

На правах рукописи

УДК xxx.xxx

Поляков Игорь Владимирович

**МОДЕЛИ И МЕТОДЫ МОНАУРАЛЬНОЙ ЛОКАЛИЗАЦИИ
НАПРАВЛЕНИЯ НА ИСТОЧНИК ЗВУКА**

Специальность 09.04.01 —

«Информатика и вычислительная техника»

Выпускная квалификационная работа

Научный руководитель:

Гай Василий Евгеньевич

Введение

Актуальность работы

Способность к звуковому поиску, т.е. определение направления на источник звука – важная вещь для биологических организмов, ведь звук может выступать и сигналом опасности и использоваться для поиска жертвы. К тому же, локализация звука имеет множество инженерных применений, начиная с определения местоположения говорящего, заканчивая автоматическими решениями куда повернуть направленный микрофон, прожектор или камеру.

В настоящее время актуальна проблема разработки систем оценки направления на источник звука. Такие системы могут использоваться как в масштабах определения местонахождения крупных объектов – например, самолетов, подводных лодок, так и в качестве сенсоров для различных устройств, роботов, охранных систем.

Во то время, когда задача поиска направления на источник звука приобрела свою актуальность, основное внимание было уделено бинауральному методу поиска, как наиболее очевидному и простому. Этот метод основан на нахождении разницы фаз и величины разности амплитуд между записями с двух приемников звука, находящихся в одной плоскости. Однако, не смотря на работоспособность, такой способ имеет ряд недостатков, среди которых малая область поиска, довольно грубый результат и относительно большие габариты установки.

Использование бинаурального метода поиска обуславливалось тем, что человеческий слух и способность воспринимать направление источника звука долгое время присваивались использованию сразу двух ушей. Однако, когда в некоторых экспериментах бинауральный принцип стал недостаточен, начала серьезно изучаться роль ушной раковины в локализации звука [1]. Некоторые эксперименты проводились с источниками звука, лежащими непосредственно на медиальной вертикальной плоскости и не отклоняющимися куда-либо по горизонтали. Невозможность использования в таких случаях бинаурального метода послужила толчком к усовершенствованию приемника-обработчика и поиску альтернативных методов решения проблемы.

Была сформулирована задача поиска с использованием одного приемника звука. Существующие решения данной задачи, основанные на монауральном принципе, рассмотрены и приняты во внимание [2] [3]. Многие из этих решений обрабатывают сигнал на уровне отсчетов. С другой стороны, известны факты, утверждающие, что, к примеру, механизмы зрительного восприятия человека целостны, при грубо-точном анализе сенсорных данных зрительной системой. В теории активного восприятия (ТАВ) описан метод грубо-точного анализа, который используется при распознавании изображений [4]. Соответственно, предполагая, что в слуховой системе работают схожие механизмы, данный метод можно применить и к анализу звуковых сигналов. Предлагаемая реализация этапов системы распознавания рассмотрена с точки зрения ТАВ.

Цель и задачи работы

Целью данной работы является разработка моделей и методов монауральной локализации направления на источник звука.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Проведение анализа известных подходов локализации источника звука, основанных на применении монаурального принципа с выделением их достоинств и недостатков;
2. Разработка системы локализация источника звука, построение её модели и решение с помощью необходимых методов;
3. Разработка программной реализации модели системы;
4. Тестирование, включающее в себя проведение экспериментов с реализованной моделью, поиск алгоритма ее работы, дающего наилучший результат, а так же сравнение конечных результатов с аналогами.

Методы исследования

Для решения поставленных задач использовались методы системного анализа, применялось совмещение монаурального принципа звуковой локации с теорией активного восприятия.

На защиту выносятся:

Научная новизна

1. Разработана информационная модели системы монауральной локализации направления на источник звука, отличающаяся использованием в качестве признакового описания сигнала полных групп и позволяющая сократить длительность сигнала для определения направления на источник звука.
2. Алгоритм вычисления полных и замкнутых групп по спектру звукового сигнала с применением ускорения формирования системы признаков с помощью использования гистограммы вероятности переходов.

Практическая значимость

Создание программно-аппаратного комплекса, решающего задачу поиска направления на источник звука. Предложенные методы и результаты проведенных исследований могут быть использованы для развития систем в данной сфере или систем других сфер, работающих по схожим принципам.

Инв. № подл.	Подп. и дата					Лист 5
	Взам. инв. №					
	Инв. № дубл.					
	Подп. и дата					
	Инв. № подл.					
Ли	Изм.	№ докум.	Подп.	Дат	ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)	

Апробация работы

Выпускная квалификационная работа и отдельные её части докладывались и обсуждались на международной научной конференции «Молодежь и современные информационные технологии», Томск, 12-14 ноября 2014 г, а так же на международных научно-технических конференциях «Информационные системы и технологии» ИСТ-2015, Нижний Новгород и «Информационные системы и технологии» ИСТ-2016, Нижний Новгород.

Публикации

1. Поляков И. В., Гай В. Е. Монауральная система локализации источника звука \\\ Томск: Молодежь и современные информационные технологии, 2014. - С. 431
2. Поляков И. В., Гай В. Е. Поиск источника звука с одним микрофоном \\\ Нижний Новгород: Международная научно-техническая конференция «Информационные системы и технологии», 2015.
3. Поляков И. В., Гай В. Е. Использование монаурального подхода в системах поиска направления на источник звука \\\ Москва: Международная научно-техническая конференция «INTERMATIC», 2015. - С. 183-185.
4. Поляков И. В., Гай В. Е. Роль формы звукоулавливателя в монауральной системе локализации источника звука \\\ Нижний Новгород: Международная научно-техническая конференция «Информационные системы и технологии», 2016.

Структура и объём работы

Диссертация состоит из введения, трех глав, заключения, списка сокращений и обозначений и списка литературы. Полный объём диссертации составляет 66 страниц с 47 рисунками и 3 таблицами. Список литературы содержит 42 наименования.

Инв. № подл.	Подп. и дата	Инв. № дубл.	Взам. инв. №	Подп. и дата						
Ли	Изм.	№ докум.	Подп.	Дат	ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)					Лист
										6

Глава 1. Обзор методов локализации источников звука

1.1. Методы бинауральной локализации источников звука

В первую очередь к вопросу о принципах работы локализации источника звука пришли биологи. Способность живых организмов извлекать наибольшее количество информации из звука - жизненно необходимая вещь, так как слуховое восприятие снабжает нас данными об окружающем мире в объеме, уступающем только данным от зрительной системы [5, 6].

Возможность осуществлять звуковой поиск, то есть определять направление на источник звука - важное умение, ведь звук может выступать и сигналом опасности и использоваться для поиска жертвы. Одна из характерных черт системы слухового восприятия живых существ, имеющих парное расположение звукоулавливателей (к таким относится и человек), обеспечивающая пространственное ориентирование - это ее бинауральное строение. Уши улавливают звуковые волны распространяющиеся во внешней среде, далее периферическая часть слуховой системы приступает к их обработке, путем спектрально-временного анализа. в результате, отделы мозга, ответственные за звуковое ориентирование, сравнивают полученные характеристики и делают предположения о местоположении источника звука.

Характерный для парных звукоулавливателей способ пространственного звукового ориентирования называется бинауральным. Долгие годы считалось, что основным механизмом этого способа является анализ временной разницы между достижением звука до одного и до другого звукоулавливателя. Однако со временем исследователи обнаружили влияние и других факторов.

К основополагающим механизмам обычно относят локализацию по разнице амплитуд arriving сигналов, временной разнице, а также спектральным различиям звука в правом и левом слуховых каналах [7]. К вспомогательным механизмам обычно относят отражения звука от туловища и плеч человека, анализ реверберационных эффектов, а также эффект психологического восприятия, приводящий слышимое расположение звукового источника в соответствие с его расположением, которое мы видим глазами. Также существует механизм монаурального слухового восприятия, который по сравнению с вышеназванными механизмами изучается относительно недавно. Его принцип работы, особенности и области применения будут рассмотрены позже.

Анализ потенциала звуковой локации у живых существ, в первую очередь проводился с помощью экспериментов с людьми-добровольцами. Первым исследователем пространственного слуха у человека считается итальянский физик Д. Вентури (1746— 1822 гг.) [7].

Для изучения пространственного слуха Вентури проделал следующий опыт. Он поместил человека с завязанными глазами на открытом лугу, а сам ходил вокруг него на расстоянии пятидесяти метров и издавал звуки с помощью флейты или колокольчика [8]. Когда источник звука был под прямым углом по направлению к линии взора, человек легко определял место расположения источника звука.

Если же звук был направлен по диагонали, а голова оставалась неподвижной, ему трудно было определить, спереди или сзади находится источник.

Когда испытуемому разрешено было поворачивать голову, он не делал ошибок при определении места звучания источника. Вентури установил также, что при односторонней глухоте человек может локализовать звуки только при повороте головы в их сторону, причем звучание должно быть продолжительным. Исследователь отметил, что при такой патологии особенно сильно страдает локализация коротких звуков.

Основной вывод Вентури состоял в том, что локализация звука основана на оценке силы раздражителей, действующих на разные уши, и что два «слуховых впечатления» от каждого из звукоулавливателей не смешиваются внутри черепа. Эти исследования Вентури не привлекли внимания ученых и вскоре были забыты.

Спустя век после работ Д. Вентури, опубликовавшего их четыре раза: дважды на немецком, на французском и на итальянском языках (1796 — 1801 гг.), в 70-х годах XIX в. известный английский физик Д. Релей, очевидно, не знакомый с работами своего предшественника, провел точно такие же опыты и сделал аналогичные выводы [8].

В конце XIX—начале XX в. исследования пространственного слуха были малоинтенсивными и интересовали преимущественно клиницистов. Только в период первой мировой войны в связи с разработкой звуковых локаторов для пеленгации самолетов начались работы по изучению временных характеристик пространственного слуха на основе объединения данных, получаемых от двух ушей (бинауральный слух) (см. рис. 1.1).

Было экспериментально установлено, что чувствительность человека к нахождению направления на источник звука можно повысить путем увеличения расстояния между слуховыми отверстиями, что привело к созданию специальных установок (см. рис. 1.2, рис. 1.3).

Активные исследования по слуховой локализации возобновились в конце 20го века и привнесли несколько новых и довольно важных моментов.

По-прежнему человеческий слух и способность воспринимать направление источника звука присваивались использованию сразу двух ушей [9] [10], [11], [12]. Субъективный опыт бинаурального слуха в процессе вычисления источника звука сначала считался результатом интерактивного анализа двух звукоулавливателей [10], [11], [12], [13]. Человек по имени Лорд Рейлиг разработал «дуплексную теорию» (1990), основанную на мнении, что локализация звука объясняется тем, что уши разделяются некоторым расстоянием и акустически непрозрачной массой (голова) и вместе это придает два отличительных свойства приходящим звукам. Во-первых, звук, приходящий не с направления медиальной плоскости головы, достигнет одно ухо быстрее, чем второе, создавая разницу в прибытии, которая может быть определена и использована при локализации. Такой процесс назван задержкой во времени между ушами (interaural time difference - ITD) Фузессери, Венstrupом и Поллаком (1990) [14]. Во-вторых, масса головы заставляет приходящий звук терять интенсивность, при прохождении от одной стороны головы к другой. Фузессери, Венstrup и Поллак назвали такой процесс задержкой в интенсивности между ушами (interaural intensity difference - IID) из-за выступления головы в роли смягчителя звука.

