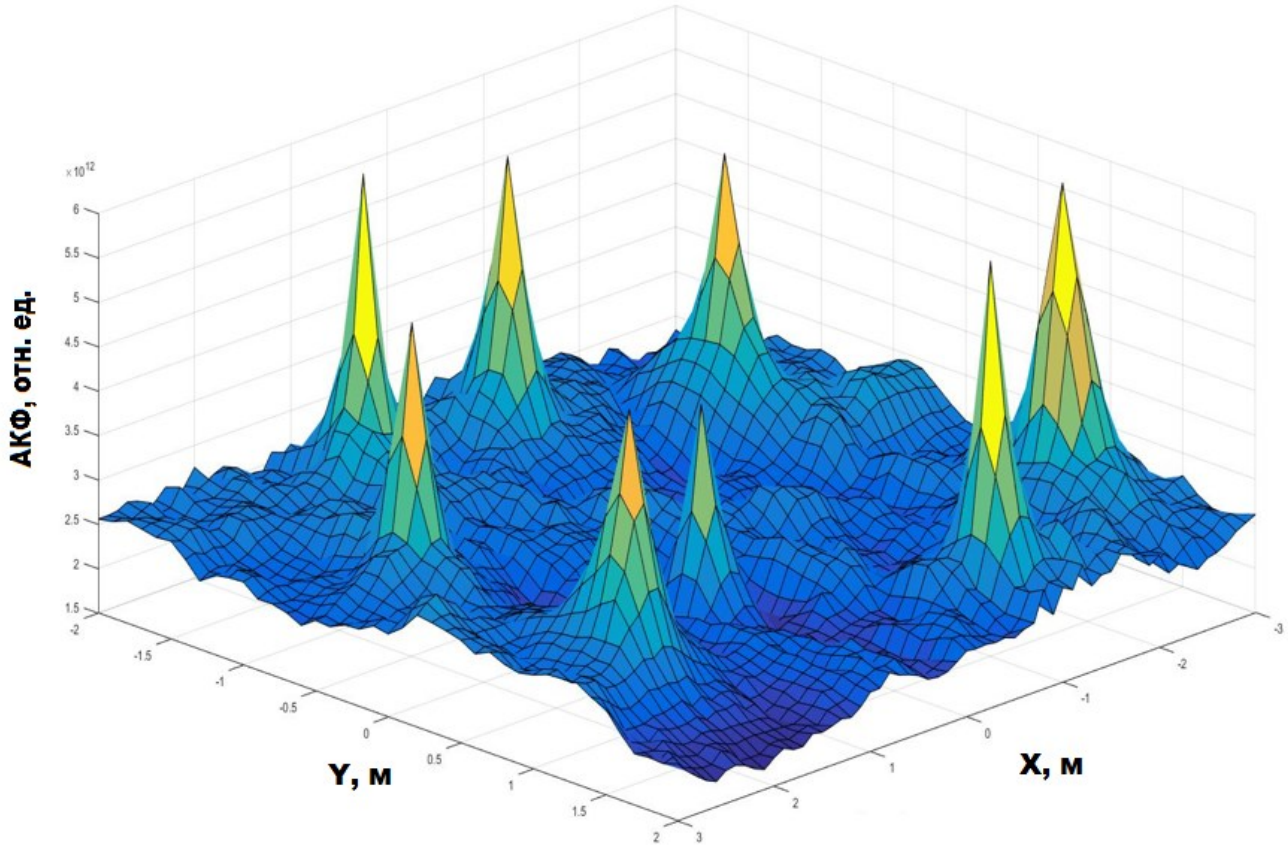


Рисунок 36 – Расчет пространственной автокорреляционной функции для восьми акустических источников речи (70-7000 Гц)

Из Рисунка 36 следует, что наличие восьми максимумов пространственной автокорреляционной функции соответствует восьми источникам речевой информации. Координаты максимумов соответствуют координатам источников. Построим картину изолиний (Рисунок 37).

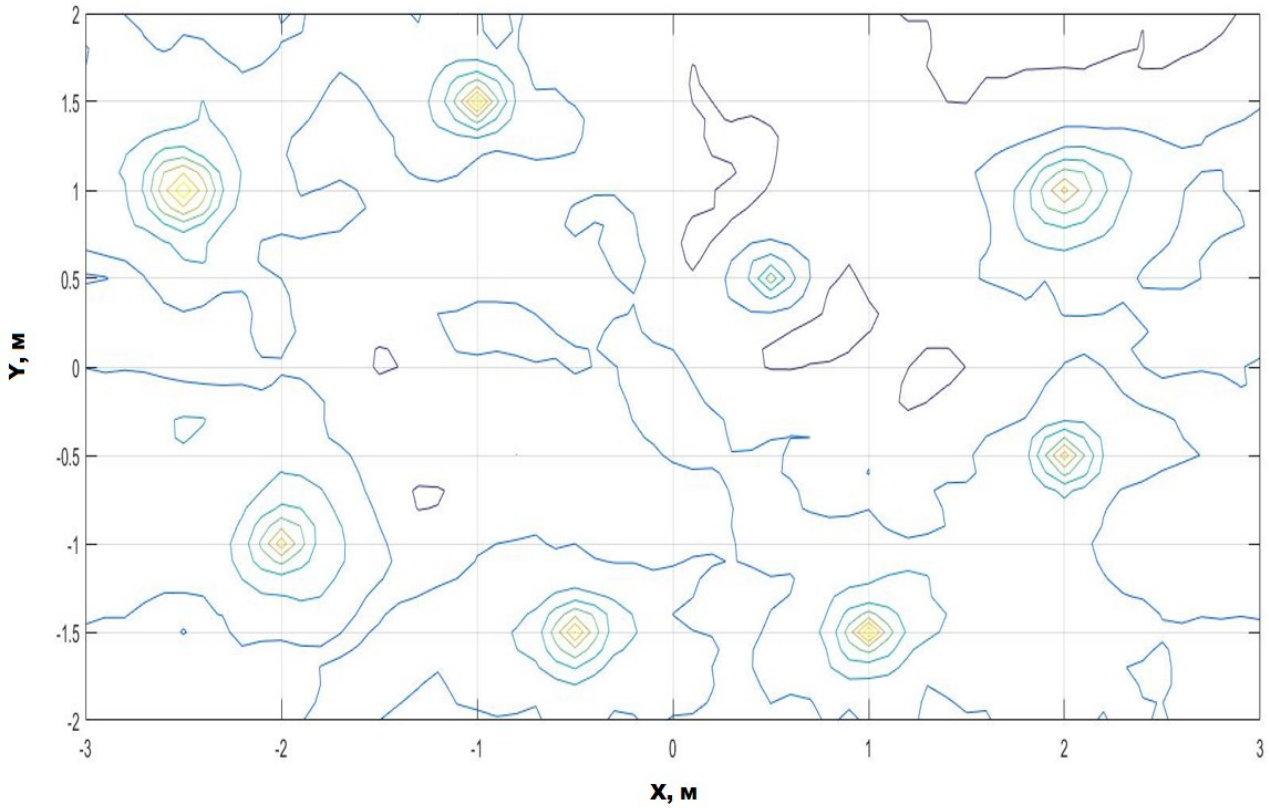


Рисунок 37 – Картина изолиний для восьми акустических источников речи (70-7000 Гц)

По данному изображению можно определить координаты восьми источников (Таблица 7).

Таблица 7 – Координаты восьми источников речевых сообщений

Диктор 1	Диктор 2	Диктор 3	Диктор 4	Диктор 5	Диктор 6	Диктор 7	Диктор 8
$x_1 = -0.5$ $y_1 = -1.5$	$x_2 = 1$ $y_2 = -1.5$	$x_3 = -2$ $y_3 = -1$	$x_4 = 2$ $y_4 = -0.5$	$x_5 = 0.5$ $y_5 = 0.5$	$x_6 = -2.5$ $y_6 = 1$	$x_7 = 2$ $y_7 = 1$	$x_8 = -1$ $y_8 = 1.5$

Таким образом, расчет пространственной автокорреляционной функции позволяет определять координаты всех акустических источников. Разрешающая способность акустической системы для обработки речи (70-7000 Гц) составила 28 см. Для исследования пространства наблюдения обрабатывать точки пространства без потери акустических источников следует с соответствующим шагом. Отметим,

что такая разрешающая способность позволяет разделить речевые сообщения двух человек, расположенных очень близко друг к другу.

3.3. Расчет оптимальных весовых коэффициентов

В п.3.2 было показано, что акустическая система из двадцати микрофонов может определить пространственные координаты восьми источников речи. Следовательно, после введения задержек, соответствующих координатам источников, могут быть выделены восемь голосов – S_{p1}, ... S_{p8}. На Рисунке 38 показано взаимное расположение микрофонной решетки и источников речи.

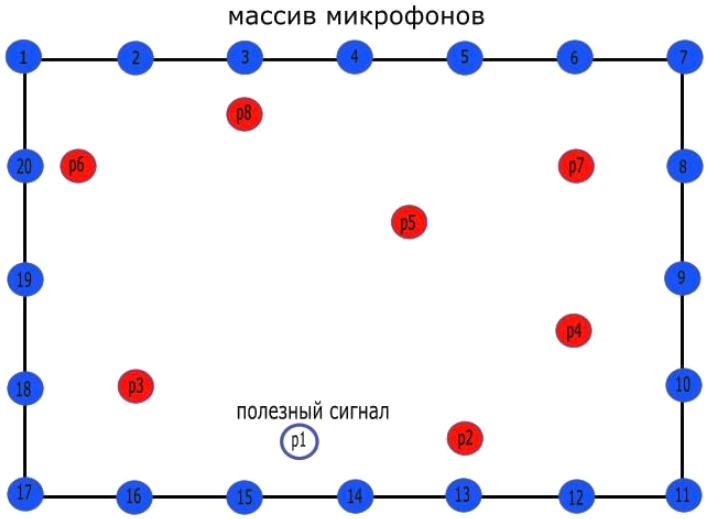


Рисунок 38 – Взаимное расположение восьми одновременно говорящих людей и решетки из двадцати микрофонов

Для расчета весовых коэффициентов микрофонов необходимо один из голосов условно назначить полезным S_{p1}, а остальные S_{p2}, ... S_{p8} выделенные голоса – считать мнимыми источниками сторонних помех. После выделения данных сигналов алгоритм обработки формирует корреляционную матрицу каждой помехи в соответствии с формулой (39):

$$\boldsymbol{M_{pi}} = \begin{pmatrix} \left(\sum \frac{A_{pi}}{r_{pi1}} S_{pi}\,(t_v - \Delta_1)\right)^2 & \dots & \sum \frac{A_{pi}}{r_{piN}} S_{pi}(t_v - \Delta_N) \sum \frac{A_{pi}}{r_{pi1}} S_{pi}(t_v - \Delta_1) \\ \sum \frac{A_{pi}}{r_{pi1}} S_{pi}(t_v - \Delta_1) \sum \frac{A_{pi}}{r_{pi2}} S_{pi}(t_v - \Delta_2) & \dots & \sum \frac{A_{pi}}{r_{piN}} S_{pi}(t_v - \Delta_N) \sum \frac{A_{pi}}{r_{pi2}} S_{pi}(t_v - \Delta_2) \\ \dots & \dots & \dots \\ \sum \frac{A_{pi}}{r_{pi1}} S_{pi}(t_v - \Delta_1) \sum \frac{A_{pi}}{r_{piN}} S_{pi}(t_v - \Delta_N) & \dots & \left(\sum \frac{A_{pi}}{r_{piN}} S_{pi}\,(t_v - \Delta_N)\right)^2 \end{pmatrix}. \quad (39)$$

И корреляционную матрицу полезного сигнала:

$$\boldsymbol{S} = \begin{pmatrix} \frac{A}{r_{p1}} \sum S_{p1}\left(t_v - \tau_{opt} F_s\right) \\ \frac{A}{r_{p2}} \sum S_{p1}\left(t_v - \tau_{opt} F_s\right) \\ \dots \\ \frac{A}{r_{pN}} \sum S_{p1}\left(t_v - \tau_{opt} F_s\right) \end{pmatrix}. \quad (40)$$

Оптимальные весовые коэффициенты рассчитываем по формуле:

$$\boldsymbol{W = M^{-1} S.} \quad (41)$$

В Таблице 8 представлены результаты расчета весовых коэффициентов многопозиционной акустической системы из двадцати микрофонов.

Таблица 8 – Результаты расчета весовых коэффициентов многопозиционной системы из двадцати микрофонов

w_1	0,04761	w_6	0,03294	w_{11}	0,00664	w_{16}	0,15032
w_2	0,00004	w_7	0,12427	w_{12}	0,17115	w_{17}	0,06687
w_3	0,11972	w_8	0,03599	w_{13}	0,20674	w_{18}	0,20680
w_4	0,12166	w_9	0,04965	w_{14}	0,48802	w_{19}	0,17697
w_5	0,22382	w_{10}	0,12916	w_{15}	0,68042	w_{20}	0,08472

Сопоставим результаты расчета весовых коэффициентов, приведенных в Таблице 8 и Рисунок 38. Наибольшие значения весов у микрофонов №14 и №15 – ближайшие микрофоны по расположению к полезному источнику речевого сигнала. Оптимальные весовые коэффициенты направлены на усиление полезного сигнала и ослабление сигналов сторонних источников. Именно поэтому

максимальное и минимальное значение весовых коэффициентов в данном конкретном случае различается в 17000 раз.