В настоящее время рассматриваются два основных условия, определяющих бинауральный слух [8].

Инв. № подл.	Подп. и дата	Инв. № дубл.	Взам. инв. №	Подп. и дата	Инв. № инв.	Подп. и дата	Ли	Изм.	№ докум.	Подп.	Дат	ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)	Лист
													8



Рисунок 1.1. Голландский персональный локатор 1930-х годов.

Для низких частот основным фактором является различие во времени попадания звука в левое и в правое ухо, для высоких — различия в интенсивности. Сначала звук достигает уха, расположенного ближе к источнику. При низких частотах звуковые волны «оггибают» голову в силу большой их длины. Звук в воздушной среде имеет скорость 330 м/с. Следовательно, 1 см он проходит за 30 мкс. Поскольку расстояние между ушами у человека составляет 17–18 см, а голову можно рассматривать как шар с радиусом 9 см, то разница между попаданием звука в разные уши составляет $9 \cdot \pi \cdot 30 = 840$ мкс.

Для того чтобы можно было отметить такую небольшую разницу во времени, необходимы очень тонкие и точные механизмы сравнения. Такое сравнение осуществляется центральной нервной системой в местах, где импульсация от правого и левого ушей сходится на одной структуре (нервной клетке). Различия во времени поступления звука сопровождаются различиями в фазе звуковой волны. Более «поздний» звук отстает по фазе от предыдущего более «раннего» звука. Это отставание может быть использовано только при относительно низких частотах звуковой волны, при которых для завершения полного цикла требуется время не менее 840 мкс. Это соответствует длительности периода частоты не более 1300 Гц.



Рисунок 1.2. Дорадарные прослушиватели самолетов противника.

Однако, речь ведется о возможностях локализации звука благодаря участию в этом процессе двух ушей, применительно к способностям человека, межушное расстояние у которого достаточно велико. В свою очередь, стоит обратить внимание, к примеру, на мелких животных, у которых расстояние между ушами составляет всего несколько сантиметров и задуматься, с какой скоростью должны происходить вычисления, если по-прежнему принимать во внимание только бинауральный способ локации. У полевой мыши, например, расстояние между ушами всего 2 см, а локализация, хищников по звуку у нее отличное. Таким образом, по логике рассуждений напрашивается вывод, что мышь должна оценивать значительно меньшую разницу во времени, нежели человек, а это соответствует частотам выше 10 кГц. Встает вопрос, действительно ли нервная система способна оценивать подобные микроинтервалы времени. Чтобы выяснить это, было необходимо выдвижение новых теорий и их экспериментальная проверка.

Согласно с мнением Батлера и Хумански (1992) [1], роль ушной раковины в локализации звука начала серьезно изучаться только когда в некоторых экспериментах бинауральный принцип стал недостаточен. Одними из таких экспериментов были те, которые проводились с источниками звука, лежащими непосредственно на медиальной вертикальной плоскости и не отклоняющимися куда-либо по горизонтали. Таким образом, теоретически бинауральный метод не давал возможности определить на какой же высоте находится источник звука, а так же спереди он или сзади. Однако подопытные вполне стабильно успешно определяли направление на источник, что говорит о наличии еще какого-то скрытого механизма в человеческой системе обработки звука.

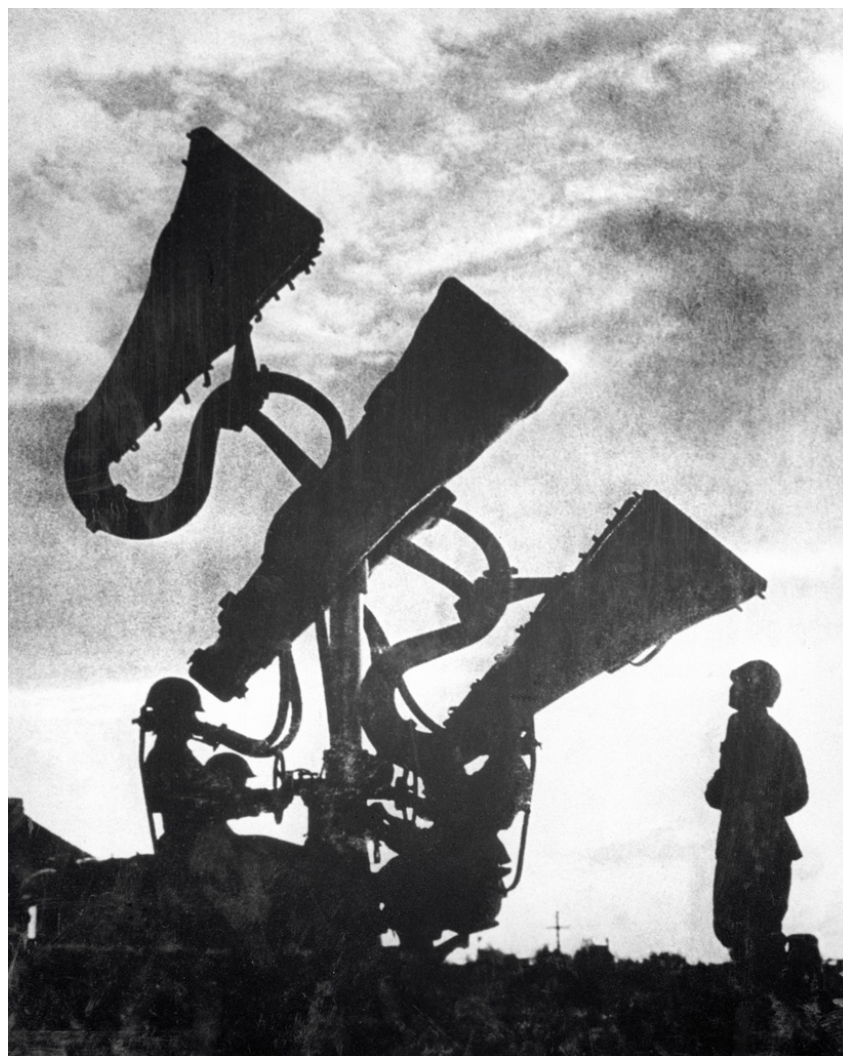


Рисунок 1.3. Звукоулавливающая установка ПВО. СССР, октябрь 1941 года.

Этот фактор послужил толчком к усовершенствованию приемника-обработчика, а так же поиску и применению дополнительных решений, кроме IID и ITD. И вот, в конце концов, в течение этих экспериментов было выявлено, что наружное ухо выступает в роли совокупности зависящих от направления фильтров, необходимых для локализации звука [1], [13], [11].

Независимые и практически параллельно идущие эксперименты другой команды исследователей изучают восприятие изменения высоты тона звука, без какого-либо перемещения источника звука [10], [13]. Обе группы ученых, Мидлбрукс и Грин, а так же Роджерс и Батлер отмечают, что при простом изменении звука на тон (при нахождении источника звука на вертикальной медиальной плоскости) слушатель ошибочно считает, что источник сдвинулся с места. Что интересно, изменения, подмеченные слушателями, находятся в пределах медиальной вертикальной плоскости. Эти наблюдения предоставляют дополнительную поддержку против дуплексной теории.

Также это дает информацию о динамическом звуковом манипулировании через внешнее ухо и о том, как оно служит для локализации монофонических сигналов или “сигналов ушной раковины” [10], [11], [14], [15], [16] .

Таким образом, в результате исследований бинаурального слуха человека были выделены принципы, использованные при построении различных систем звуковой локализации. Начиная с установок, где главным анализатором являлся человек, заканчивая современными массивными

системами определения направления на звуки работы двигателей различной техники (в основном военной), системами подводной эхолокации, а так же системами-помощниками и охранными системами с многомикрофонными массивами для определения направления на источник голоса.

1.2. Методы монауральной локализации источников звука

Большинство источников информации, рассмотренных по ключевому слову «монауральный» интерпретирует этот термин в качестве способа воспроизведения аудио записей. Такой способ требует для работы всего один громкоговоритель и не обеспечивает для слушателя такого эффекта погружения, как занявшее главенство стерео воспроизведение. Однако, исследуя не излучение, а восприятие звуковых волн, термин «монауральный» не стоит рассматривать, как устаревший или малофункциональный.

Как было рассмотрено выше, исследования бинаурального принципа показали, что его не достаточно для полного объяснения способностей человека ориентироваться в пространстве. Это подтолкнуло исследователей к обнаружению еще одного метода определения направления на источник звука.

У биологических организмов, в частности людей, «задержка во времени между ушами» (ITD) обычно используется в качестве сигнализатора для звуковой ориентации; в организмах, имеющих два уха (звукоулавливателя), это называется «бинауральной задержкой во времени» или бинауральной разницей [2]. Как бы то ни было, человек не может использовать бинауральную разницу ни для того, чтобы определить высоту нахождения источника звука, ни для того, чтобы узнать спереди приходит звук или сзади. К примеру, если источник звука непосредственно впереди нас, задержка во времени между ушами будет равна нулю, независимо от высоты положения источника. То же самое касается и расположения источника спереди/сзади.

Люди могут определять нахождение источника звука во всех направлениях трехмерного пространства с достаточно высокой степенью точности, включая и высоту нахождения источника и то, спереди он или сзади. Более того, человек может все это, используя всего одно ухо; это явление известно как монауральная локализация.

Монауральный метод нахождения направления на источник звука начал изучаться относительно недавно. В его основе лежит изучение роли ушной раковины в локализации звука человеком. Внешнее человеческое ухо, т.е. совокупность ушной раковины и наружной части ушного канала, формирует фильтры выборки направлений. В зависимости от угла направления приходящего звука, различные фильтры входят с ним в резонанс. Такие резонансы внедряют зависящие от направления шаблоны в частотные характеристики звука, которые могут оцениваться слуховой системой (диапазоны направлений) для пространственной локализации звука. Вместе с остальными отражениями, испытывающими влияние от направления, от головы, плеч и торса, они формируют передаточные функции внешнего уха. Вся эта монауральная информация необходима для определения, откуда пришел звук. Однако, для того, чтобы иметь возможность пользоваться данным методом и оценить положение источника звука, необходимо иметь базовое знакомство с тем звуком, с которым можно сравнить, что актуализирует проблему использования монаурального принципа в искусственных системах. Вообще, даже если принять во внимание, что звуки изменяются ухом в зависимости от угла падения, нужно отметить, что из одной единственной монофониче-

Инв. № подл.	Подп. и дата						
	Взам. инв. №						
	Инв. № дубл.						
	Подп. и дата						
Инв. № подл.	Ли	Изм.	№ докум.	Подп.	Дат	ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)	Лист
							12

ской записи невозможно определить направление на источник, ведь мы не можем знать есть ли в ней какое-либо специфическое отличие, привнесенное ушной раковиной или она вообще никак не изменена. Типичные звуки в нашем окружении (и в природе) не случайны – они имеют определенную структуру. Таким образом, с помощью наших предварительных знаний (возможно, благодаря многолетнему опыту) о том, на что должны быть похожи конкретные звуки, мы и можем наиболее вероятно предсказать направление. Также, психологическими исследованиями установлено, что у новорожденных монауральная локализация хуже, чем у взрослых. Одно из возможных объяснений кроется в том, что младенцы не имеют достаточного опыта с натуральными звуками, чтобы создать достаточную базу типичных звуков.