Необходимо убедиться, что при изменении помеховой обстановки весовые коэффициенты будут изменяться. Расположение источников речевых сообщений не изменяется, источник полезного сигнала расположен в том же месте, изменяется начитываемый всеми дикторами текст. Из Таблицы 9 можно убедиться в том, что из-за смены помеховой обстановки изменяются и весовые коэффициенты, но по-прежнему максимальные значения остаются у микрофонов, которые расположены ближе к источнику полезного сигнала.

Таблица 9 – Результаты расчета весовых коэффициентов многопозиционной системы из двадцати микрофонов

w_1	0,00516	w_6	0,03218	w_{11}	0,04604	w_{16}	0,09673
w_2	0,01310	w_7	0,04518	w_{12}	0,12340	w_{17}	0,13054
w_3	0,09563	w_8	0,08841	w_{13}	0,24374	w_{18}	0,23935
w_4	0,22733	w_9	0,01245	w_{14}	0,58146	w_{19}	0,12085
w_5	0,18048	w_{10}	0,03900	w_{15}	0,61531	w_{20}	0,04562

Произведем расчет для другой точки пространства наблюдения. В качестве полезного сигнала обозначим сигнал S_{p6} (Рисунок 39). В Таблице 10 представлен расчет оптимальных весовых коэффициентов для точки пространства с координатами полезного источника.

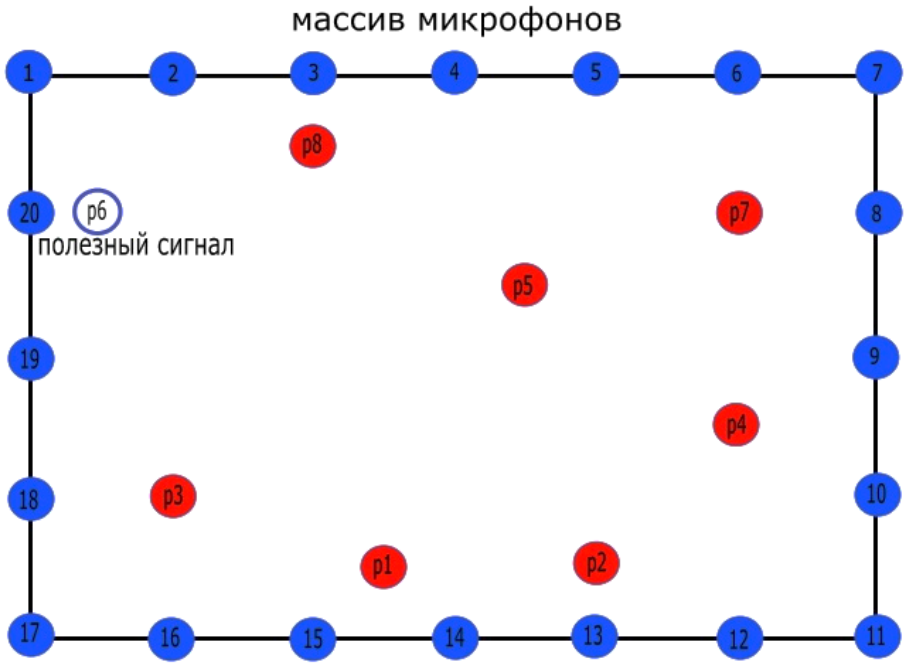


Рисунок 39 – Взаимное расположение восьми одновременно говорящих людей и решетки из двадцати микрофонов

Таблица 10 – Результаты расчета весовых коэффициентов многопозиционной системы из двадцати микрофонов

w_1	0,22974	w_6	0,00074	w_{11}	0,00920	w_{16}	0,22115
w_2	0,11411	w_7	0,00186	w_{12}	0,01321	w_{17}	0,12748
w_3	0,30255	w_8	0,00041	w_{13}	0,00316	w_{18}	0,10700
w_4	0,06816	w_9	0,02420	w_{14}	0,02811	w_{19}	0,16710
w_5	0,01269	w_{10}	0,02892	w_{15}	0,05348	w_{20}	0,85322

Максимальное значение весового коэффициента микрофона № 20 свидетельствует о корректной работе предложенного алгоритма.

Для усиления сигнала полезного источника и уменьшения влияния помех алгоритм обработки определяет максимальное значение у того микрофона, который наиболее близко расположен относительно полезного источника.

3.4. Исследование эффективности фильтрации полезного речевого сигнала на фоне интенсивных распределенных помех

Эффективность фильтрации полезного сигнала определяется по отношению сигнал/помеха выделенного речевого сообщения, а также по показателю разборчивости речи.

Взаимное расположение микрофонной решетки и источников речи соответствует Рисунку 38. На Рисунке 40 показана реализация акустической обстановки – запись голосов восьми одновременно говорящих людей одним микрофоном.

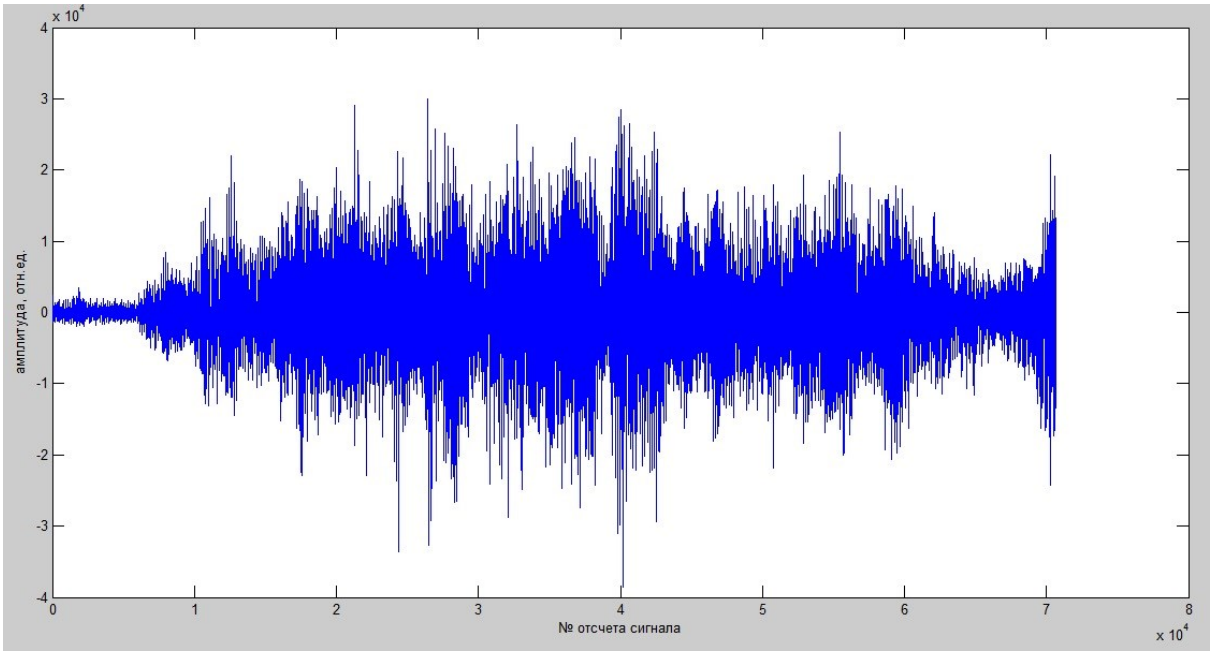


Рисунок 40 – Запись голосов восьми одновременно говорящих людей одним микрофоном

Рисунок 41 иллюстрирует сигнал полезного источника до внесения искажений.

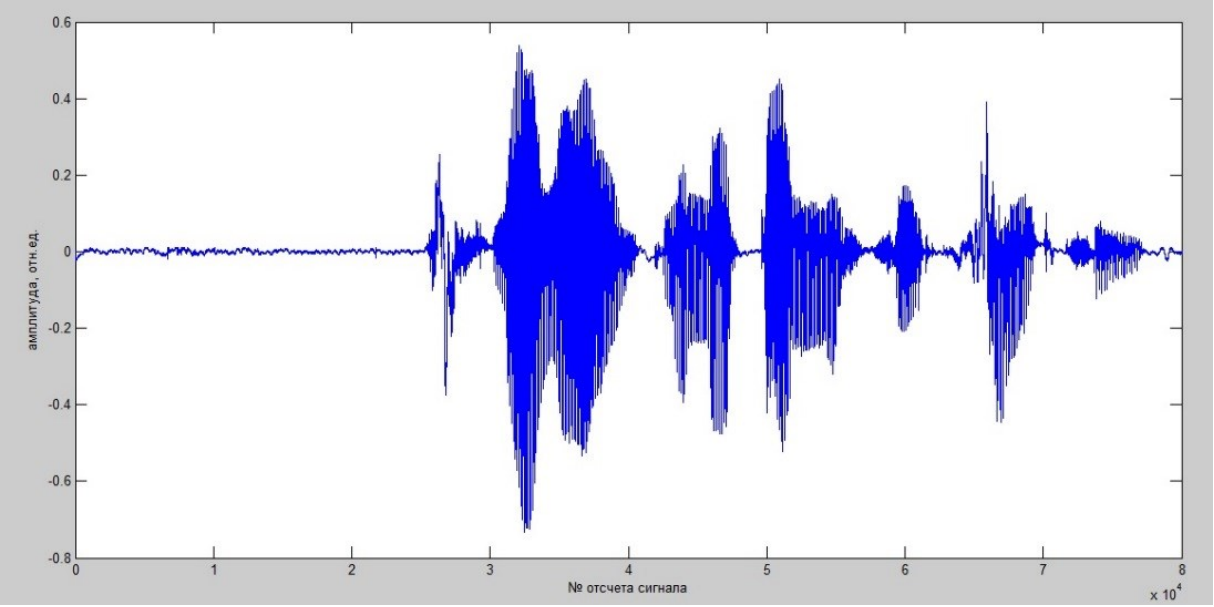


Рисунок 41 – Исходный сигнал полезного источника

На первом этапе разработанный алгоритм обработки речевых сообщений выделяет с помощью метода пространственной фильтрации, основанного на введении задержек, зависящих от пространственных координат, все зашумленные речевые сообщения. Реализация выделенного из помех сигнала полезного источника приведена на Рисунке 42. Остальные семь речевых сообщений необходимы для формирования корреляционной матрицы помехи.

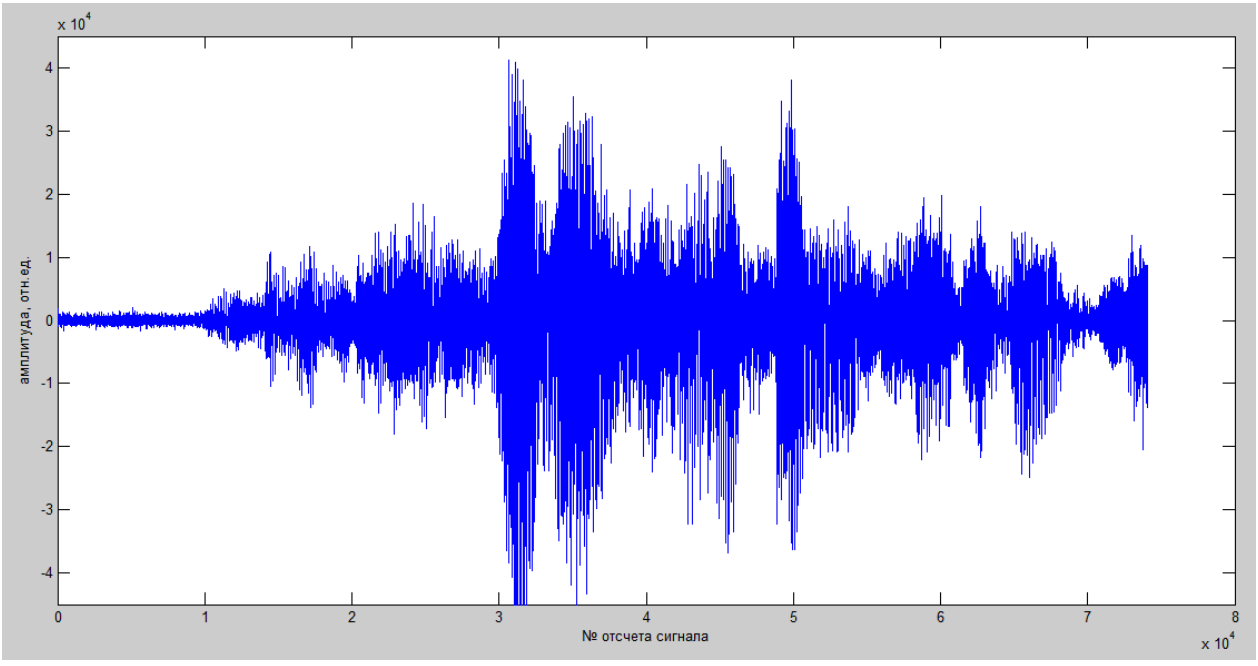


Рисунок 42 – Реализация выделенного с помощью алгоритма введения задержек из помех полезного сигнала

Эффективность выделения речевого сообщения возрастает за счет дальнейшего расчета оптимальных на интервалах стационарности весовых коэффициентов микрофонов. Оптимальный весовой вектор решетки определяется через корреляционную матрицу помехи (см. п.3.3).

Таким образом, предложенный алгоритм позволяет увеличить эффективность выделения речи целевого диктора: с помощью введения временных задержек выделяется полезное речевое сообщение, отношение сигнал/помеха которого увеличивается за счет применения весовых коэффициентов. Реализация выделенного речевого сообщения полезного источника с учетом весовых коэффициентов показана на Рисунке 43.

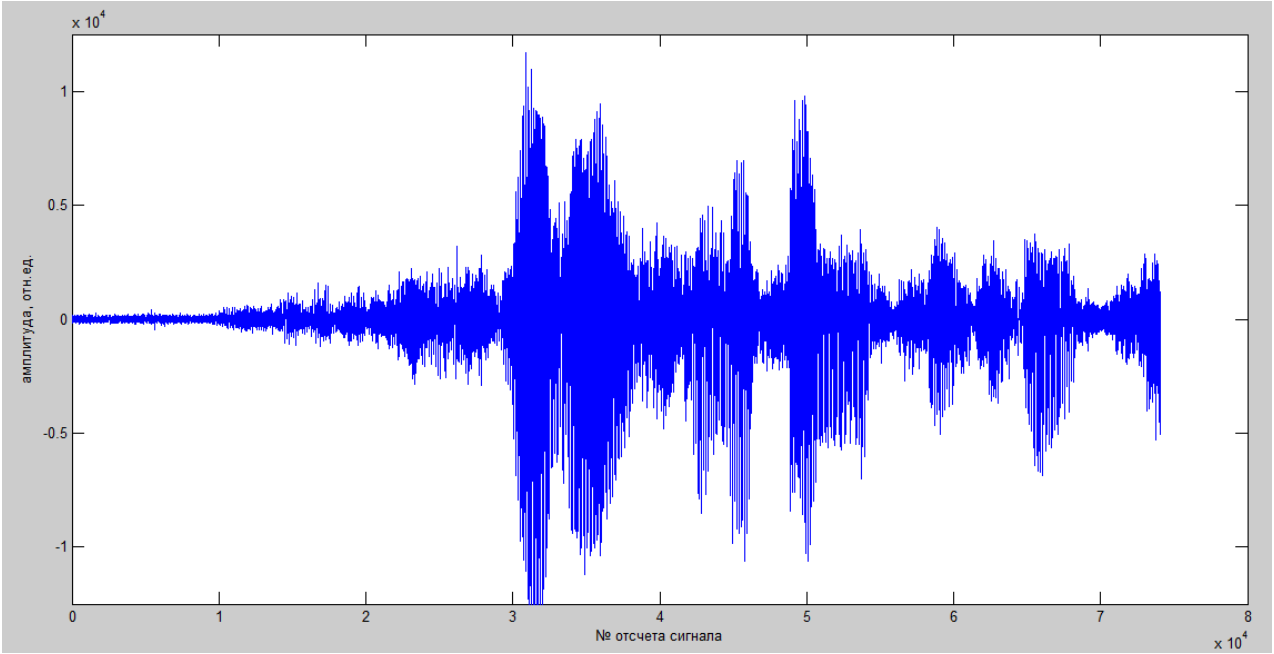


Рисунок 43 – Реализация выделенного полезного сигнала предложенным алгоритмом

Степень влияния помех от сторонних источников при применении весовых коэффициентов уменьшается, огибающая речевого сообщения становится практически идентичной огибающей исходного сигнала.

В Таблице 11 приведены результаты расчета отношения сигнал/помеха для сигналов, изображенных на Рисунках 40, 42 и 43. Таблица 12 демонстрирует расчет разборчивости речевых сообщений.

Таблица 11 – Результаты расчета отношения сигнал/помеха для сигналов, изображенных на Рисунках 40, 42 и 43

Название алгоритма	Отношение сигнал/помеха, разы
Запись одним микрофоном без пространственной фильтрации (Рисунок 40)	0,147
Алгоритм пространственной фильтрации, основанный на введении точных временных задержек, зависящих от пространственных координат (Рисунок 42)	1,972
Алгоритм обработки речевого сигнала микрофонной решеткой во временной области, максимизирующий отношение сигнал/помеха (Рисунок 43)	6,641

Таблица 12 – Расчет разборчивости выделенных речевых сообщений

Название алгоритма	Разборчивость речевого сообщения (%)		Характеристика класса качества
	Слоговая S	Словесная W	
Запись одним микрофоном без пространственной фильтрации (Рисунок 40)	10	44	Полная неразборчивость связного текста
Алгоритм пространственной фильтрации, основанный на введении точных временных задержек, зависящих от пространственных координат (Рисунок 42)	50	91	Понимание передаваемой речи с напряжением внимания без переспросов и повторений
Алгоритм обработки речевого сигнала микрофонной решеткой во временной области, максимизирующий отношение сигнал/помеха (Рисунок 43)	72	97	Понимание передаваемой речи без затруднений

При выделении речи целевого диктора алгоритмом пространственной фильтрации без весовых коэффициентов достигается выигрыш (для определенного случая) в 11,2 дБ по отношению к записи акустической обстановки одним микрофоном (Таблица 11). Результат согласовывается с теорией (см. п.2.1), согласно которой при использовании двадцати микрофонов алгоритм пространственной фильтрации без весовых коэффициентов обеспечивает выигрыш в отношении сигнал/помеха не более чем в 20 раз, т.е. 13 дБ. Дальнейшее применение весовых коэффициентов увеличивает выигрыш в отношении сигнал/помеха с 11,2 до 16,5 дБ.

Таким образом, эффективность предложенного в работе алгоритма зависит от количества микрофонов в решетке, количества одновременно говорящих дикторов и их взаимного расположения в пространстве.

ВЫВОДЫ ПО ТРЕТЬЕЙ ГЛАВЕ

Чем ближе расположен полезный источник к одному из микрофонов решетки, тем эффективнее работа алгоритма. При вынесении микрофонов из плоскости размещения источников речевых сообщений, расстояние от источника полезного сигнала до ближайшего микрофона увеличивается, а разность расстояний до всех микрофонов решетки уменьшается, что приводит к снижению эффективности предлагаемого подхода. Таким образом, для предлагаемого алгоритма определена оптимальная конфигурация микрофонной решетки для выделения речевых сообщений из помех: микрофонная решетка с размещением микрофонов по периметру помещения в плоскости локализации речевых источников, позволяет выделять речевые сообщения с наибольшим отношением сигнал/помеха.

Исследована пространственная разрешающая способность акустической системы. Для исследуемого диапазона частот 70-7000 Гц разрешающая способность составила 28 см. При значительном изменении значений верхней и нижней граничных частот диапазона речи разрешающая способность акустической

системы изменяется слабо. Такая разрешающая способность системы позволяет разделить речевые сообщения двух человек, расположенных очень близко друг к другу.

Проведенные исследования по применению предлагаемого алгоритма обработки для выделения речевых сообщений из помех для определенной точки пространства наблюдения свидетельствуют о возможном достижении выигрыша по отношению сигнал/помеха для конкретного случая более, чем в N раз, где N – число микрофонов в решетке.

ГЛАВА 4. ЧИСЛЕННЫЙ ЭКСПЕРИМЕНТ ПО ВЫДЕЛЕНИЮ РЕЧЕВОГО СООБЩЕНИЯ ИЗ ГОЛОСОВОЙ СМЕСИ С УЧЕТОМ РЕАЛЬНЫХ УСЛОВИЙ

В четвертой главе приводятся результаты компьютерного моделирования реальных условий выделения речевого сообщения из голосовой смеси микрофонной решеткой. Основные результаты четвертой главы опубликованы в работах автора [А3, А17].

При решении задачи выделения речевого сообщения полезного источника из смеси голосов необходимо рассмотреть ряд проблемных вопросов.

Звуковая волна, распространяясь в замкнутом помещении, многократно отражается от границ рассматриваемого пространства. За счет отражения звуковая волна теряет часть своей энергии, поэтому амплитуда отраженных сигналов уменьшается. Процесс затухания колебаний в помещении носит название реверберации [82]. Интервал между отражениями очень короткий, поэтому человек слышит все отраженные звуки вместе. В реальных условиях в помещении создается диффузное поле – поле, в котором энергия отраженных звуковых волн преобладает над энергией прямого звука. Направление распространения отраженных звуков различно. Если затухание сигналов происходит не слишком быстро, то в любой точке помещения происходит наложение большого числа звуковых волн с различными направлениями волнового вектора. Поле становится изотропным и однородным – средние потоки звуковой энергии по различным направлением равны друг другу и в различных точках помещения средние значения плотности энергии одинаковы. Такой эффект необходимо учитывать в расчетах.

Другим проблемным вопросом является различие энергии голосов говорящих людей. Даже при спокойном темпе и обычной громкости голоса звуковая энергия у двух разных человек будет отличаться. Если человек хочет, чтобы его услышал только один слушатель, он будет разговаривать шепотом, а, значит, на конечном временном интервале звуковая энергия его речевого сообщения будет значительно меньше. В реальных условиях необходимо

учитывать то, что голос полезного источника может быть как очень громким, так и очень тихим.

Когда решение задачи сводится к выделению голоса определенного человека, то необходимо учитывать тот факт, что человек может не стоять на месте, а двигаться по определенной траектории. В реальной акустической обстановке учет нестационарности рассматриваемой обстановки крайне актуален.

Система, позволяющая выделять речевые сообщения из помех, должна проводить обработку акустических сигналов в реальном масштабе времени. Поскольку акустическая система содержит большое число, например, двадцать ненаправленных микрофонов необходимо обеспечить ускорение работы такой системы, а также обеспечить снижение вычислительных затрат.

Таким образом, для моделирования работы предложенного алгоритма в реальных условиях необходимо учитывать следующие факторы:

1) При распространении звуковой волны в замкнутом помещении происходят многократные отражения, что приводит к увеличению общего шума – при моделировании необходимо учитывать эффект реверберации звука;

2) При большом числе одновременно действующих источников звука (толпа говорящих людей) громкость звука различных источников будет отличаться, поэтому необходимо проверить работоспособность алгоритма в случаях, когда уровень сигнала полезного источника в несколько раз меньше уровня окружающего шума.

3) В реальной акустической обстановке диктор (полезный источник) может перемещаться. Необходимо предложить решение задачи для нестационарной обстановки.

4) Практическая ценность современной цифровой обработки состоит в обработке данных в режиме реального времени. Необходимо предложить решение для реализации алгоритма в реальном масштабе времени.

4.1. Учет эффекта реверберации звука в помещении

Известно [83], что «при суммировании хотя бы 5-6 гармонических колебаний со случайными и взаимно независимыми фазами получается стационарный случайный процесс, близкий к нормальному. В случае же суммирования гармонических колебаний не только со случайными начальными фазами, но и с различными частотами получается процесс не только стационарный, но и эргодический». Таким образом, для адекватного моделирования мешающего действия эффекта реверберации помещения нет необходимости имитировать точные значения временных задержек, частотной и угловой зависимости коэффициентов отражения конструкций помещения. Достаточно обеспечить эквивалентность энергетических характеристик суммы отражённых звуковых колебаний в реальном помещении и компьютерной модели.

Поэтому для исследования влияния эффекта реверберации на отношение сигнал/помеха предложена следующая модель: *v*-й отсчет сигнала, регистрируемого *i*-м микрофоном представляется в виде суперпозиции сдвинутых во времени и уменьшенных по амплитуде сигналов одного микрофона без учета дополнительных эффектов:

$$Q_{ex_i}(t_v) = Q_i(t_v) + \sum_{n=1}^{10} \gamma^n \, Q_i(t_v - n\Delta F_s), \tag{42}$$

где γ – коэффициент отражения по амплитуде, Δ – характерное время распространения звуковой волны, прямо пропорциональное расстоянию, проходимому звуковой волной от центра к углу прямоугольного помещения с линейными размерами *a* и *b* определяемое как:

$$\Delta = \frac{\sqrt{a^2 + b^2}}{2V_s}. \tag{43}$$

Характерное время распространения звуковой волны много меньше времени реверберации – важнейшего параметра, характеризующего общую гулкость помещения [84]. Время реверберации – время, за которое уровень звукового давления уменьшается на 60 дБ [85].

Время реверберации зависит от объема помещения и от материалов поглощения стен. Так, например, для исследуемого помещения объемом V=6x4x3 м^3 с поглощающим материалом поверхностей – дерево (усредненный коэффициент поглощения 0,1) – время реверберации по формуле Сэбина [86] будет равно:

$$T = \frac{kV}{A} = \frac{0{,}16 \cdot 72}{0{,}1 \cdot 108} = 1{,}07 \text{ (с)}, \quad (44)$$

где k [с/м] – коэффициент пропорциональности, зависящий от формы помещения, A [м^2] – полное поглощение помещения:

$$A = \alpha_{\text{ср}} S_{\text{пов}}, \quad (45)$$

где $\alpha_{\text{ср}}$ – усредненный коэффициент поглощения, $S_{\text{пов}}$ – суммарная площадь поверхностей помещения.