Звук приобретает некоторые отличия, в зависимости от направления, с которого он явился, из-за воздействия внешнего уха (см. рис. 1.4). Запутанная структура внешнего уха воздействует на звуковые волны, группирующие и продвигающиеся по ушному каналу. В процессе волны перекрываются, гасят друг друга и реверберируют. Эти воздействия влияют на спектральную характеристику звука по-разному, в зависимости от направления на источник. В характере такого влияния заключается монофоническая информация, необходимая для определения того, откуда пришел звук. Как бы то ни было, это порождает интересный факт: чтобы иметь возможность воспринимать меняющийся звук в качестве исходящего от одного и того же источника и поддерживать восприятие этого звука достаточно качественно, необходимо иметь с таким звуком предварительное знакомство.

Строение уха

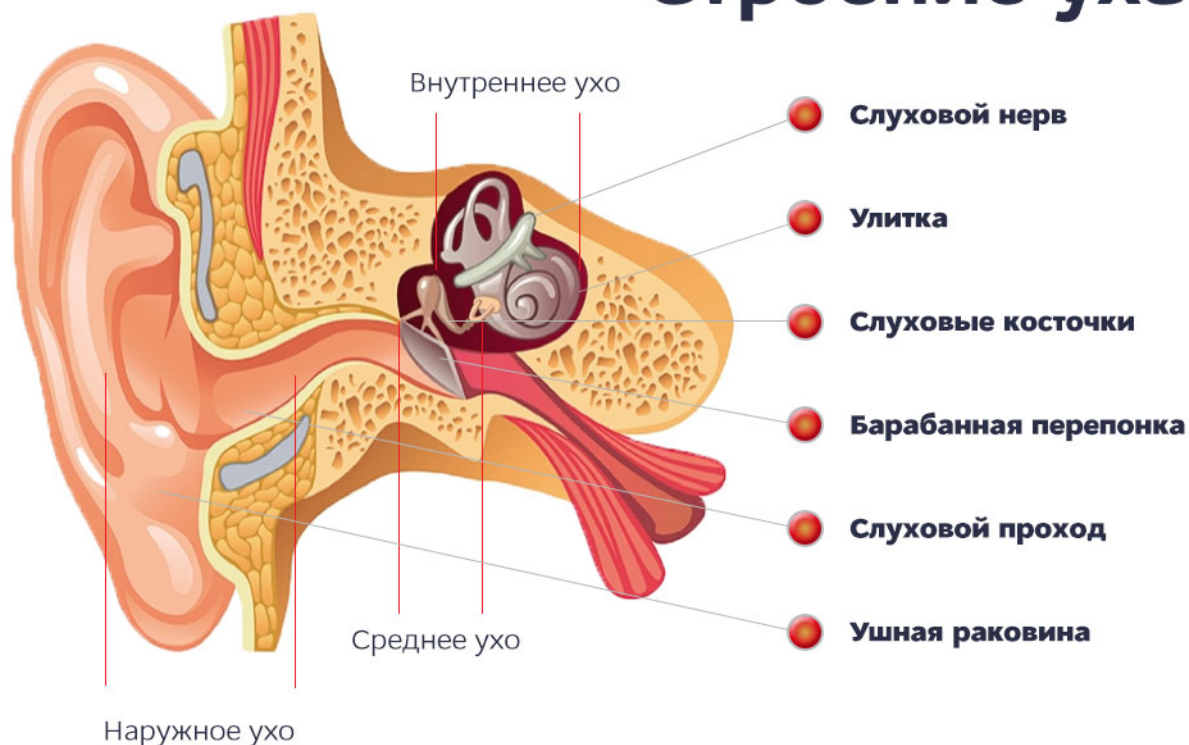


Рисунок 1.4. Строение уха человека.

Инва. № подл.	Подп. и дата	Инва. № дубл.	Взам. инв. №	Подп. и дата

Ли	Изм.	№ докум.	Подп.	Дат

Известно [17], что наша способность локализовать какой-либо звук, напрямую связана с количеством нашего опыта, со степенью знакомства с этим звуком. Предполагается, что мы "настраиваемся" на акустические модели, которые привносит в звук наше ухо и сперва мы составляем общую модель какого-либо звука, чтобы потом суметь распознать его измененные варианты. Другими словами, мы должны составить для себя эталонный образец звука, от которого будем отталкиваться, сравнивая с тем, как этот же звук проявляет себя в пространстве в разных местах.

Также, помимо эффектов, накладываемых внешним ухом, не стоит забывать о других факторах, влияющих на звук, в зависимости от направления на источник. Это эффекты, возникающие вследствие отражения от плеч, а так же от процесса огибания звуковой волной головы, на пути до дальнего уха. В психологической литературе обычно считается, что бинауральный принцип используется человеком в основном для определения азимута (направление вправо-влево), в то время как монофонические сигналы, по большей части, используются для определения высоты (направления выше-ниже) [1], [11], [13]. Некоторые исследователи [12], однако, верят, что монауральный принцип более важен, чем бинауральный, даже для определения азимута. Как было сказано ранее, в момент, когда источник звука располагается на медиальной вертикальной плоскости (то есть на равном расстоянии от каждого из ушей), использование бинаурального принципа не принесет никаких результатов. Таким образом, определение расположения источника звука на плоскости опирается на монофонические сигналы, претерпевшие изменения от влияния внешнего уха. Рассмотрев внешнее ухо человека с его сложной и несимметричной, относительно центра ушного канала, формой становится понятным, каким образом обеспечивается необходимая точность нахождения угла направления приходящего звука.

В биологических системах поиск направления на источник звука и понимание содержания, которое этот звук в себе несет представляются как два противоречащих друг другу требования. То есть, чтобы понимать содержание звука, логично желать наименьших искажений; однако, чтобы локализовать звук, нужно иметь различия в звуке при различных углах направления на его источник. Уши живых существ получили свои особенности в развитии, чтобы удовлетворить оба этих требования. К примеру, летучие мыши и совы [18], [19], совершающие эхо локацию, имеют уши, сильно влияющие на звук, в зависимости от того, откуда он пришел, в то время как уши свиньи относительно мягко влияют на изменения, происходящие со звуком в результате смены направления на его источник.

Когда задача локализации направления на источник звука не ограничивала размеры принимающих установок (локация авиации и подводных судов, многомикрофонные массивы для поиска направления на открытой местности), бинауральный принцип применялся достаточно успешно. Однако, когда возникла потребность уменьшить размеры приборов, их стоимость, а так же устранить некоторые недостатки, исследователи все чаще стали уделять внимание вопросам разработки и улучшению систем, основанных на монауральном принципе.

Из современных работ по этой тематике стоит выделить исследования Тетсуи Такигучи [20] и Нишимуры [21], в которых были поставлены эксперименты по монауральному распознаванию направления на диктора. В их работе для отделения речи диктора от сопровождающих шумов была использована Скрытая Марковская модель. В их следующем исследовании [3] вместо Скрытой модели Маркова они используют Гауссовскую смешанную модель, что по их отчетам повысило чистоту выделяемых речевых звуков и позволило добиться лучших результатов при конечной

Инв. № подл.	Подп. и дата	Инв. № дубл.	Взам. инв. №	Подп. и дата						
					ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)					Лист
Ли	Изм.	№ докум.	Подп.	Дат						14

локализации. Эти исследования очень поучительны, однако в них для придания отличительных особенностей приходящему звуку используется акустическая форма комнаты, в которой производилась запись (акустические особенности комнаты выполняют роль сложного звукоотражателя и заменяют, в данном случае, внешнее ухо человека). В формате устройства, от которого не требуется производить локализацию в помещении, отличающемся от того, где была произведена первоначальная настройка (обучение), такой подход вполне неплох, также он позволяет избавиться от дополнительного оснащения микрофона акустическими фильтрами.

Однако, планируя более универсальную систему в плане расположения на местности и используя принципы схожести с биологическими системами (внешнее ухо человека), можно получить систему, похожую на представленную в статье Learning Sound Localization from a Single Microphone [2]. В данном исследовании для разделения голоса и постороннего шума также использовалась Скрытая модель Маркова, однако роль объекта, меняющего параметры звука в зависимости от расположения источника, играла не комната, где проводились эксперименты, а небольшие акустические фильтры, установленные рядом с микрофоном. Эти фильтры имели форму небольших стенок различной фактуры и формы, на некотором уровне имитируя внешнее ухо человека. Таким образом, результаты исследования показывают интересные зависимости в точности локализации, при экспериментах с различными акустическими фильтрами и различными типами звуков.

Проанализировав вышеприведенные исследования можно сделать вывод, что вполне реально добиться повышения результатов, проведя работы по улучшению как физической части установки (более сложный акустический фильтр у микрофона, имеющий большую схожесть с человеческим внешним ухом), так и программной ее части, используя алгоритмы, предположительно превосходящие те, что упомянуты в вышеназванных статьях. Реализация монаурального принципа локализации в стандартном представлении осложняется моделированием различных передаточных функций, ответственных за изменение эталонного звука, в зависимости от направления, с которого он пришел. Однако, с развитием методик анализа данных появилась возможность реализовать систему, схожую по принципу действия с человеческой системой слуха.

1.3. Методы монауральной локализации направления на источник звука

Подробно рассмотрены несколько известных методов локализации направления на источник звука, основанных на монауральном подходе. Первый из рассмотренных методов - Learning Sound Location from a Single Microphone [2].

Формальное описание модели. Представим $s(t)$ как источник звука, а $h_{\theta}(t)$ как передаточную функцию, зависящую от направления приходящего сигнала (по аналогии с зависимой от направления передаточной функцией, привносящуюся человеческим внешним ухом). Сигнал, записанный микрофоном $y(t)$, будет представлен: $y(t) = h_{\theta}(t) * s(t) + \omega(t)$ (1), где «*» обозначает свертку, а $\omega(t) \in N(0, \sigma_{\omega}^2)$ - это добавочный белый шум Гаусса.