Характерное время распространения звуковой волны в том же помещении Δ=11 мс.

Модель, описываемая формулой (42), учитывает десятикратное отражение. Дальнейшими отражениями можно пренебречь, так как амплитуда сигналов, полученных после десятого отражения крайне мала. В реальных условиях при одновременном разговоре нескольких дикторов полное отражение (γ=1) невозможно, так как часть звуковой энергии будет поглощаться дикторами [87].

Для оценки энергетических характеристик реверберации будем считать, что энергии многократно отраженных волн суммируются некогерентно. Амплитудные характеристики многократно отраженных волн образуют геометрическую прогрессию:

$$a_n = a_0 \gamma^n, \quad (46)$$

где, a_0 – амплитуда падающей волны, a_{n} – амплитуда волны после n отражений.

Мощность многократно отраженной волны:

$$a_n^2 = a_0^2 \gamma^{2n}. \quad (47)$$

При бесконечном числе отражений, сумма геометрической прогрессии равна:

$$S_\infty = \frac{a_0^2 \gamma^2}{1 - \gamma^2}. \quad (48)$$

При числе отражений, равном 10:

$$S_{10} = a_0^2 \gamma^2 \frac{1-(\gamma^2)^{10}}{1-\gamma^2}. \quad (49)$$

Отношение для γ=0,9:

$$\frac{S_\infty}{S_{10}} = \frac{1}{1-(\gamma^2)^{10}} \cong 1{,}138 \ \cong 0{,}56 \text{ дБ}. \quad (50)$$

То есть суммирование до 10-кратного отражения волн дает отличие от бесконечной суммы не более 0,56 дБ при $\gamma \leq 0{,}9$.

Каждая копия исходного сигнала сдвинута во времени на характерное время распространения в 11 мс. За счет относительно малого сдвига копий структура исходного сигнала (Рисунок 44, верхняя реализация) в сравнении с сигналом с учетом реверберации (Рисунок 44, нижняя реализация) не разрушается.

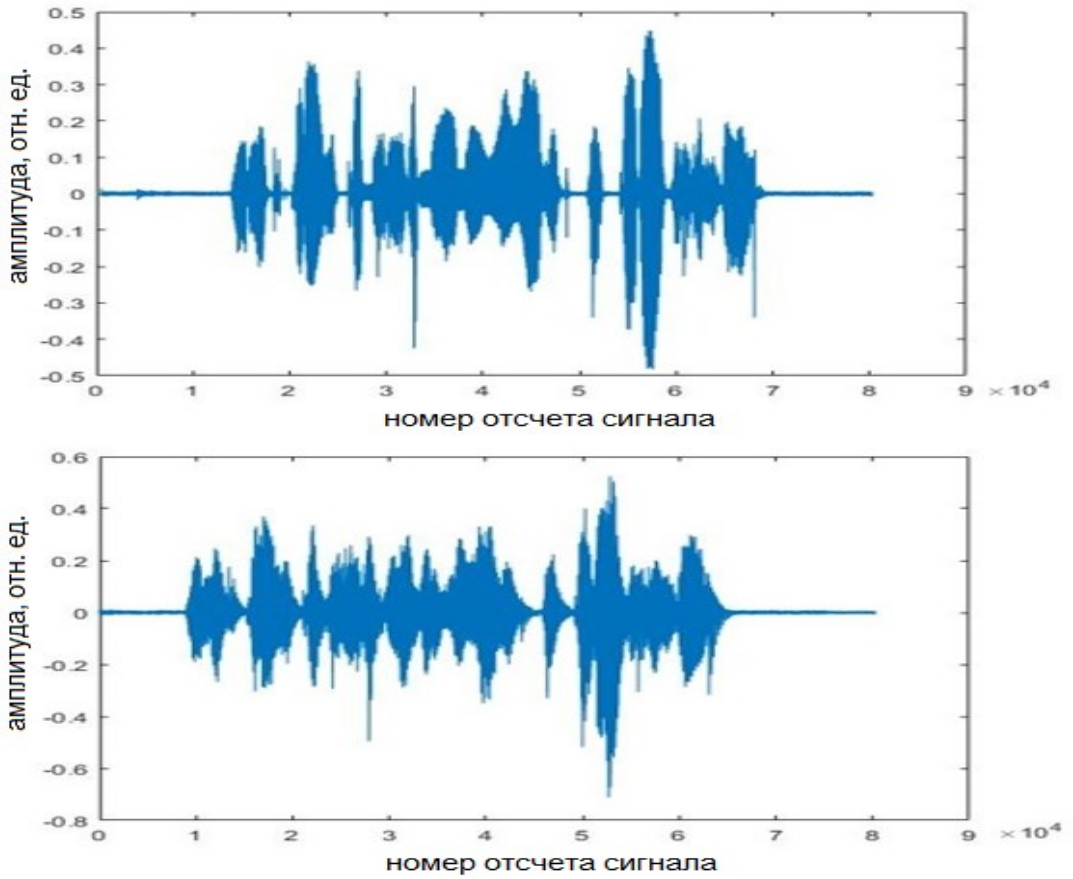


Рисунок 44 – Искажение исходного сигнала за счет учета эффекта реверберации

Выходной сигнал многопозиционной акустической системы представляет собой сумму сигналов всех приемников и выражается как:

$$Q_{ex}(t_v) = \sum_{i=1}^{N} Q_{ex_i}(t_v). \quad (51)$$

Разберем влияние реверберации на отношение сигнал/помеха и на разборчивость речевого сообщения.

На Рисунке 45 представлен график зависимости отношения сигнал/помеха от коэффициента отражения γ для *N*=20 микрофонов. Нижняя линия на графике показывает снижение отношения сигнал/помеха при использовании алгоритма без учета весовых коэффициентов микрофонов. Верхняя – соответствует предложенному алгоритму.

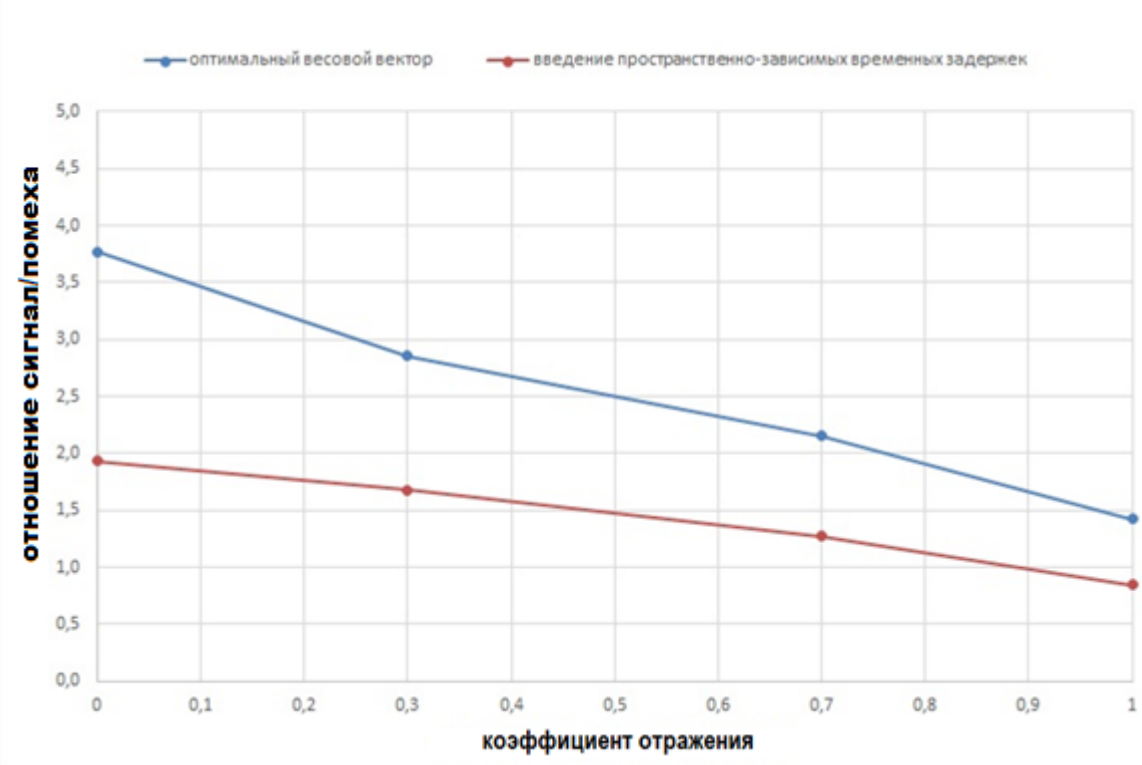


Рисунок 45 – Зависимость отношения сигнал/помеха от коэффициента отражения по амплитуде

На Рисунке 46 показано изменение разборчивости выделенного речевого сообщения в зависимости от коэффициента отражения γ. При выделении речевого сообщения из смеси восьми равномощных голосов акустической системой из двадцати микрофонов, разборчивость речевого сообщения не опускается ниже 95%. Такой уровень разборчивости обеспечивается тем, что мощность

выделенного речевого сообщения для выбранного алгоритма при всех γ всегда больше мощности сторонних акустических помех. Поскольку удовлетворительной считается разборчивость выше 87% [76], Рисунок 46 свидетельствует об устойчивости алгоритма к эффекту реверберации.

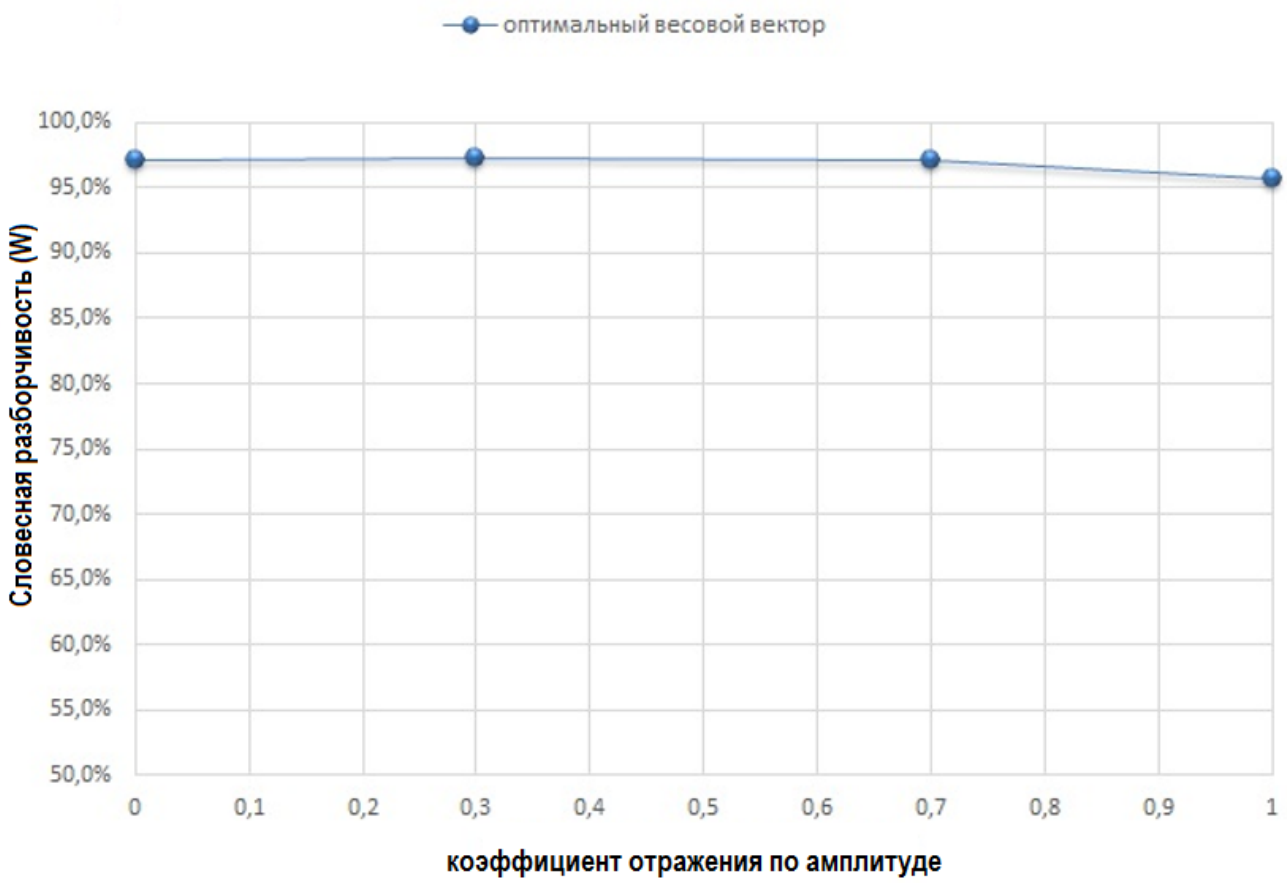


Рисунок 46 – Зависимость словесной разборчивости выделенного речевого сообщения от коэффициента отражения по амплитуде

Данный результат согласуется с результатами, приведенными в работе [88]. Если энергия сигнала полезного источника значительно больше общей энергии интерференции, то обеспечивается уровень разборчивости речевого сообщения, соответствующий пониманию передаваемой речи без затруднений. На Рисунке 47 показана зависимость словесной разборчивости от отношения сигнал/шум для одного, двух и трех сторонних голосов.

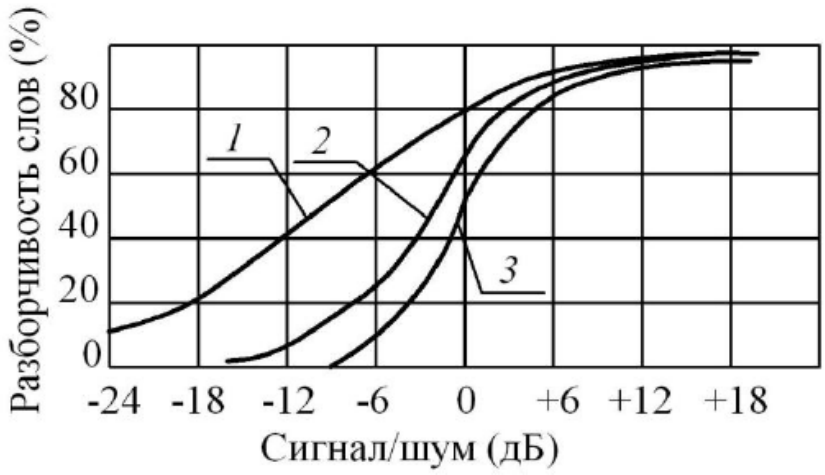


Рисунок 47 – Зависимость словесной разборчивости от отношения сигнал/шум при различном количестве речевых помех [88]

Таким образом, предложенный алгоритм выделения речевых сообщений из помех для заданного числа сторонних источников и определенного числа микрофонов является устойчивым к эффекту реверберации.

4.2. Выделение «тихих» речевых сообщений на фоне громкого разговора

Предложенный в работе алгоритм обработки речевых сообщений позволяет выделить голос определенного человека из смеси голосов. Алгоритм увеличивает отношение сигнал/помеха, тем самым, разборчивость речевого сообщения повышается. Необходимо определить минимальное пороговое значение энергии источника полезного сигнала по отношению к энергии шумовой обстановки, при котором уровень разборчивости выделенного речевого сообщения остается удовлетворительным.

Численный эксперимент был проведен без учета реверберации звука. Мощность семи источников помех считалась одинаковой, а мощность исходного полезного сигнала изменялась.

На Рисунке 48 показана зависимость отношения сигнал/помеха выделенного речевого сообщения от отношения сигнал/помеха исходного полезного сигнала к мощности всех помех.

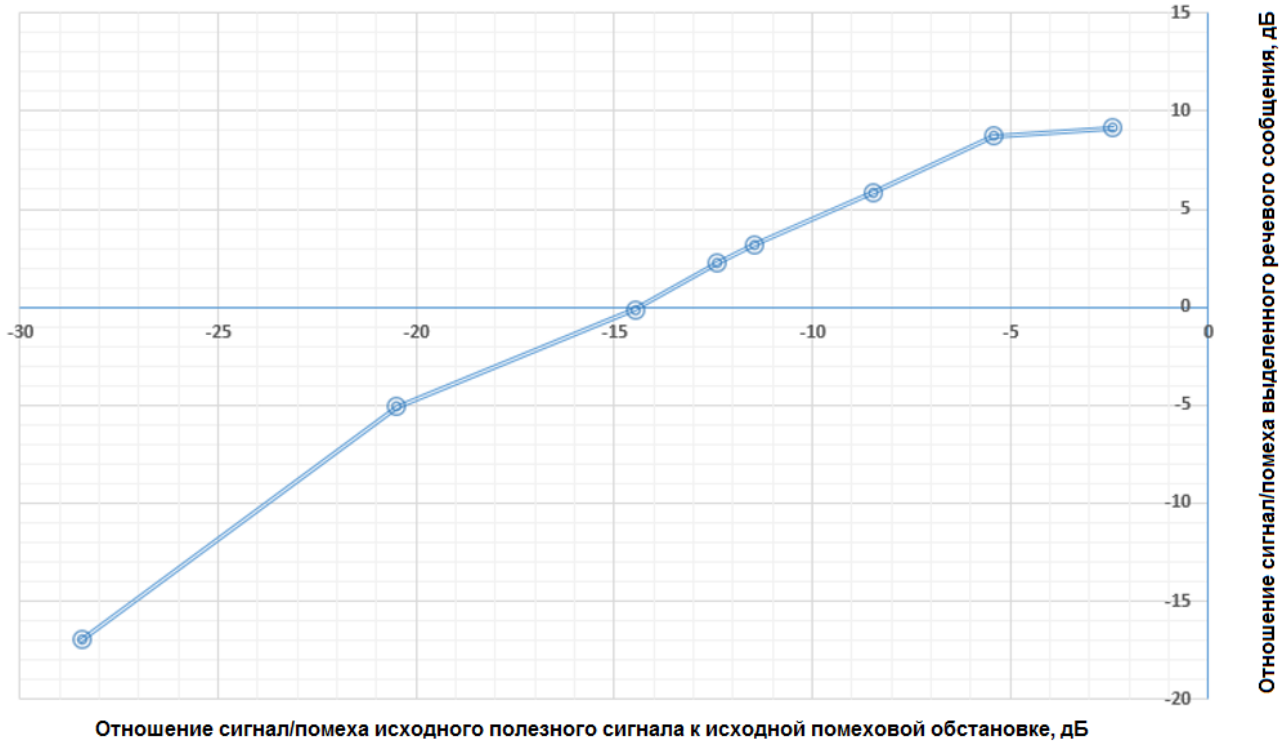


Рисунок 48 – Зависимость отношения сигнал/помеха выделенного речевого сообщения от исходного отношения сигнал/помеха

Отношение сигнал/помеха выделенного речевого сообщения равно единице (0 дБ) при исходном отношении сигнал/помеха в -14,5 дБ (мощность исходного сигнала в 4 раза меньше мощности одной из семи равномощных помех).

Рисунок 49 иллюстрирует зависимость разборчивости выделенного речевого сообщения из помех от исходного отношения сигнал/помеха. Удовлетворительный уровень разборчивости достигается при исходном значении отношения сигнал/помеха не менее -20,5 дБ (энергия исходного сигнала в 16 раз меньше энергии одной из семи равномощных помех). Поэтому предельно допустимое значение исходного отношения сигнал/помеха для корректной работы алгоритма равно -20,5 дБ при рассмотренных условиях.

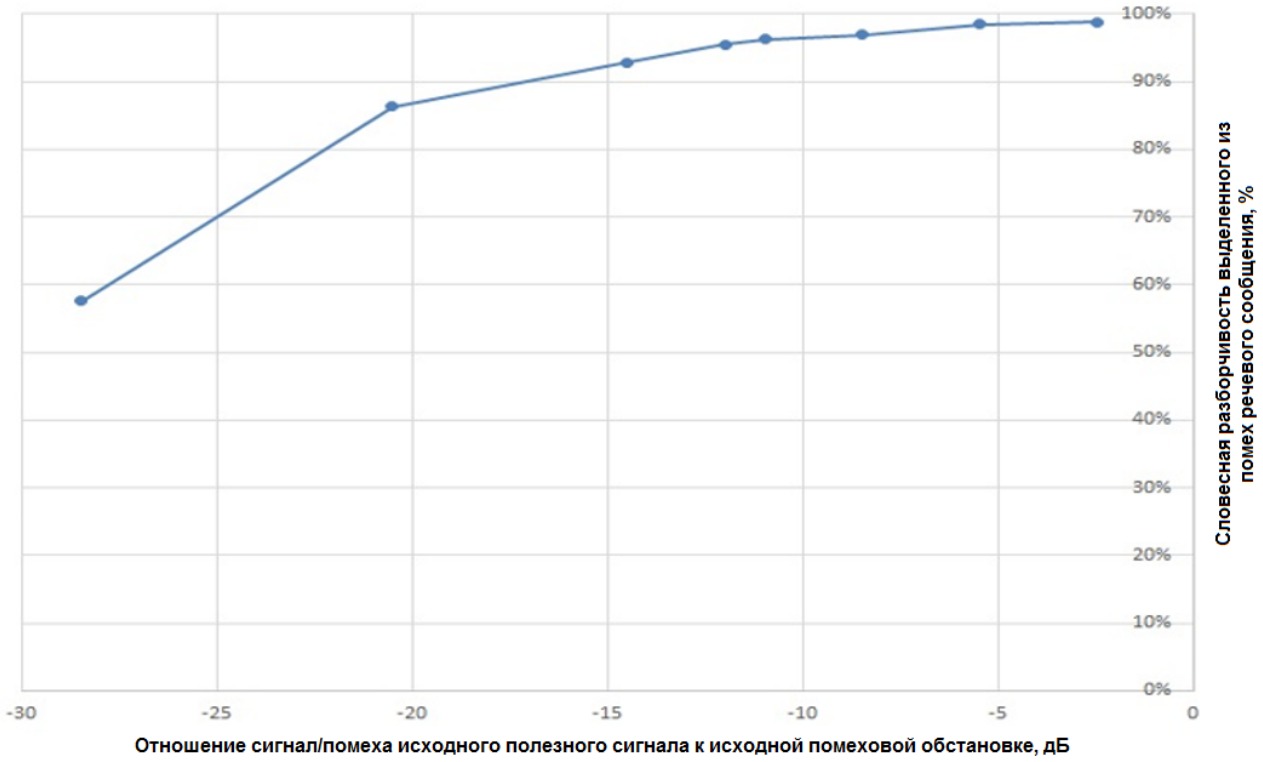


Рисунок 49 – Зависимость словесной разборчивости речевого сообщения от исходного отношения сигнал/помеха

Полученные результаты работы алгоритма по выделению речевых сообщений с разным исходным уровнем мощности дают право сделать вывод о том, что предложенный в работе алгоритм выделения речевых сообщений позволяет выделять «тихие» речевые сигналы на фоне «громкого» разговора.

4.3. Выделение голоса движущегося диктора

При работе алгоритма используется пост-обработка. На временном интервале в 2,4 с происходит определение координат акустических источников и расчет оптимального весового вектора. При средней скорости движения человека в 5 км/ч за время пост-обработки человек смещается на 3,4 м. Отсюда следует важный вывод: алгоритм не способен «следить» за движущимся человеком в отсутствии априорной информации о траектории и координатах его движения.

Для выделения голоса движущегося источника необходимо обладать информацией о траектории его движения. Такую информацию можно получить, например, используя системы видеонаблюдения (Рисунок 50). Так, например, в работе [89] предложен метод определения координат, курса и скорости перемещения объекта по результатам обработки изображения объекта на экране телевизионной камеры, основанный на геометрических соотношениях и пропорциональности размера изображения и расстояния до объекта.

Рисунок 50 – Размеры высоты изображения объекта и поперечного отклонения от оси телекамеры [89]

Также имеются решения по определению траекторий одновременного движения нескольких объектов, которые связаны с использованием системы анализа и обработки видеоданных, полученных с нескольких камер [90].

Для проведения численного эксперимента по выделению голоса движущегося диктора предполагаем, что информация о траектории его движения известна через использование систем видеонаблюдения.

Численный эксперимент был проведен при следующих условиях: на акустической сцене площадью 24 м^2 одновременно разговаривают восемь человек, причем один из них – полезный источник – движется по известной траектории (Рисунок 51), задаваемой уравнениями:

$$\begin{cases} x_{\text{движ}} = -2.5 + 1.4t, \\ y_{\text{движ}} = 1, \end{cases} \quad (52)$$

где x, y и t – безразмерные величины.

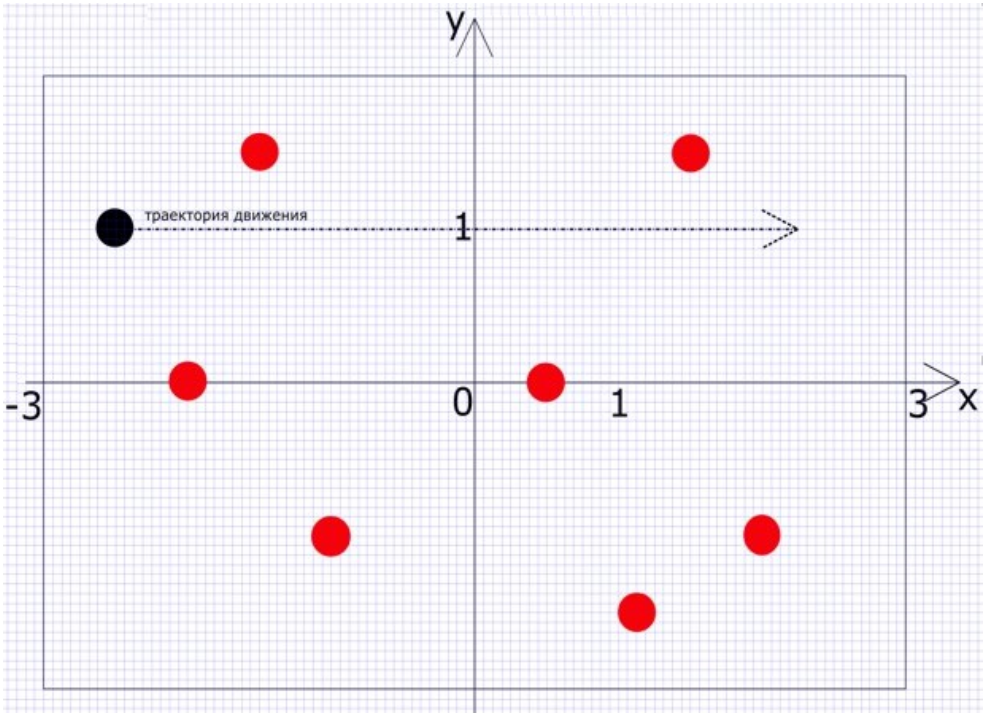


Рисунок 51 – Взаимное расположение источников помех и движущегося источника

Энергия речевых сообщений одинакова. Микрофонной решеткой из двадцати элементов, размещенной по периметру помещения на высоте 1,7 м выделено голосовое сообщение движущегося источника за счет перемещения точки фокусировки системы по известной траектории движения без расчета весовых коэффициентов микрофонов ($w_i = 1$).