Зависимый от направления отклик h_{θ} не зависит от времени, таким образом примем систему микрофона и отражателя звука как линейную. Следовательно, можно представить h_{θ} в частотной области (Фурье) [22] как $H_{\theta}(f) = F[h_{\theta}(t)]$. В общем смысле, фурье-преобразование $F[\cdot]$ представляет $h_{\theta}(t)$ влияющим на степень усиления или ослабления исходного звука на различных частотах f . Принимая некоторые допущения, далее $H_{\theta}(f)$ будет обозначать вектор H_{θ} . Затем бы-

Инв. № подл.	Подп. и дата						
	Взам. инв. №						
	Инв. № дубл.						
	Подп. и дата						
Инв. № подл.	Ли	Изм.	№ докум.	Подп.	Дат	ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)	Лист
							15

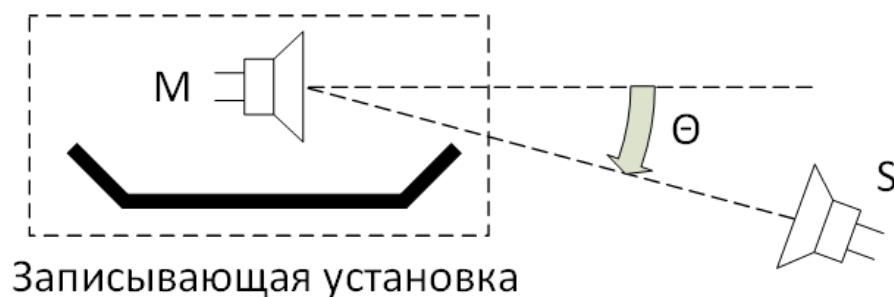


Рисунок 1.5. Иллюстрация записывающей системы, показан микрофон (М), источник звука (S) и угол направления на источник.

ли применены некоторые стандартные методы привнесения шума [23] для того, чтобы оценить H_θ для различных направлений θ , исследуемых на установке. Таким образом, это дает довольно точную модель системы зависимости от направления на источник звука, в применении к различным звуковым источникам. Далее следует описание модели источника звука $s(t)$. Различные свойства звуков $s(t)$ изменяются во времени. Например, когда произносится слово «привет», акустические свойства частей «п», «р», «и», «в», «е» и «т» сильно различаются. Похожий эффект справедлив и для других природных звуков. Чтобы охватить все эти зависящие от времени эффекты, применяется оконное преобразование Фурье (STFT) для вычисления представления частотной области $St(f)$ сигнала в момент времени t . Такая величина, возведенная в квадрат, $S_t^2(f)$, называется спектрограммой. Также в тексте S_t^2 будет обозначать вектор, содержащий все компоненты $S_t^2(f)$.

Несмотря на то, что звуковой сигнал $S_t^2(f)$ стохастический и его статистика (первого и второго порядка) изменяется как функция от времени, множество звуков в природе можно смоделировать в виде стационарных, при рассмотрении только коротких временных промежутков. Например, каждая из частей сигнала «привет» (каждая фонема) может быть довольно точно смоделирована, в качестве стационарной (для фонемы значение и корреляция остаются относительно постоянными), даже в случае, если это нельзя применить ко всему слову целиком. Следуя практике обработки аудио сигналов [24], [25], $S_t^2(f)$ будет моделироваться при помощи скрытой марковской модели (СММ) [26]. Используется модель наблюдения $P(S_t^2|z_t) = N(S_t^2; \mu_{z_t}, \Sigma_{z_t})$, где $z_t \in \{1, \dots, K\}$ является состоянием СММ в момент времени t . Переходы НММ из z_t состояния в z_{t+1} соответствует переходам между состояниями вероятностей $P(z_{t+1}|z_t)$. СММ модель обучается с помощью алгоритма ЕМ (Баума-Вельша), с параметрами, сочетающимися с использованием метода кластеризации k-средних, для группирования наблюдений S_t в обучающей выборке в различные дискретные состояния z_t . Модель обучалась с использованием множества природных и искусственных звуков.

Оптимизация $\arg\max$ производится в два шага. Во-первых, дискретизируется θ с шагом в 15° и вычисляется результат для каждого из значений $\theta \in 0, 15, 30, \dots, 345$ для вычисления вероятности состояний СММ. Следующим шагом в локальной окрестности проводятся вычисления с шагом угла в 2 градуса, для повышения точности результата.

Второй из рассмотренных методов - Monaural Sound Localization [27].

Оценка звукового направления основывается на наборе передаточных функций головы (HRTF), созданных при использовании головы манекена, а так же на статистической модели речи. Предложенный метод требует обучения только одной модели гауссовой смеси с помощью речи. Оценка выполняется в спектральной области, при использовании подхода максимального правдоподобия.

Инв. № подл.	Подп. и дата
Инв. № дубл.	Взам. инв. №
Подп. и дата	Инв. № дубл.
Инв. № подл.	Подп. и дата

Ли	Изм.	№ докум.	Подп.	Дат
----	------	----------	-------	-----

Звукозаписывающая установка также представляет из себя микрофон, закрепленный на месте уха макета головы человека.

В качестве моделей, на которых основывается получение характерных особенностей направлений, с которых приходит звук, использовалась одна из баз записей HRTF, находящаяся в свободном доступе. Из этих записей выделяются признаки, характерные для направлений приходящего звука, располагающихся в 5° друг от друга. На звукозаписывающей установке производится запись речи нескольких дикторов с различных направлений. Из записей дикторов каждого направления также вычисляются признаки, а затем описываются с помощью моделей гауссовых смесей (GMM). Далее, модели, описывающие признаки записи эталона (HRTF) сравниваются с моделями, полученными при описании записи какого-либо направления, выполненной диктором в эксперименте. На основании результатов сравнения делается вывод о том, с какого направления приходил звук в экспериментальной записи (см. рис. 1.6).

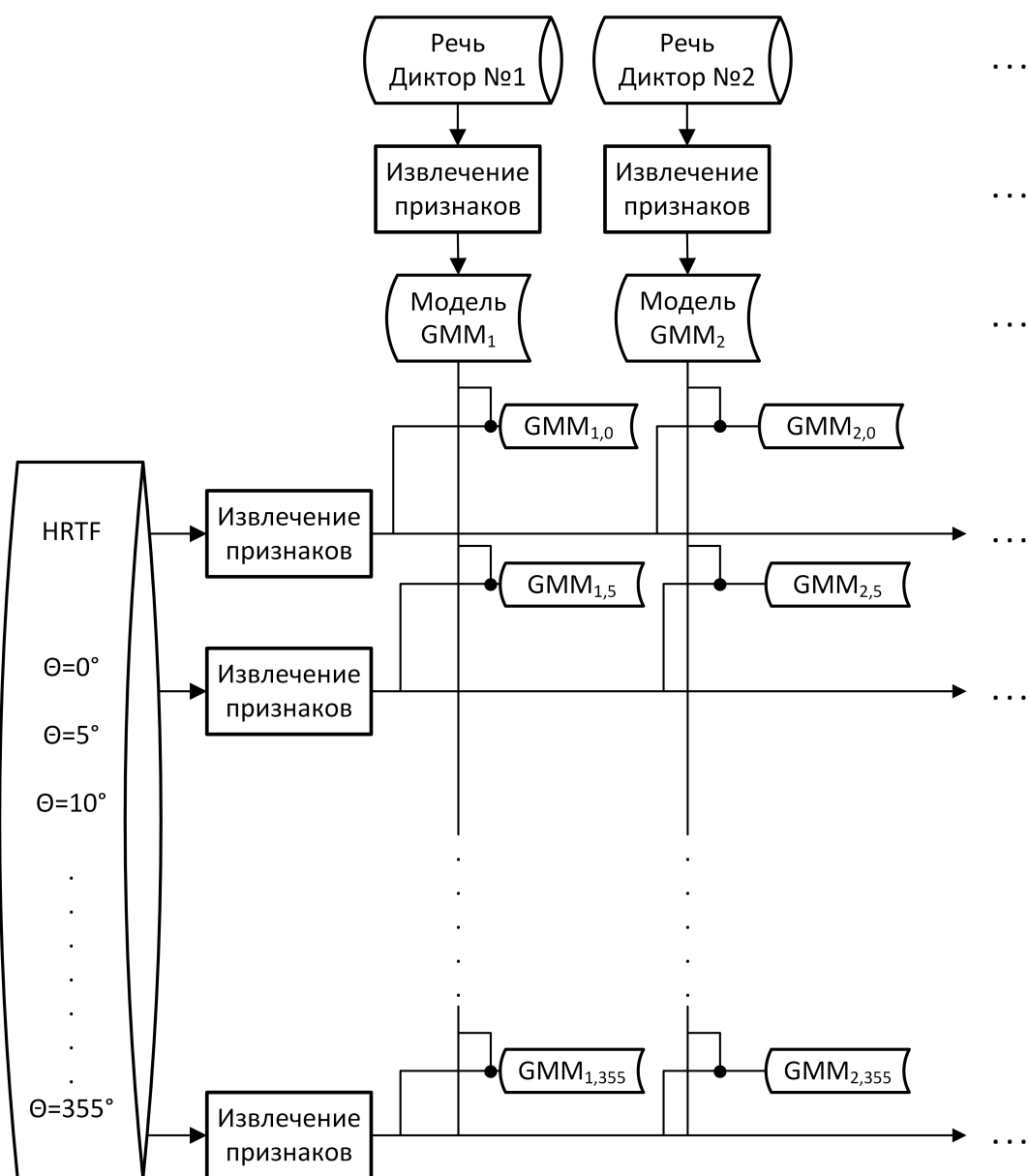


Рисунок 1.6. Иллюстрация модели работы системы, основанной на использовании передаточных функций головы.

Инв. № подл.	Подп. и дата
Инв. № дубл.	Взам. инв. №
Подп. и дата	Подп. и дата
Инв. № подл.	Подп. и дата

Ли	Изм.	№ докум.	Подп.	Дат
----	------	----------	-------	-----

1.4. Выводы по главе 1

Эффективность работы системы локализации источника звука в большей степени зависит от метода анализа приходящих звуковых сигналов, выбранного для реализации. Одним из методов локализации источника звука является бинауральный метод. Данный метод использует для работы учет временных и амплитудных различий характеристик звуковых сигналов, приходящих в отдельные звукоприемники. Такой метод требует относительно крупных габаритов звукоприемных установок, имеет относительно невысокую точность при возрастании расстояния до источника звука, а так же имеет ограничения на корректное определение направления на источник звука, при нахождении источника на равноудаленном расстоянии от звукоприемников.

Также для осуществления поставленной цели возможно использование другого метода, имеющего название монаурального. Данный метод использует оценку влияния частотных и амплитудных искажений на звук, привносимых акустической формой звукоулавливателя (или окружающей обстановки).

Для решения задачи монауральным способом, могут быть использованы алгоритмы и методы цифровой обработки и анализа данных (например, анализ изображений). При работе системы поиска источника звука важным ее качеством является степень сложности обработки сигнала, что напрямую сказывается на общем быстродействии. Алгоритмы, используемые в известных системах, включают в процесс вычислений довольно сложные математические действия (умножение, извлечение корня), таким образом, выбрав алгоритм, основанный на простых математических действиях (сложение, вычитание), можно повысить быстродействие всей системы.