На Рисунке 52 показаны реализации исходного сигнала (верхняя реализация) и выделенного из помех голоса движущегося диктора (нижняя реализация). Коэффициент взаимной корреляции двух данных сигналов составил 0,77. Отношение сигнал/помеха выделенного речевого сообщения равно 1,54 и уровень словесной разборчивости – 93,23%.

Для обеспечения понимания выделяемого с помощью рассматриваемого алгоритма речевого сообщения движущегося источника из стационарных в пространстве помех необходимо обладать информацией о траектории его

движения. Совместное применение разработанного алгоритма и системы видеонаблюдения позволяет решить поставленную задачу.

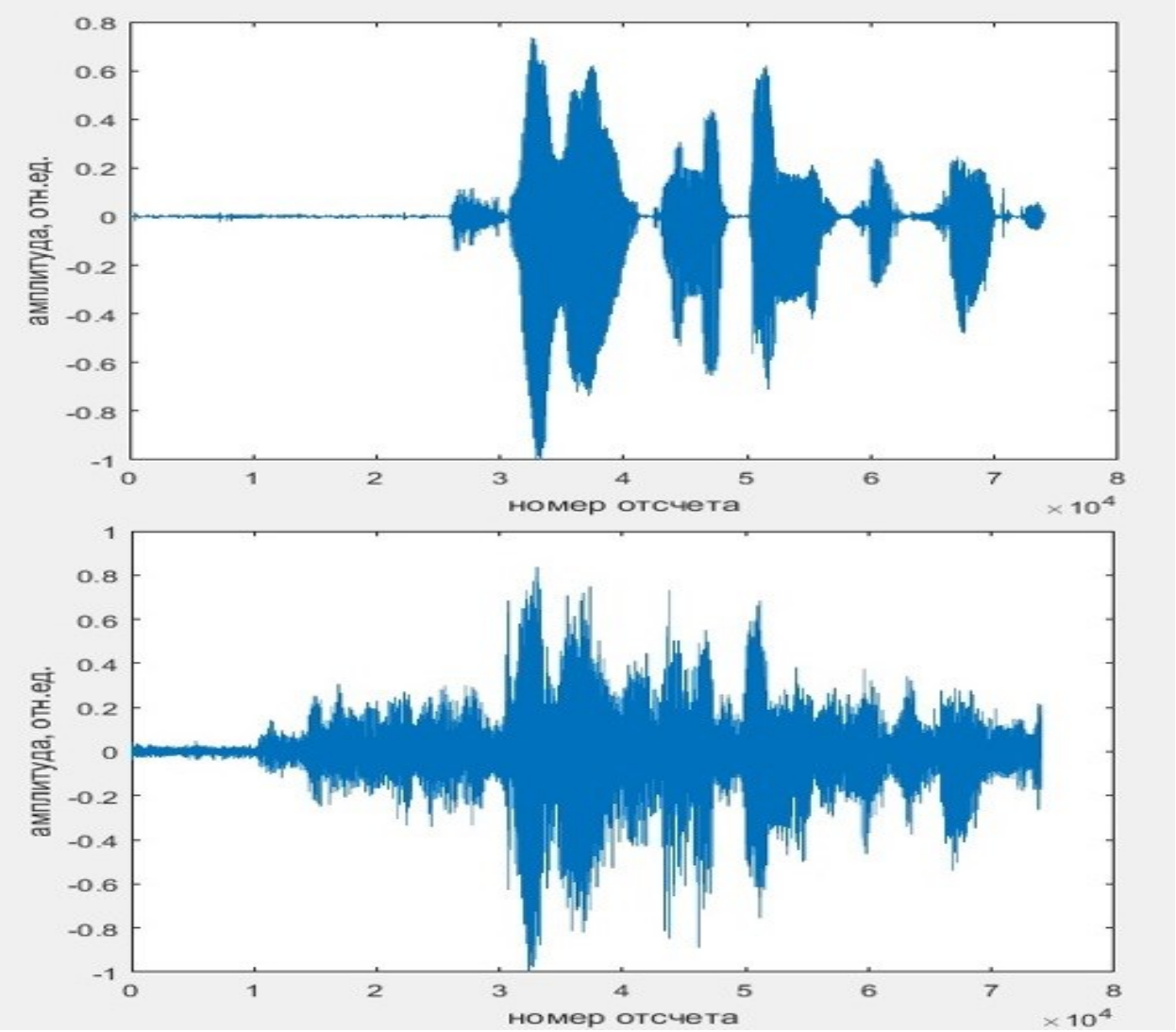


Рисунок 52 – Временная реализация исходного речевого сообщения и выделенного из помех речевого сообщения движущегося источника

4.4. Апробация работы алгоритма в реальном масштабе времени

На обработку аудиосмеси из восьми голосов длительностью 2,4 с последовательный алгоритм обработки затрачивал порядка нескольких десятков секунд (в зависимости от используемого аппаратного обеспечения). Данное время затрачивалось на сканирование пространства, определение координат всех

источников звука в замкнутом пространстве, расчет и дальнейшее применение весовых коэффициентов микрофонного массива. Анализ скорости вычислений показал, что 90% времени алгоритм затрачивает именно на определение координат акустических источников, поскольку он обрабатывает все пространственные координаты акустической сцены с шагом, равным разрешающей способности (28 см для сигналов диапазона 70-7000 Гц).

Поскольку практическую ценность представляет цифровая обработка в реальном масштабе времени (обработка аудиосмеси не более 2,4 секунды), то для ускорения работы алгоритма были применены параллельные вычисления. Возможность применения таких вычислений обусловлена тем, что каждая пространственная координата точки фокусировки обрабатывается микрофонным массивом независимо от другой. Реализация параллельных вычислений была выполнена в среде MATLAB с помощью встроенной поддержки *ParallelComputingToolbox* [91]. Данная поддержка позволяет уменьшать время вычислений за счет запуска схожих заданий на независимых процессорах в одно и то же время [92]. Выигрыш в скорости обработки данных обеспечивается выполнением вычислений в нескольких независимых потоках, а не последовательным выполнением всех инструкций алгоритма в рамках одного потока. Каждый независимый поток программы получил название «работник». На Рисунке 53 показан принцип запуска восьми «работников» системы Matlab при использовании параллельных вычислений на одном компьютере.

После использования встроенной поддержки ParallelComputingToolbox вычисления были произведены на четырех разных конфигурациях оборудования (Таблица 13).

Рисунок 53 – Запуск восьми «работников» на одном компьютере [92]

Таблица 13 – Конфигурации используемого аппаратного обеспечения

№	Процессор	КЭШ L1	КЭШ L2	КЭШ L3	ОЗУ
1	Intel Core i7-4770 CPU 3.40 Ghz	4x32 Kbytes	4x256 Kbytes	10 Mbytes	12 Gb RAM
2	Intel® Xeon® CPU E5-1410 0 @ 2.80 Ghz	4x32 Kbytes	4x256 Kbytes	8 Mbytes	8 Gb RAM
3	Intel® Xeon® CPU E5-4660v3 @ 2.1Ghz 2.1 Ghz (2 процессора)	14x32 Kbytes	14x256 Kbytes	35 Mbytes	32 Gb RAM
4	Intel® Xeon® Gold 6130 CPU @ 2.1 Ghz 2.1 Ghz (2 процессора)	16x32 Kbytes	16x1 Mbytes	22 Mbytes	128 Gb RAM

В Таблице 14 приведены результаты обработки голосовой смеси из восьми сообщений длительностью 2,4 секунды микрофонным массивом для разного количества потоков.

Таблица 14 – Время обработки в зависимости от количества потоков

№ конфигурации оборудования	Время вычисления, сек			
	1 поток	4 потока	8 потоков	16 потоков
1	31,976	8,164	-	-
2	51,328	12,818	-	-
3	52,967	15,415	9,139	5,752
4	25,029	4,442	2,865	2,260

Исходя из данных Таблицы 14, наилучшие результаты достигаются на конфигурации оборудования № 4. Рисунок 54 иллюстрирует зависимость времени вычислений от количества потоков на данном оборудовании.

По данным Рисунка 54 можно сделать вывод, что при использовании не менее 16 потоков алгоритм обрабатывает данные в реальном масштабе времени.

Рисунок 54 – Зависимость времени обработки голосовой смеси от количества потоков на конфигурации №4

Анализируя данные, приведенные в Таблицах 13 и 14, можно сделать заключение о том, что при работе с небольшим количеством потоков процессоры для персональных компьютеров (конфигурация № 1) могут показывать результаты выше специализированных серверных процессоров (конфигурации № 2 и 3), однако данные результаты недостаточны, чтобы алгоритм мог считаться работающим в режиме реального времени. Для данного случая определяющей характеристикой процессора является тактовая частота, а КЭШ-память и количество ядер играют менее значимую роль. Однако при увеличении числа потоков мы наблюдаем, что серверные процессоры начинают показывать лучшие результаты. На первый план выходят количество ядер, которые позволяют реализовать реальную многопоточность, и объем КЭШ-памяти, который позволяет предоставить больший объем данных на обработку без обращения к более медленной оперативной памяти.

Для обработки голосовой смеси восьми источников, регистрируемой массивом из двадцати ненаправленных микрофонов, предложенным алгоритмом в ограниченном пространстве объемом 72 $м^3$ в режиме реального времени потребовался серверный процессор Intel® Xeon® Gold 6130 CPU. При изменении условий численного эксперимента (количество микрофонов, объем исследуемого пространства, количество источников звука) соответственно изменятся требования к аппаратному обеспечению, необходимому для работы алгоритма в режиме реального времени. Но эти требования удовлетворимы на современном уровне развития вычислительной техники, что показал проведенный численный эксперимент.

Подводя итоги, хотелось бы отметить, что использование параллельных вычислений MATLAB и выполнение расчета на современном аппаратном обеспечении позволяет осуществлять многоканальную обработку голосовой смеси во временной области и выделять речевые сообщения с наибольшим отношением сигнал/помеха в режиме реального времени.

ВЫВОДЫ ПО ЧЕТВЕРТОЙ ГЛАВЕ

Предложенный в работе алгоритм обработки акустических сообщений показал устойчивость к эффекту реверберации. При учете эффекта реверберации при одинаковой мощности акустических источников, мощность выделенного речевого сообщения всегда больше суммарной мощности семи сигналов помех. Это позволяет поддерживать понимание передаваемой речи (разборчивость более 87 %).

Рассчитано предельно допустимое значение исходного отношения сигнал/помеха для корректной работы алгоритма: -20,5 дБ (для конкретных условий численного эксперимента). Численные эксперименты показали возможность выделять речевые сообщения предложенным алгоритмом, мощность которых значительно меньше мощности сигналов помех.

При совместном применении алгоритма с системами видеонаблюдения показан хороший результат в выделении голоса движущегося источника за счет априорно известной траектории движения. Выделено речевое сообщение движущегося источника с уровнем словесной разборчивости 93,23%.

Несмотря на вычислительную сложность реализации алгоритма во временной области, использование параллельных вычислений и современного аппаратного обеспечения позволяет выделять речевые сообщения из помех из любой точки пространства наблюдения разработанным алгоритмом в режиме реального времени.

ЗАКЛЮЧЕНИЕ

Исходя из цели работы в ходе диссертационного исследования были решены следующие задачи:

1. Проведен анализ существующих методов разделения акустических сигналов: одноканальные, двухканальные и методы, использующие микрофонные массивы с различной пространственной геометрией. Проведен обзор известных алгоритмов обработки сигналов микрофонными решетками.

2. Разработан оригинальный алгоритм обработки речевого сигнала микрофонной решеткой во временной области, максимизирующий отношение сигнал/помеха на выходе решетки за счет введения точных временных задержек, зависящих от пространственных координат и управления весовыми коэффициентами микрофонов.

3. Численный эксперимент по выделению голоса одного человека из смеси голосов показал устойчивость предложенного алгоритма к эффекту реверберации в помещении, работоспособность при выделении слабых сигналов на фоне более мощных распределенных в пространстве источников помех, возможность выделения голоса движущегося по известной траектории диктора, реализуемость в режиме реального времени.

В диссертационном исследовании определена оптимальная конфигурация микрофонной решетки для выделения речевых сообщений из помех предложенным алгоритмом. Показано, что микрофонная решетка с размещением микрофонов по периметру помещения позволяет выделять речевые сообщения с наибольшим отношением сигнал/помеха.

Определена пространственная разрешающая способность предложенной акустической системы. Для исследуемых сигналов (диапазон частот 70-7000 Гц) разрешающая способность составила 28 см.

Получены количественные оценки эффективности предлагаемого решения для конкретной конфигурации модели помещения с источниками речевых сообщений. Для микрофонной решетки из двадцати микрофонов, выделяющей

одно речевое сообщение из семи равномощных помех предложенным алгоритмом, выигрыш составил 16,5 дБ. Рассчитано предельно допустимое значение исходного отношения сигнал/помеха для корректной работы алгоритма: -20,5 дБ. Выделено речевое сообщение движущегося диктора из семи сигналов источников помех с уровнем словесной разборчивости 93,23%.

На основании полученных результатов можно утверждать, что цель исследования «разработка алгоритма обработки речевого сигнала микрофонной решеткой во временной области, позволяющего выделять речевые сообщения из любой точки пространства наблюдения с максимальным отношением сигнал/помеха, независимо от взаимного расположения целевого диктора и других дикторов, являющихся источниками речевых помех» достигнута в полной мере. Таким образом, показано, что алгоритмы обработки речевого сигнала микрофонной решетки во временной области состоятельны и эффективны, имеют свои преимущества, и могут найти применение в большом числе речевых приложений.

Решения, предложенные в диссертационном исследовании, легко комплексируются с моноуральными алгоритмами разделения речи и могут быть полезны специалистам, занимающимся разработкой акустических систем мониторинга ограниченного пространства.

СПИСОК ЛИТЕРАТУРЫ

1. Cherry, C. Some experiments on the recognition of speech, with one and with two ears / C. Cherry // Journal of the Acoustical Society of America. – 1953. – V. 25. – № 5. – P. 975–979.
2. Cherry, C. On human communication: a review, survey, and a criticism / C. Cherry. – The Technology Press: Massachusetts Institute of Technology, 1957. – 333 p.
3. Arons, B. A review of the cocktail party effect / B. Arons // Journal of the American Voice I/O Society. – 1992. – V. 12. – P. 35–50.
4. Haykin, S. The cocktail party problem / S. Haykin, Z. Chen // Journals of Neural Computation. – 2005. – V. 17. – № 9. – P. 1875–1902.
5. McDermott, J. H. The cocktail party problem / J. H. McDermott // Current Biology. – 2009. – V. 19. – № 22. – P. 1024–1027.
6. Speech Processing in Modern Communication / I. Cohen, J. Benesty, S. Gannot (Eds.) – Springer, 2010. – 360 p.
7. Столбов, М. Б. Применение микрофонных решеток для дистанционного сбора речевой информации / М. Б. Столбов // Научно-технический вестник информационных технологий, механики и оптики. – 2015. – Т. 15. – № 4. – С. 661–675.
8. Microphone arrays: signal processing techniques and applications / M. Brandstein, D. Ward (Eds.). – Springer, 2001. – 398 p.
9. Benesty, J. Microphone array signal processing / J. Benesty, J. Chen, Y. Huang. – Springer, 2008. – 245 p.
10. Springer handbook of speech processing / J. Benesti, J. M. Sondhi, Y. Huang (Eds.). – Springer, 2008. – 1159 p.
11. Tashev, I. Improving meetings with microphone array algorithms [Электронный ресурс] / I. Tashev // Microsoft. – Режим доступа: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/Tashev_MeetingsMicrophoneArray_NIPS_03.pdf (дата обращения: 12.12.2019).

12. Jaeckel, O. Transient noise source localization / O. Jaeckel, G. Heilmann // Euronoise. – 2006. – P. 1–6.
13. How to solve the cocktail party problem [Электронный ресурс] // Future learning. – Режим доступа: https://futurelearning.ai/how-to-solve-the-cocktail-party-problem/ (дата обращения: 12.12.2019).
14. Blauert, J. Spatial hearing: the psychophysics of human sound localization / J. Blauert. – Cambridge: MIT Press, 1983. – 427 p.
15. Bregman, A. S. Auditory scene analysis: the perceptual organization of sound / A. S. Bregman. – Cambridge: MIT Press, 1990. – 854 p.
16. Гусев, А. Н. Общая психология: в 7 т. / А. Н. Гусев. – М.: Академия, 2007. – т.2. – 416 с.
17. Marr, D. Vision / D. Marr. – W. H. Freeman and Company, 1982. – 415 p.
18. Chait, M. Auditory scene analysis [Электронный ресурс] / M. Chait // UCL psychology and language sciences. – Режим доступа: https://www.phon.ucl.ac.uk/courses/spsci/AUDL4007/Scene_analysis.pdf (дата обращения: 12.12.2019).
19. Wang, D. L. Computational auditory scene analysis: principles, algorithms, and applications / D. L. Wang, G. J. Brown. – IEEE Press: Wiley, 2006. – 395 p.
20. Wang, D. Computational auditory scene analysis and its potential application to hearing aids [Электронный ресурс] / D. Wang // Slideserve. – Режим доступа: https://www.slideserve.com/aradia/computational-auditory-scene-analysis-and-its-potential-application-to-hearing-aids (дата обращения: 12.12.2019).
21. Hu, G. Monaural speech segregation based on pitch tracking and amplitude modulation / G. Hu, D. Wang // IEEE Transactions on Neural Networks. – 2004. – V. 15. – № 5. – P. 1135–1150.
22. Wang, D. Monaural and binaural speech separation [Электронный ресурс] / D. Wang // The Laboratory for the Recognition and Organization of Speech and Audio. – Режим доступа: https://labrosa.ee.columbia.edu/Montreal2004/talks/deliang1.pdf (дата обращения: 12.12.2019).

23. Gu, L. Single-channel speech separation based on modulation frequency / L. Gu, R. M. Stern // IEEE ICASSP. – 2008. DOI: 10.1109/ICASSP.2008.4517537
24. Doclo, S. Binaural speech enhancement and cue preservation algorithms [Электронный ресурс] / S. Doclo // UCL psychology and language sciences. – Режим доступа: https://www.phon.ucl.ac.uk/events/elobes2019/Elobes2019doclo.pdf (дата обращения: 12.12.2019).
25. Lyon, F. A computational model of binaural localization and separation / F. Lyon // ICASSP. – 1983. – P. 1148–1151.
26. Yost, W. A. Auditory Perception of Sound Sources / W. A. Yost, A. N. Popper, R. R. Fay. – Springer Handbook of Auditory Research, 2008. – V. 29. – 337 p.
27. Алдошина, И. Основы психоакустики [Электронный ресурс] / И. Алдошина // Digital music academy. – Режим доступа: http://www.digitalmusicacademy.ru/sites/default/files/content/aldoshina-psihoakustika.pdf (дата обращения: 12.12.2019).
28. Hidri, A. About multichannel speech signal extraction and separation techniques / A. Hidri, S. Meddeb, H. Amiri // Journal of Signal and Information Processing. – 2012. – V. 3. – №2. – P. 238–247.
29. Divenyi, P. Speech separation by humans and machines / P. Divenyi // Springer US, 2005. – 319 p.
30. Wang, D.L. Time–Frequency masking for speech separation and its potential for hearing aid design / D. L. Wang // Trends in Amplification. – 2008. – V. 12. – № 4. – P. 332–353.
31. Yilmaz, O. Blind separation of speech mixtures via time-frequency masking / O. Yilmaz, S. Rickard // IEEE Transactions on Signal Processing. – 2004. –V. 52. – №7. – P. 1830–1847.
32. Makino, S. Blind Speech Separation / S. Makino, T.-W. Lee, H. Sawada. – Springer, 2007. – 438 p.
33. Подвительский, А. Н. Исследование методов слепого разделения сигналов [Электронный ресурс] / А. Н. Подвительский // Электронная библиотека БГУ.

– Режим доступа: http://elib.bsu.by/bitstream/123456789/7403/1/44.pdf (дата обращения: 12.12.2019).

34. Comon, P. Handbook of blind source separation: Independent Component Analysis and Applications / P. Comon, C. Jutten. – Academic Press, 2010. – 840 p.
35. Топников, А. И. Введение в слепое разделение речевых сигналов: практикум для студентов, обучающихся по направлению Радиотехника / А. И. Топников. – Ярославль: ЯрГУ, 2015. – 44 с.
36. Stone, J. V. Independent component analysis. A tutorial introduction / J.V. Stone. – Cambridge: MIT Press, 2004. – 206 p.
37. Hyvarinen, A. Independent component analysis: algorithms and applications / A. Hyvarinen, E. Oja // Neural Networks. – 2000. – V. 13. – P. 411–430.
38. Chatterjee, S. Dimensionality Reduction — PCA, ICA and Manifold learning [Электронный ресурс] / S. Chatterjee // Towards data science. – Режим доступа: https://towardsdatascience.com/dimensionality-reduction-pca-ica-and-manifold-learning-65393010253e (дата обращения: 12.12.2019).
39. A regularized weighted smoothed L_0 norm minimization method for underdetermined blind source separation [Электронный ресурс] // L. Wang, X. Yin, H. Yue, J. Xiang // Mdpi. – Режим доступа: https://www.mdpi.com/1424-8220/18/12/4260 (дата обращения: 12.12.2019).
40. Gannot, S. Introduction to distributed speech enhancement algorithms for ad hoc microphone arrays and wireless acoustic sensor networks [Электронный ресурс] / S. Gannot, A. Bertrand // EUSIPCO 2013. – Режим доступа: https://hobbydocbox.com/Radio/70876737-Introduction-to-distributed-speech-enhancement-algorithms-for-ad-hoc-microphone-arrays-and-wireless-acoustic-sensor-networks.html (дата обращения: 12.12.2019).
41. Seltzer, M. L. Calibration of microphone arrays for improved speech recognition [Электронный ресурс] / M. L. Seltzer, B. Raj // School of computer science. – Режим доступа: https://www.cs.cmu.edu/~mseltzer/talks/mls_eurosp01_talk.pdf (дата обращения: 12.12.2019).

42. Woelfel, M. Distant speech recognition / M. Woelfel, J. McDonough. – Wiley, 2009. – 600 p.
43. McCowan, I. Microphone arrays: a tutorial [Электронный ресурс] / I. McCowan // Режим доступа: http://www.aplu.ch/home/download/microphone_array.pdf (дата обращения: 12.12.2019).
44. Microphone Array Beamforming [Электронный ресурс] // Invensense. – Режим доступа: http://www.invensense.com/wp-content/uploads/2015/02/Microphone-Array-Beamforming.pdf (дата обращения: 12.12.2019).
45. Beamforming Techniques for Multichannel audio Signal Separation [Электронный ресурс] / A. Hidri, S. Meddeb, A. Abdulqadir, H. Amiri // Режим доступа: https://arxiv.org/ftp/arxiv/papers/1212/1212.6080.pdf (дата обращения: 12.12.2019).
46. Bourgeois, J. Time-domain Beamforming and Blind Source Separation / J. Bourgeois, W. Minker. – Springer, 2009. – 228 p.
47. Столбов, М. Б. Исследование двухканального алгоритма MVDR для выделения речи из когерентного шума / М. Б. Столбов, Ч. Т. Куан // Научно-технический вестник информационных технологий, механики и оптики. – 2019. – Т. 19. – № 1. – С. 180–183.
48. Ермолаев, В. Т. Современные методы пространственной обработки сигналов в информационных системах с антенными решетками: учебно-методический материал по программе повышения квалификации / В. Т. Ермолаев, А. Г. Флаксман. – Нижний Новгород, 2007. – 99 с.
49. Design and calibration of large microphone arrays for robotic applications / F. Perrodin, J. Nikolic, J. Busset, R. Siegwart // IEEE International Conference on Intelligent Robots and Systems, 2012. – P. 4596–4601.
50. Kodrasi, I. Microphone position optimization for planar superdirective beamforming / I. Kodrasi, T. Rohdenburg, S. Doclo // ICASSP, 2011. – P. 109–112.
51. Hodjat, F. Nonuniformly spaced linear and planar array antennas for sidelobe reduction / F. Hodjat, S. A. Hovahessian // IEEE Transactions on Antennas and Propagation. – 1978. – V. 26. – № 2. – P. 198–204.