Учитывая особенности монаурального способа при решении задачи поиска источника звука, создаваемая система принимает вид системы распознавания образов. Необходимо решить, каким образом будут проводиться этапы обработки сигнала, характерные для данного типа систем: предварительная обработка сигнала, составление системы признаков и классификация полученных элементов.

Инв. № подл.	Подп. и дата	Инв. № дубл.	Взам. инв. №	Подп. и дата						
					ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)					Лист
Ли	Изм.	№ докум.	Подп.	Дат						18

Глава 2. Монауральная локализация источника звука с позиции ТАВ

2.1. Обработка звуковых сигналов с позиций теории активного восприятия

Пусть $f(t)$ – анализируемый звуковой сигнал, наблюдаемый на конечном отрезке времени. Результат применения U -преобразования к сигналу f : $D = U[f]$ – многоуровневое (грубо-точное) спектральное представление $D = \{d_{i,j}\}$, $i = \overline{0, K-1}$, $j \in \overline{1, M_i}$, где K – число уровней разложения, M_i – количество сегментов сигнала на i -ом уровне разложения, $d_{i,j}$ – спектр, включающий L спектральных коэффициентов (число спектральных коэффициентов равно числу используемых фильтров), $d_{i,j}(k)$ – k -ый спектральный коэффициент ($k = \overline{1, L}$), $H = \{h_{i,j}\}$ – множество сегментов, полученных из исходного сигнала f , $h_{i,j}$ – сегмент сигнала f , по которому вычислен спектр $d_{i,j}$, $h_{i,j}(l)$ – l -ый отсчёт сегмента, T_i – количество отсчётов сегмента сигнала $h_{i,j}$ на i -ом уровне разложения, $F = \{f_i\}$ – набор фильтров Уолша системы Хармута, используемых для построения разложения, где $i \in \overline{1, L}$, $f_i(k)$ – k -ый отсчёт i -го фильтра.

В приведенном примере для вычисления спектрального представления сигнала используется четыре фильтра [28]. Однако для решения задачи оценки частоты основного тона требуется более тонкий анализ сигнала. В связи с этим, при построении спектрального представления сигнала будет использовано 16 фильтров (см. рис. 2.1).

Рассмотрим случай, когда отсчёты сигнала не нормированы по амплитуде к отрезку $[0; 1]$ и могут принимать отрицательные значения. Значения отсчётов сигнала, принимают как положительные, так отрицательные значения. Результат, полученный после вычисления массы такого сигнала является некорректным, так как масса равна нулю. Из этого следует, что в данном сигнале не содержится информации и его дальнейший анализ, с позиций ТАВ, можно не проводить. После нормализации амплитуд сигнала к отрезку $[0; 1]$, получим, что массы сигнала отлична от нуля и, следовательно, необходим его дальнейший анализ.

2.2. Алгоритмы формирования спектрального представления звукового сигнала

Рассмотрим варианты построения U -преобразования анализируемого сигнала [29].

1. алгоритм формирования R -ичного дерева разложения сигнала (параметры алгоритма: P – количество уровней разложения, L_i – количество сегментов, на которые разбивается сигнал на i -ом уровне), в вершине дерева расположено разложение исходного сигнала;
2. алгоритм построения разложения сигнала на одном уровне с использованием сегмента произвольной длины (параметр алгоритма: Z – длина сегмента);
3. алгоритм построения разложения сигнала на одном уровне, при формировании которого сигнал разбивается на заданное количество сегментов (параметр алгоритма: L – количество сегментов, на которые разбивается сигнал).

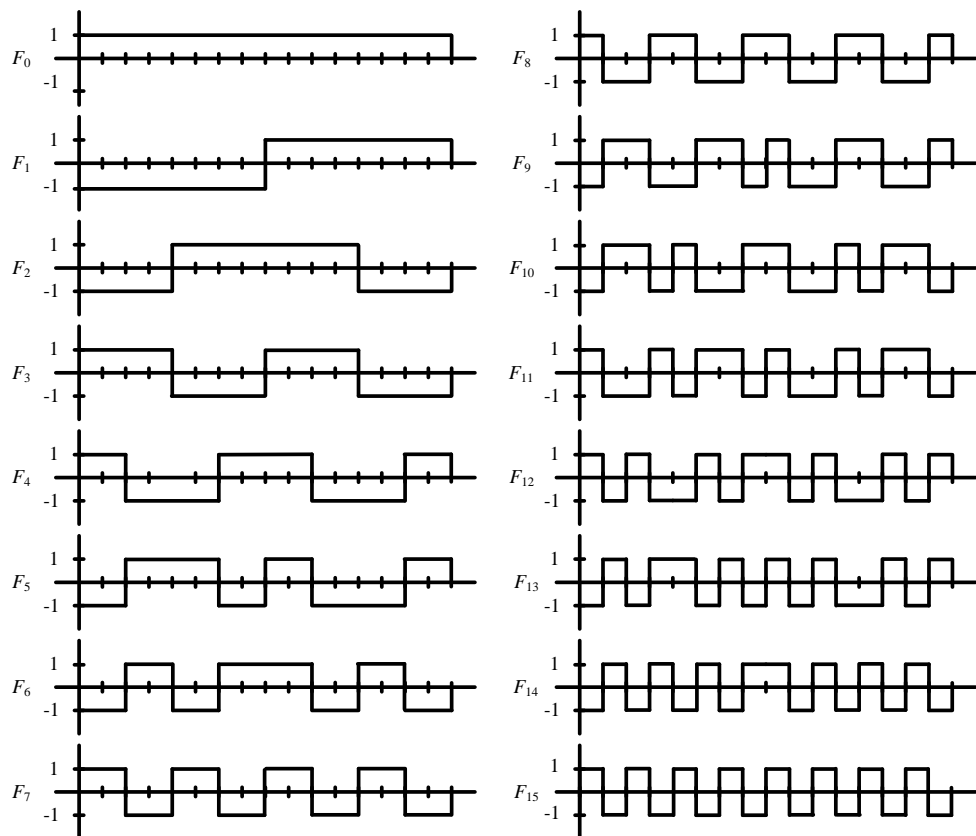


Рисунок 2.1. Система фильтров

В ряде случаев может использоваться комбинация первого и второго алгоритмов: на основе второго алгоритма формируется первоначальное разбиение сигнала на сегменты, после чего для каждого из них формируется R -ичное дерево разложения с помощью первого алгоритма. Каждый из алгоритмов должен учитывать возможность того, что длина сегмента может не быть кратной длине фильтра.

2.2.1. Генерация многоуровневого представления сигнала

Алгоритм состоит из двух этапов:

1. формирование представления сигнала f в виде дерева H , в вершине дерева (на нулевом уровне) находится исходный сигнал f , на i -ом уровне дерева находится L_i сегментов, полученных в результате разбиения на R частей сигнала (или его сегмента) находящегося на $i - 1$ -ом уровне;
2. вычисление спектрального представления D сигналов, входящих в дерево H , с помощью U -преобразования.

Рассмотрим алгоритм подробнее:

1. формирование разложения исходного сигнала (вершины дерева):

(а) $i = 0$, $H_{i,1} = f$, $Z_0 = |f|$, где $|f|$ – количество отсчётов в сигнале f ;

Вследствие выполнения операции округления возможна ситуация, в которой длина фильтра будет больше длины анализируемого сегмента. В этом случае нужно дополнить сегмент до длины фильтра нулями или отсчётами, относящимися к следующему сегменту.

Примечания:

- i. на нулевом уровне сегмент $h_{0,1}$ дополняется нулями, при этом, при генерации первого уровня разложения добавленные нули из сегмента должны быть удалены;
- ii. если используется методика дополнения сегмента следующими отсчётами, то для последнего сегмента на i -ом уровне не существует отсчётов, которыми его можно дополнить, в этом случае координата начала сегмента в сигнале смещается назад на количество отсчётов, которые надо добавить к сегменту, т.е. отсчёты выбираются из предыдущего сегмента.

2. формирование сегмента $h_{i,j}$ сигнала f :

(a) $k = 1, i = 1$;

(b) $\forall j \in [1, L_{i-1}]$:

$$\{h_{i,k}, \dots, h_{i,k+K-1}\} = DIV(h_{i-1,j})$$

$k = k + K$, где $DIV(\bullet)$ – оператор деления сигнала на R частей. Оператор $DIV(\bullet)$ выполняет следующие действия:

- i. определяет координаты сегментов и выделяет сегменты из сигнала;
- ii. дополняет сегменты отсчётами так, чтобы длина сегмента была кратна длине фильтра, используемого при построении разложения.

(c) вычисление структурных элементов сегмента h_{ij} (деление сегмента h_{ij} на N частей и применение операции интегрирования к каждой части):

$$m_{i,j}(k) = \sum_{l=(k-1)*(Z_i/N)+1}^{k*(Z_i/N)} h_{i,j}(l)$$

(d) $i = i + 1$.

Шаги (a) – (d) должны выполняться до тех пор, пока не будут сгенерированы P уровней дерева разложения и соответствующие им фильтры.

Минимальная длина сегмента сигнала (в отсчётах), по которому может быть сформировано спектральное представление равняется числу используемых фильтров. В связи с этим, построение дерева разложения сигнала должно быть остановлено, если значение Z_i станет равным N .

3. вычисление спектрального представления каждого сегмента дерева H :

$$d_{i,j}(n) = \frac{1}{N} \cdot \sum_{k=1}^N m_{i,j}(k) \cdot F_i(k), n \in \overline{1, N}.$$

Инв. № подл.	Подп. и дата	Инв. № дубл.	Взам. инв. №	Подп. и дата						
Ли	Изм.	№ докум.	Подп.	Дат	ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)					Лист
										21

1. амплитуда анализируемого сигнала для выполнения операции восстановления должна быть нормирована к отрезку $[0; 1]$. Данная операция должна быть выполнена до построения разложения сигнала;
2. при построении разложения набор фильтров генерируется для каждого уровня разложения, вследствие чего снижается скорость построения разложения;
3. полученное разложение избыточно из-за перекрытий сегментов на одном уровне. В результате, один и тот же отсчёт исходного сигнала может входить в два сегмента одновременно ($Z_P \cdot L_P > |f|$);
4. результаты тестирования показали, что после выполнения всех операций по дополнению сегментов отсчётами, каждый сегмент на i -ом уровне разложения будет включать одинаковое количество отсчётов;
5. выше предложен вариант дополнения сегмента отсчётами из следующего сегмента, в качестве альтернативы, можно использовать отсчёты не только из следующего, но и одновременно из предыдущего и следующего сегментов.
6. части сигнала не будут пересекаться, только в том случае, если длина анализируемого участка сигнала кратна количеству коэффициентов фильтра.

Рассмотрим предлагаемый алгоритм:

1. выборка из сигнала f очередного сегмента h_i (величина сдвига сегмента кодируется параметром S);
2. сравнение длины фильтра N и длины сегмента Z (их размеры могут быть различны вследствие применения операции округления на предыдущем шаге):
 - (а) если длина сегмента меньше длины фильтра, то сегмент дополняется отсчётами;
 - (б) если длина сегмента больше длины фильтра, сегмент обрезается до длины фильтра.
3. вычисление структурных элементов сегмента h_i ;
4. формирование спектрального представления d_i сегмента h_i ;
5. шаги 1 – 4 повторяются до тех пор, пока не будут обработаны все сегменты сигнала f .

1. при тестировании алгоритма оказалось, что число добавляемых (удаляемых) отсчётов не превышает половины длины фильтра из набора G ;

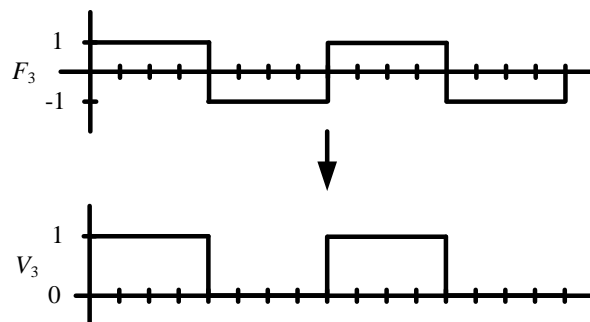


Рисунок 2.2. Вычисление оператора по фильтру из 16 отсчётов

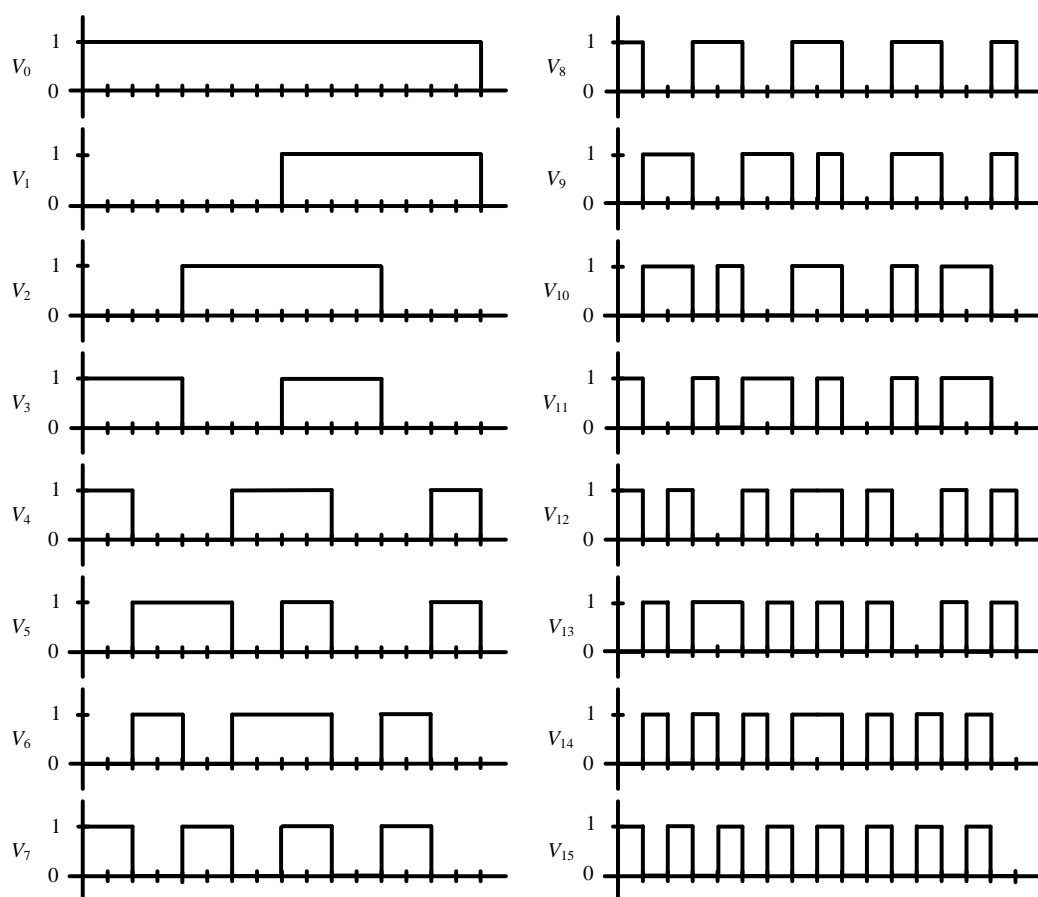


Рисунок 2.3. Система операторов

Инв. № подл.	Подп. и дата			
	Взам. инв. №			
Инв. № дубл.	Подп. и дата			
	Взам. инв. №			
<div style="display: flex; justify-content: space-between;"> <div> <p>Ли</p> <p>Изм.</p> <p>№ докум.</p> <p>Подп.</p> <p>Дат</p> </div> <div> <p>ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)</p> </div> <div> <p>Лист</p> <p>24</p> </div> </div>				

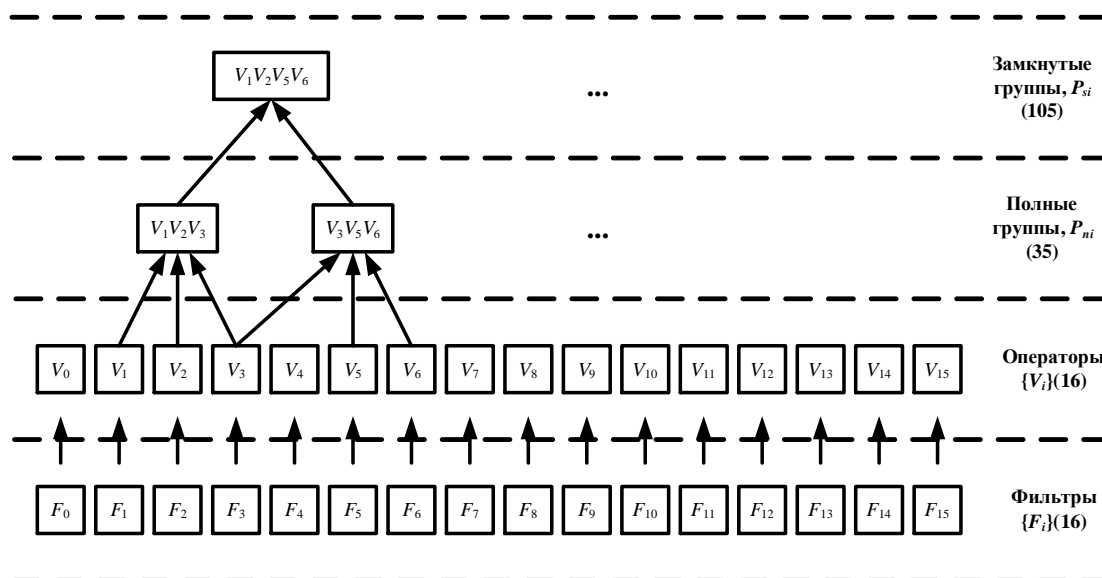


Рисунок 2.4. Иерархия операторов и групп

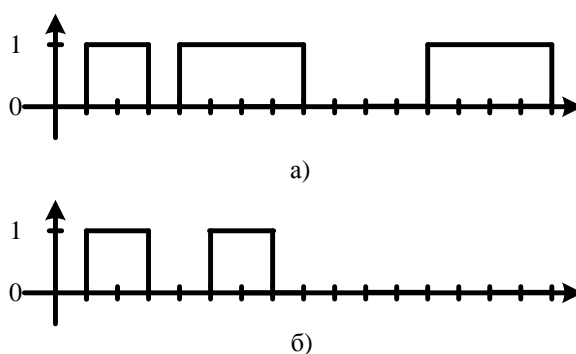


Рисунок 2.5. Вычисление массы полной группы: а) анализируемый сигнал f ; б) бинарный образ группы $\bar{V}_1\bar{V}_8V_9(P)$

Изм. № подл.	Подп. и дата	Изм. инв. №	Взам. инв. №	Изм. № дубл.	Подп. и дата	Изм. инв. №	Взам. инв. №	Изм. № дубл.	Подп. и дата

Ли	Изм.	№ докум.	Подп.	Дат	ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)	Лист
						26

2.3.1. Полные группы

Свойства полных групп

1. каждая полная группа образованна на тройке спектральных компонент. Поэтому она позволяет выявлять корреляционные (энергетические) связи между этими компонентами;
2. полная группа допускает два описания – на операции пересечения $V_i V_j V_k$ (операция умножения, число инверсий – чётно) и операции объединения $\bar{V}_i + V_j + V_k$ (операции сложения, число инверсий – нечётно);
3. полная группа на операции сложения выделяет сегмент сигнала, на котором сумма его отсчётов минимальна, на операции умножения – сегмент сигнала, на котором сумма его отсчётов максимальна.

Алгоритм генерации полных групп

Алгоритм генерации полных групп заключается в следующем:

1. выбираются три неинверсных оператора V_i, V_j, V_k такие, что $V_i + V_j + V_k \equiv e_1$ ($e_1 \equiv V_0$);
2. перебираются все возможные сочетания операторов – в результате генерируются все возможные полные группы.

Алгоритм вычисления полных групп, описывающих сигнал

Рассмотрим алгоритм вычисления полных групп, описывающих сигнал:

1. вычислить для сигнала f спектр;
2. каждому элементу спектра ставится в соответствие оператор из множества $\{V_i\}$ (элементы спектра нумеруются с нуля, номера операторов и элементов спектра совпадают).
3. выполняется проверка наличие i -ой полной группы в сигнале f :
 - (a) из спектра выбираются три элемента, номера которых соответствуют номерам операторов, входящих в i -ую группу;
 - (b) определяется количество инверсных операторов: если значение j -го элемента спектра меньше нуля, то считается, что соответствующий элементу оператор – инверсный;
 - (c) рассматривается две ситуации:
 - i. число инверсных операторов чётно: вычисляется образ полной группы на операции умножения. Пусть из спектра выбраны три элемента, соответствующие операторам V_1, V_8, V_9 . На рис. 2.8 показан пример вычисления образа полной группы $\overline{V}_1 \overline{V}_8 V_9$ (см. рис. 2.6);
 - ii. число инверсных операторов нечётно: вычисляется образ полной группы на операции сложения. На рис. 2.9 показан пример вычисления образа полной группы $\overline{V}_1 + V_8 + V_9$ (см. рис. 2.7).