52. Aarabi, P. The Fusion of Distributed Microphone Arrays for Sound Localization / P. Aarabi // EURASIP Journal on Applied Signal Processing. – 2003. – V.4. – P. 338–347.
53. Linear and Circular Microphone Array for Remote Surveillance: Simulated Performance Analysis [Электронный ресурс] / A. AlShehhi, M. L. Hammadih, M. S. Zitouni, S. AlKindi, N. Ali, L. Weruaga // Audio Engineering Society. – Режим доступа: https://arxiv.org/pdf/1703.02318.pdf (дата обращения: 12.12.2019).
54. Microphone array geometry for two dimensional broadband sound field recording [Электронный ресурс] / W-H. Liao, Y. Mitsufuji, K. Osako, K. Ohkuri // Режим доступа: http://www.aes.org/e-lib/browse.cfm?elib=19808 (дата обращения: 12.12.2019).
55. Meyer, J. A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield / J. Meyer, G. W. Elko // ICASSP. – 2002. – P. 1781–1784.
56. Meyer, J. Spherical microphone arrays for 3D sound recording / J. Meyer, G. W. Elko // in Audio Signal Processing for Next-Generation Multimedia Communication Systems. – Kluwer Academic Publishers. – 2004. – P. 67–89.
57. A spherical microphone array system for traffic scene analysis / Z. Li, R. Duraiswami, E. Grassi, L. S. Davis // IEEE Conference on Intelligent Transportation Systems. – 2004. DOI: 10.1109/ITSC.2004.1398921
58. Haykin, S. Handbook on array processing and sensor networks // S. Haykin, K. J. R. Liu. – IEEE Press: Wiley, 2009. – 924 p.
59. Prandi, G. Acoustic source localization by fusing distributed microphone arrays measurements / G. Prandi, G. Valenzise, M. Tagliasacchi (Eds.) // IEEE European Signal Processing Conference. – 2008. – P. 1–5.
60. Borra, F. Localization of acoustic sources in the ray space for distributed microphone sensors / F. Borra, F. Antonacci, A. Sarti, S. Tubaro // IEEE WASPAA. – 2017. DOI: 10.1109/WASPAA.2017.8170017

61. Spriet, A. Robustness analysis of multichannel Wiener filtering and generalized sidelobe cancellation for multimicrophone noise reduction in hearing aid applications / A. Spriet, M. Moonen, J. Wouters // IEEE Transactions on Speech and Audio Processing. – 2005. – V. 13. – P. 487-503.
62. Acoustic beamforming for hearing aid applications / S. Doclo, S. Gannot, M. Moonen, A. Spriet. In Handbook on array processing and sensor networks. – Wiley, 2010. – P. 269-302.
63. Markovich-Golan, S. A weighted multichannel Wiener filter for multiple sources scenarios / S. Markovich-Golan, S. Gannot, I. Cohen // IEEE Convention of Electrical and Electronics Engineers in Israel. – 2012. DOI: 10.1109/EEEI.2012.6376958
64. Lawin-Ore, C. Reference microphone selection for mwf-based noise reduction using distributed microphone arrays / C. Lawin-Ore, S. Doclo // ITG Symposium on Speech Communication. – 2012. – P. 31–34.
65. Stenzel, S. A multichannel Wiener filter with partial equalization for distributed microphones / S. Stenzel, C. Lawin-Ore, J. Freudenberger, S. Doclo // IEEE WASPAA. – 2013. DOI: 10.1109/WASPAA.2013.6701874
66. Kenoshita, K. Blind source separation using spatially distributed microphones based on microphone-location dependent source activities / K. Kenoshita, M. Souden, T. Nakatani // Interspeech. – 2013. – P. 822-826.
67. Plinge, A. Multi-speaker tracking using multiple distributed microphone arrays / A. Plinge, G. A. Fink // IEEE ICASSP. – 2014. DOI: 10.1109/ICASSP.2014.6853669
68. Demontis, H. 3D Identification of Acoustic Sources in Rooms Using a Large-Scale Microphone Array / H. Demontis, F. Olivier, J. Marchal // IEEE IWAENC. – 2018. DOI: 10.1109/IWAENC.2018.8521264
69. Araki, S. Hybrid Approach for Multichannel Source Separation Combining Time Frequency Mask with Multi-Channel Wiener Filter / S. Araki, T. Nakatani // ICASSP, 2011. – P. 225–228.

70. Wang L. Target Speech Extraction in Cocktail Party by Combining Beamforming and Blind Source Separation / L. Wang, H. Ding, F. Yin // Journal Acoustics Australia. – 2011. – V. 39. – № 2. – P. 64– 68.
71. Ермолаев, В. Т. Методы обработки сигналов в адаптивных антенных решетках и компенсаторах помехи: учебное пособие / В. Т. Ермолаев, А. Г. Флаксман. – Нижегородский госуниверситет, 2015. – 194 с.
72. Монзинго, Р. А. Адаптивные антенные решетки: введение в теорию / Р. А. Монзинго, Т. У. Миллер; Пер. с англ. под ред. В. А. Лексаченко. – М.: Радио и связь, 1986. – 448 с.
73. ГОСТ Р 51061-97 Системы низкоскоростной передачи речи по цифровым каналам. – М.: Госстандарт России, 1997. – 24 с.
74. Продеус, А. Н. Сравнительный анализ некоторых методов оценки разборчивости речи / А. В. Гавриленко, В. С. Дидковский, А. Н. Продеус // Акустический симпозиум: Консонанс-2007, 2007. – С.54–65.
75. Железняк, В. К. Некоторые методические подходы к оценке эффективности защиты речевой информации / В. К. Железняк, Ю. К. Макаров, А. А. Хорев // Спецтехника. – 2000. – № 4. – С. 39.
76. Покровский, Н. Б. Расчёт и измерение разборчивости речи / Н. Б. Покровский. – М.: Гос. изд-во литературы по вопросам связи и радио, 1962. – 392 с.
77. Кропотов, Ю. А. Оценивание эффективности телекоммуникаций аудиообмена в условиях внешних акустических помех / Ю. А. Кропотов, А. А. Белов, А. А. Колпаков, А. Ю. Проскуряков // Системы управления, связи и безопасности. – 2019. – № 1. – С.193–203.
78. ГОСТ 16600-72 Передача речи по трактам радиотелефонной связи. Требования к разборчивости речи и методы артикуляционных измерений. – Межгосударственный стандарт, 1974. – 74 с.
79. ГОСТ Р 50840-95 Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости. – М.: Госстандарт России, 1997. – 234 с.
80. Михайлов, В. Г. Стандартизация измерений качества передачи / В. Г. Михайлов // Акустический журнал. – 2006. – Т. 52. – № 2. – С. 243–250.

81. Рабинер, Л. Р. Цифровая обработка речевых сигналов / Л. Р. Рабинер, Р. В. Шафер; Пер. с англ. под ред. М. В. Назарова, Ю. Н. Прохорова. – М.: Радио и связь, 1981. – 496 с.
82. Акустика: справочник / под ред. М. А. Сапожкова. – М.: Радио и связь, 1989. – 336 с.
83. Гоноровский, И. С. Радиотехнические цепи и сигналы: учебник для вузов / И. С. Гоноровский. – М.: Радио и связь, 1986. – 512 с.
84. Макриненко, Л. А. Акустика помещений общественных зданий / Л. А. Макриненко. – М.: Стройиздат, 1986. – 173 с.
85. Емельянов, Е. Д. Звукофикация театров и концертных залов / Е. Д. Емелянов. – М.: Искусство, 1989. – 272 с.
86. Измерения в акустике зданий [Электронный ресурс] // Режим доступа: http://asm-tm.ru/wp-content/uploads/2014/05/Measurements_in_Building_Acoustics.pdf (дата обращения: 16.12.2019).
87. Фурдуев, В. В. Поглощение звука публикой (методы и результаты исследований). Обзор / В. В. Фурдуев // Акустический журнал. – Т. 16. – № 3. – С. 332–344.
88. Теоретические основы защиты информации от утечки по акустическим каналам: учеб. пособие / Ю. А. Гатчин, А. П. Карпик, К. О. Ткачев, К. Н. Чиков, В. Б. Шлишевский. – Новосибирск: СГГА, 2008. – 194 с.
89. Мартынова, Л. А. Определение координат и параметров движения объекта на основе обработки изображений / Л. А. Мартынова, А. В. Корякин, К. В. Ланцов, В. В. Ланцов // Компьютерная оптика. – 2012. – Т. 36. – № 2. – С. 266–273.
90. Коробков, А. Отслеживание объектов в видеопотоке. Методы построения траекторий / А. Коробков // Системы безопасности. – 2014. – №3. – С. 98–100.
91. Parallel Computing Toolbox [Электронный ресурс] // Mathworks. – Режим доступа: https://www.mathworks.com/products/parallel-computing.html (дата обращения: 17.12.2019).

92. Туревский, А. Параллельные вычисления в MATLAB [Электронный ресурс] / А. Туревский // Matlab. – Режим доступа: https://matlab.ru/news/Introduction%20to%20Parallel%20Computing%20with%20MATLAB_final_russian.pdf (дата обращения: 17.12.2019).

Список работ, опубликованных автором по теме диссертации

Статьи, опубликованные в журналах, включенных в перечень ВАК:

A1. Миронов, Н. А. Выделение речевого сообщения из помех, вносимых сторонними распределёнными источниками / В. А. Канаков, Н. А. Миронов // Известия высших учебных заведений. Радиофизика. – 2017. – Т. 60. – № 3. – С. 281–287.

Mironov, N. A. Speech-Message Extraction from Interference Introduced by External Distributed Sources / V. A. Kanakov, N. A. Mironov // Radiophysics and Quantum Electronics. – 2017. – V. 60. – № 3. – P. 252–257.

A2. Миронов, Н. А. Пространственная обработка широкополосных сигналов на примере речевых сообщений / В. А. Канаков, Н. А. Миронов // Известия высших учебных заведений. Радиофизика. – 2018. – Т. 61. – № 1. – С. 85–91.

Mironov, N. A. Spatial Processing of Broadband Signals Using Speech Messages as Examples // V. A. Kanakov, N. A. Mironov // Radiophysics and Quantum Electronics. – 2018. – V. 61. – № 1. – P. 77–82.

A3. Миронов, Н. А. Моделирование реальных условий выделения речевого сообщения из голосовой смеси / Н. А. Миронов // Радиотехника. – 2019. – Т. 83. – № 6 (7). – С. 81–86.

A4. Миронов, Н. А. Выбор наилучшей конфигурации микрофонной решетки для выделения речевых сообщений из помех / В. А. Канаков, Н. А. Миронов // Радиотехника. – 2019. – Т. 83. – № 8 (11). – С. 13–19.

Статьи в сборниках трудов конференций и материалы докладов:

A5. Миронов, Н. А. О выделении акустического сигнала на фоне интенсивных пространственно-распределенных помех / В. А. Канаков, Н. А. Миронов // Современное состояние естественных и технических наук. – 2014. – № XVII. – С. 8–11.

A6. Миронов, Н. А. Оценка пространственной разрешающей способности многопозиционной акустической системы / Н. А. Миронов // Труды XVIII научной конференции по радиофизике. – 2014. – С. 238–239.

A7. Миронов, Н. А. Оценка разборчивости речевого сообщения в помеховой обстановке в зависимости от количества приемных устройств / В. А. Канаков, Н. А. Миронов // Тезисы докладов XXI международной научно-технической конференции «Информационные системы и технологии». – 2015. – С.31.

A8. Миронов, Н. А. Выделение речевого сообщения из помех от распределенных в пространстве источников / В. А. Канаков, Н. А. Миронов/ Труды XIX научной конференции по радиофизике. – 2015. – С. 141–142.

A9. Миронов, Н. А. Пространственная фильтрация речевого сигнала на фоне интенсивных помех / Н. А. Миронов // Доклады 17-й Международной конференции «Цифровая обработка сигналов и ее применение». – 2015. – С. 183–186.

A10. Миронов, Н. А. Локализация источников звуковых сигналов в ограниченном пространстве / Н. А. Миронов // Сборник научных трудов по материалам Международной научно-практической конференции «Наука 21 века: открытия, инновации, технологии» – 2016. – С. 100–102.

A11. Миронов, Н. А. О способе фильтрации речевого сигнала в помеховой обстановке / Н. А. Миронов // Тезисы докладов XXII международной научно-технической конференции «Информационные системы и технологии». – 2016. – С.71.

A12. Миронов, Н. А. Реализация параллельной схемы наблюдения за большим числом источников широкополосных сигналов / Н. А. Миронов // Материалы

докладов XXI Нижегородской сессии молодых ученых (естественные, математические науки). – 2016. – С. 29–32.

А13. Миронов, Н. А. Поиск оптимального весового вектора криволинейной антенной решетки / Н. А. Миронов // Сборник научных трудов по материалам XV международной научно-практической конференции «Перспективы развития науки и образования». – 2017. – С. 146–149.

А14. Миронов, Н. А. Применение методов пространственной обработки сигналов для выделения речевых сообщений из помех / Н. А. Миронов // Тезисы докладов XXIII международной научно-технической конференции «Информационные системы и технологии». – 2017. – С.1176–1179.

А15. Миронов, Н. А. Применение адаптивных алгоритмов пространственной обработки широкополосных сигналов в отсутствие априорной информации о помеховой обстановке / В. А. Канаков, Н. А. Миронов // Труды XXI научной конференции по радиофизике. – 2017. – С.232–235.

А16. Миронов, Н. А. Выделение речевых сообщений из помех многопозиционной системой микрофонов / Н. А. Миронов // Материалы докладов XXIII Нижегородской сессии молодых ученых (технические, естественные, математические науки). – 2018. – С.94–97.

А17. Миронов, Н. А. Выделение речевого сообщения из помех движущегося источника / Н. А. Миронов // Труды XXII научной конференции по радиофизике. – 2018. – С.282–284.