1. вычислить для сигнала f спектр;
2. каждому элементу спектра ставится в соответствие оператор из множества $\{V_i\}$ (элементы спектра нумеруются с нуля, номера операторов и элементов спектра совпадают).
3. выполняется проверка наличие i -ой полной группы в сигнале f :
 - (a) из спектра выбираются три элемента, номера которых соответствуют номерам операторов, входящих в i -ую группу;
 - (b) определяется количество инверсных операторов: если значение j -го элемента спектра меньше нуля, то считается, что соответствующий элементу оператор – инверсный;
 - (c) рассматривается две ситуации:
 - i. число инверсных операторов чётно: вычисляется образ полной группы на операции умножения. Пусть из спектра выбраны три элемента, соответствующие операторам V_1, V_8, V_9 . На рис. 2.8 показан пример вычисления образа полной группы $\overline{V}_1 \overline{V}_8 V_9$ (см. рис. 2.6);
 - ii. число инверсных операторов нечётно: вычисляется образ полной группы на операции сложения. На рис. 2.9 показан пример вычисления образа полной группы $\overline{V}_1 + V_8 + V_9$ (см. рис. 2.7).

4. вычисляется масса группы.

Шаги 3-4 выполняются для всех 35 полных групп. В результате, формируется два массива масс полных групп: для масс полных групп на операции сложения и операции умножения.

Каждой полной группе, присутствующей в сигнале, в зависимости от числа инверсий можно поставить в соответствие число от 1 до 140. Данная операция выполняется отдельно для полных групп на операции сложения и полных групп на операции умножения.

V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}	V_{13}	V_{14}	V_{15}
100	-12	-6	4	3	9	11	23	-2	5	21	11	3	0	7	8

Рисунок 2.6. Спектр сигнала f (для определения полной группы на операции умножения)

V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}	V_{13}	V_{14}	V_{15}
100	-12	-6	4	3	9	11	23	2	5	21	11	3	0	7	8

Рисунок 2.7. Спектр сигнала f (для определения полной группы на операции сложения)

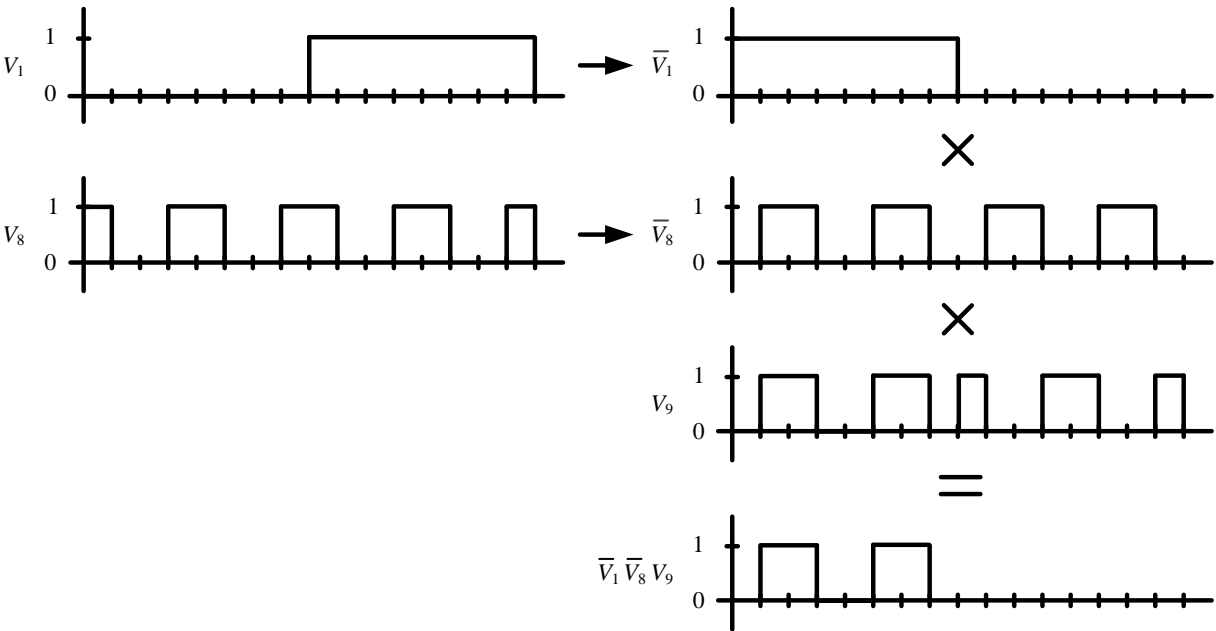


Рисунок 2.8. Вычисление образа полной группы на операции умножения

На рис. 2.11 показан пример выделения участка сигнала с максимальной массой с помощью полной группы на операции умножения (анализируемый сигнал приведён на рис. 2.10).

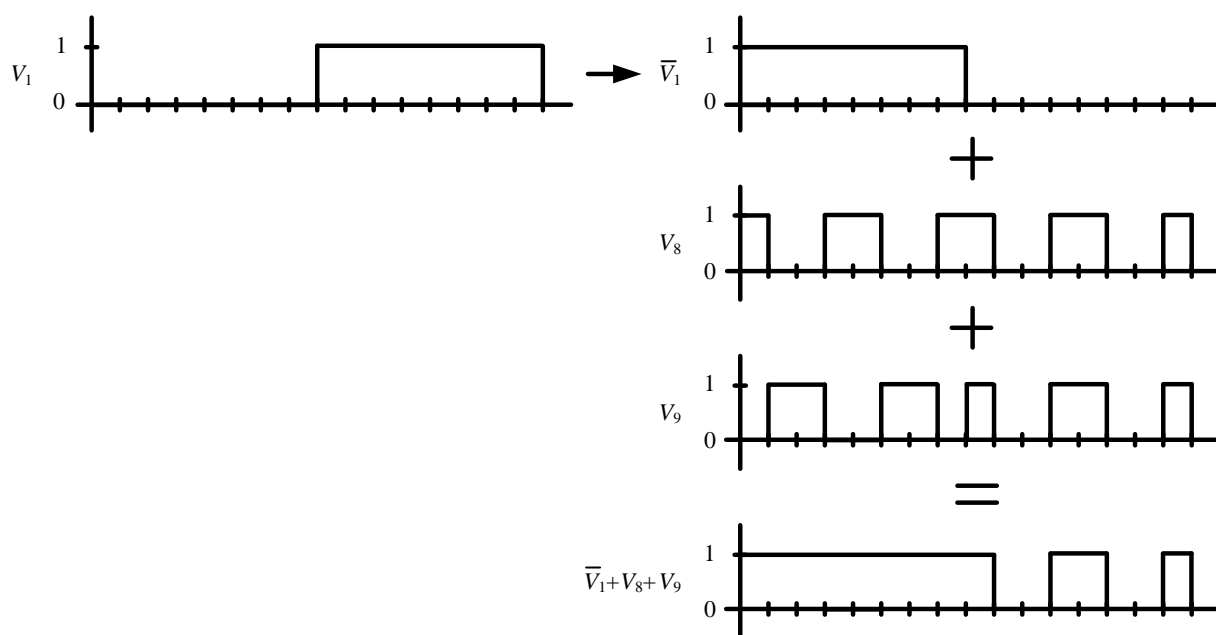


Рисунок 2.9. Вычисление образа полной группы на операции сложения

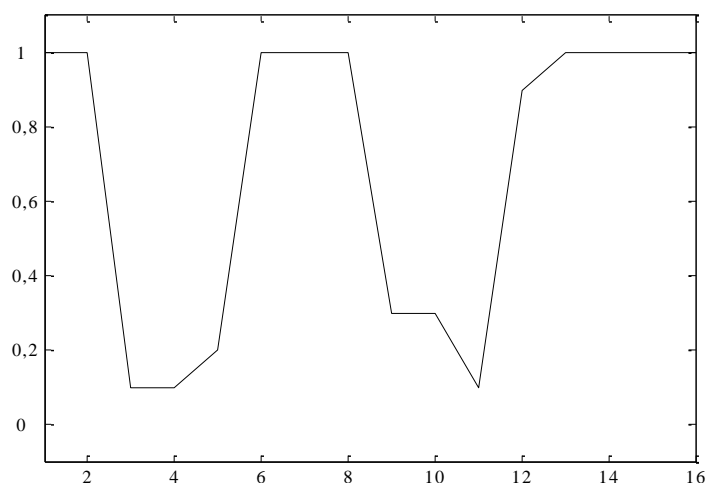


Рисунок 2.10. Анализируемый сигнал (1, 1, 0.1, 0.1, 0.2, 1, 1, 1, 0.3, 0.3, 0.1, 0.9, 1, 1, 1, 1)

Инв. № подл.	Подп. и дата			
	Взам. инв. №			
	Инв. № дубл.			
	Подп. и дата			
<div style="display: flex; justify-content: space-between;"> <div> <p>Ли</p> <p>Изм.</p> <p>№ докум.</p> <p>Подп.</p> <p>Дат</p> </div> <div> <p>ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)</p> </div> <div> <p>Лист</p> <p>29</p> </div> </div>				

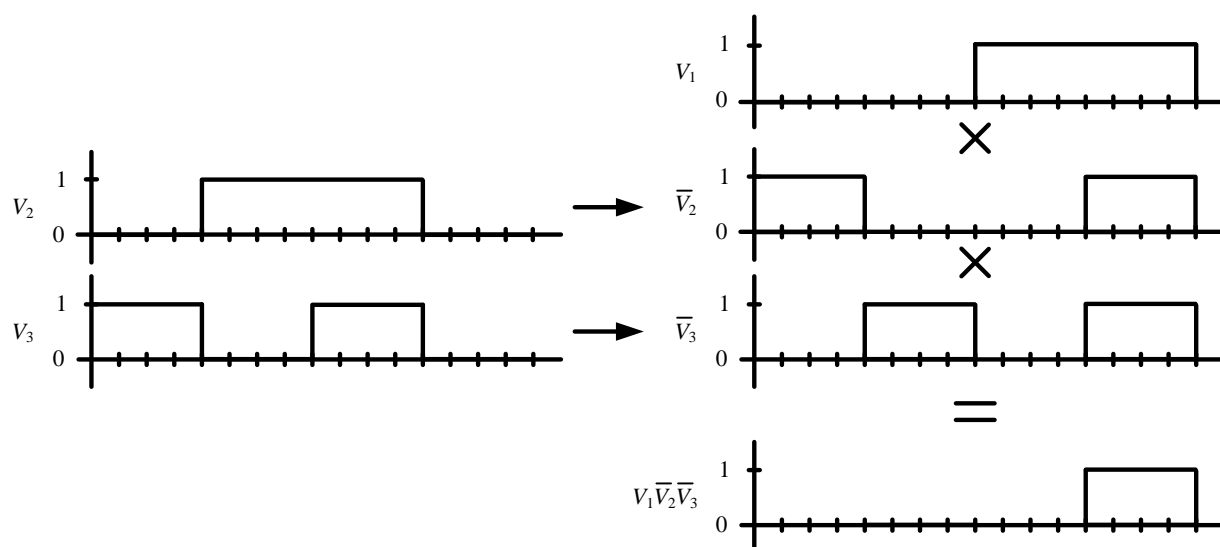


Рисунок 2.11. Вычисление участка сигнала с максимальной массой

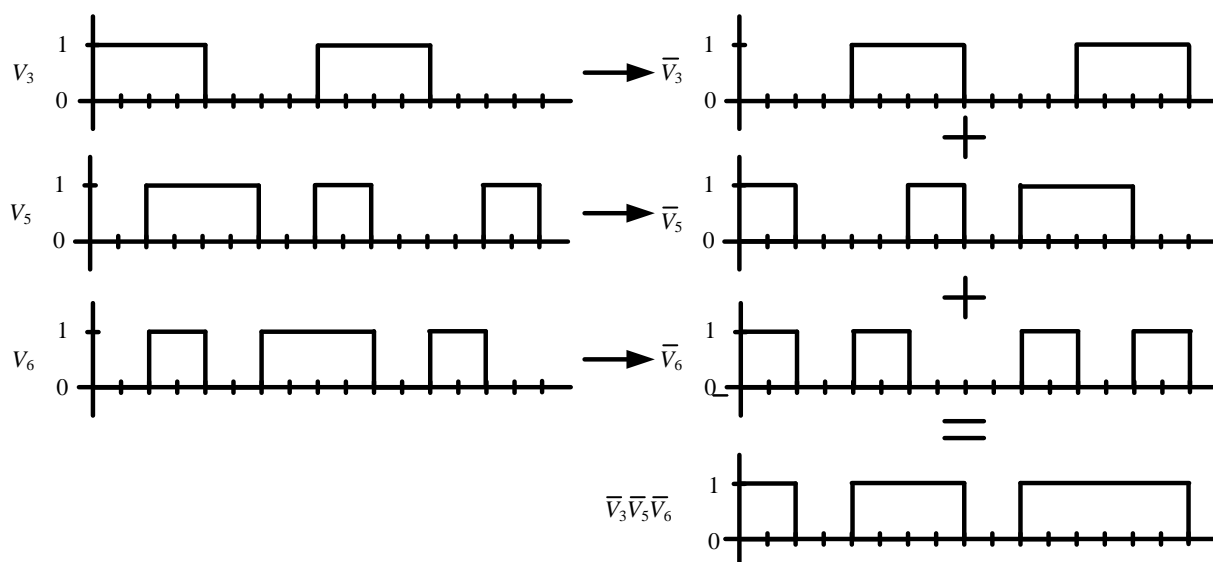


Рисунок 2.12. Вычисление участка сигнала с минимальной массой

Инв. № подл.	Подп. и дата	Инв. № дубл.	Взам. инв. №	Подп. и дата					
Ли	Изм.	№ докум.	Подп.	Дат	ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)				
					Лист 30				

На рис. 2.12 показан пример определения участка сигнала с минимальной массой с помощью полной группы на операции сложения.

Несмотря на то, что с учётом возможных инверсий существует всего 140 вариантов полных групп ($140 = 35 \cdot 4$), для рассматриваемого сегмента сигнала всегда срабатывает только одна из четырёх групп.

2.3.2. Замкнутые группы

Свойства замкнутых групп

1. Каждая замкнутая группа образована на паре полных групп путем их склейки на общем операторе. Поэтому она позволяет выявлять корреляционные (энергетические) связи между данными анализа этих полных групп. Результатом является иерархия анализа спектральных компонент (откликов фильтров);
2. Каждая замкнутая группа образована на паре полных групп путем их склейки на общем операторе. В качестве такого оператора может выступать любой из трех операторов некоторой новой полной группы P_{ns} (коммутирующей), элементы которой образуют полные группы с исходными, входящими в состав замкнутой группы. Результатом коммутации является выбор трех вариантов структурно связанных подобластей в составе образа замкнутой группы;
3. Для замкнутой группы $V_i V_j V_k V_l$ существует восемь вариантов записи с учётом инверсий: $V_i V_j + V_k \bar{V}_l$, $V_i V_j + \bar{V}_k V_l$, $V_i \bar{V}_j + V_k V_l$, $\bar{V}_i V_j + V_k V_l$, $V_i \bar{V}_j + \bar{V}_k \bar{V}_l$, $\bar{V}_i V_j + \bar{V}_k \bar{V}_l$, $\bar{V}_i \bar{V}_j + V_k \bar{V}_l$, $\bar{V}_i \bar{V}_j + \bar{V}_k V_l$;
4. Если один из операторов замкнутой группы – инверсный, то сумма операторов представляет собой единичный вектор;
5. Образы замкнутых групп, получаемые при разном порядке операторов – одинаковые;
6. Замкнутая группа выделяет сегмент сигнала, на котором сумма отсчётов сигнала максимальна;
7. Для замкнутого множества $V_i V_j V_k V_l$ существует восемь вариантов записи с учётом инверсий: $V_i V_j V_k V_l$, $\bar{V}_i \bar{V}_j V_k V_l$, $\bar{V}_i V_j \bar{V}_k V_l$, $V_i \bar{V}_j \bar{V}_k V_l$, $\bar{V}_i V_j V_k \bar{V}_l$, $V_i \bar{V}_j V_k \bar{V}_l$, $\bar{V}_i \bar{V}_j \bar{V}_k \bar{V}_l$, $V_i V_j \bar{V}_k \bar{V}_l$ (знак операции: сложение или умножение, в этой записи не указан);

Алгоритм генерации замкнутых групп

Алгоритм построения замкнутой группы:

1. выбирается полная группа и формируется четыре её варианта на операции умножения (с инверсиями): $V_i V_j V_k$, $\bar{V}_i \bar{V}_j V_k$, $\bar{V}_i V_j \bar{V}_k$, $V_i \bar{V}_j \bar{V}_k$;
2. полученные варианты образов полных групп отображаются на прямой (см. рис. 2.13);
3. образ полной группы, который в области оператора V_0 принимает ненулевое значение, не может использоваться при формировании замкнутой группы;

4. проанализировав варианты отображения полных групп на прямую, формируются замкнутые группы: замкнутую группу составляют операторы, в области которых образ полной группы принимает ненулевое значение.

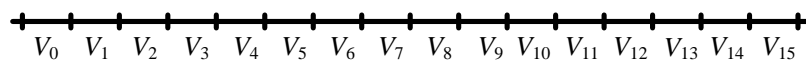


Рисунок 2.13. Расположение операторов на прямой

Рассмотрим полную группу $V_1 V_8 V_9$. Возможны следующие варианты записи данной группы с учётом инверсий: $V_1 V_8 V_9$, $\bar{V}_1 \bar{V}_8 V_9$, $\bar{V}_1 V_8 \bar{V}_9$, $V_1 \bar{V}_8 \bar{V}_9$. Образ каждого варианта группы показан на рис. 2.14.

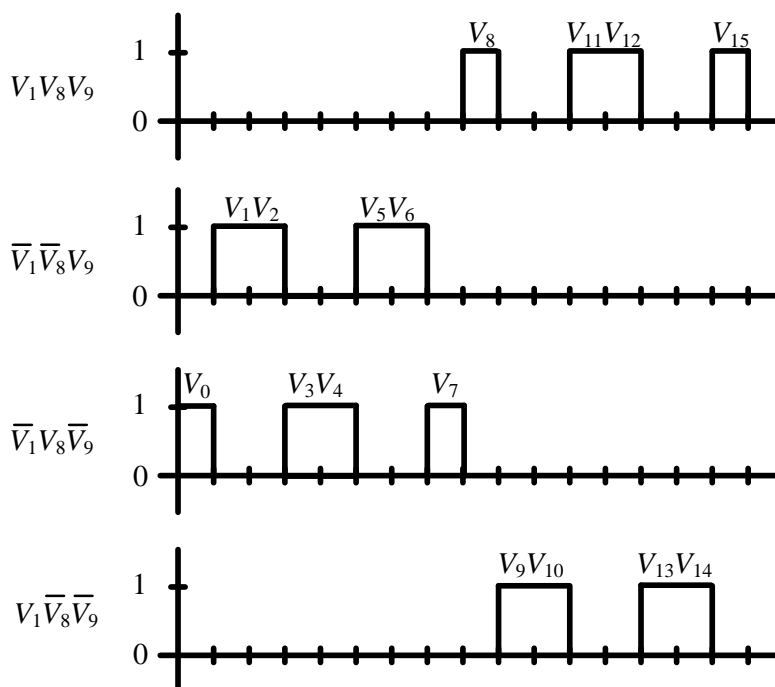


Рисунок 2.14. Образы полных групп

Алгоритм вычисления замкнутых групп, описывающих сигнал

Рассмотрим алгоритм вычисления замкнутых групп, описывающих сигнал:

1. вычислить для сигнала f спектр;
2. каждому элементу спектра ставится в соответствие оператор из множества $\{V_i\}$ (элементы спектра нумеруются с нуля, номера операторов и элементов спектра совпадают).
3. выполняется проверка наличие i -ой замкнутой группы в сигнале f :

- (а) из спектра выбираются четыре элемента, номера которых соответствуют номерам операторов, входящих в i -ую группу;
- (б) определяется количество инверсных операторов: если значение j -го элемента спектра меньше нуля, то считается, что соответствующий элементу оператор – инверсный;
- (с) рассматривается две ситуации:
- число инверсных операторов чётно: вычисляется образ замкнутого множества. Пусть из спектра выбраны четыре элемента, соответствующие операторам V_8, V_9, V_{10}, V_{11} . На рис. 2.18 показан пример вычисления образа замкнутого множества на операции сложения ($V_8 + V_9 + V_{10} + V_{11}$), и на операции умножения ($V_8 V_9 V_{10} V_{11}$) (см. рис. 2.16);
 - число инверсных операторов нечётно: вычисляется образ замкнутой группы на операции сложения. На рис. 2.17 показан пример вычисления образа замкнутой группы $\bar{V}_8 V_9 + V_{10} V_{11}$ (см. рис. 2.15).

4. вычисляется масса группы.

V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}	V_{13}	V_{14}	V_{15}
100	-12	-6	4	3	9	11	23	-2	5	21	11	3	0	7	8

Рисунок 2.15. Спектр сигнала f (для определения замкнутой группы)

V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}	V_{13}	V_{14}	V_{15}
100	-12	-6	4	3	9	11	23	2	5	21	11	3	0	7	8

Рисунок 2.16. Спектр сигнала f (для определения замкнутого множества)

Шаги 3-4 выполняются для всех 140 замкнутых групп. В результате, формируется три массива масс замкнутых групп: для масс замкнутых групп, замкнутых множеств на операции сложения и замкнутых множеств на операции умножения.

Каждой замкнутой группе (замкнутому множеству), присутствующей в сигнале, в зависимости от числа инверсий, можно поставить в соответствие число от 1 до 840. Данная операция выполняется отдельно для замкнутых групп и замкнутых множеств.

2.3.3. Анализ полных и замкнутых групп

Анализируя спектры сигналов, можно отметить, что не для каждого сигнала можно вычислить полную и замкнутую группу.

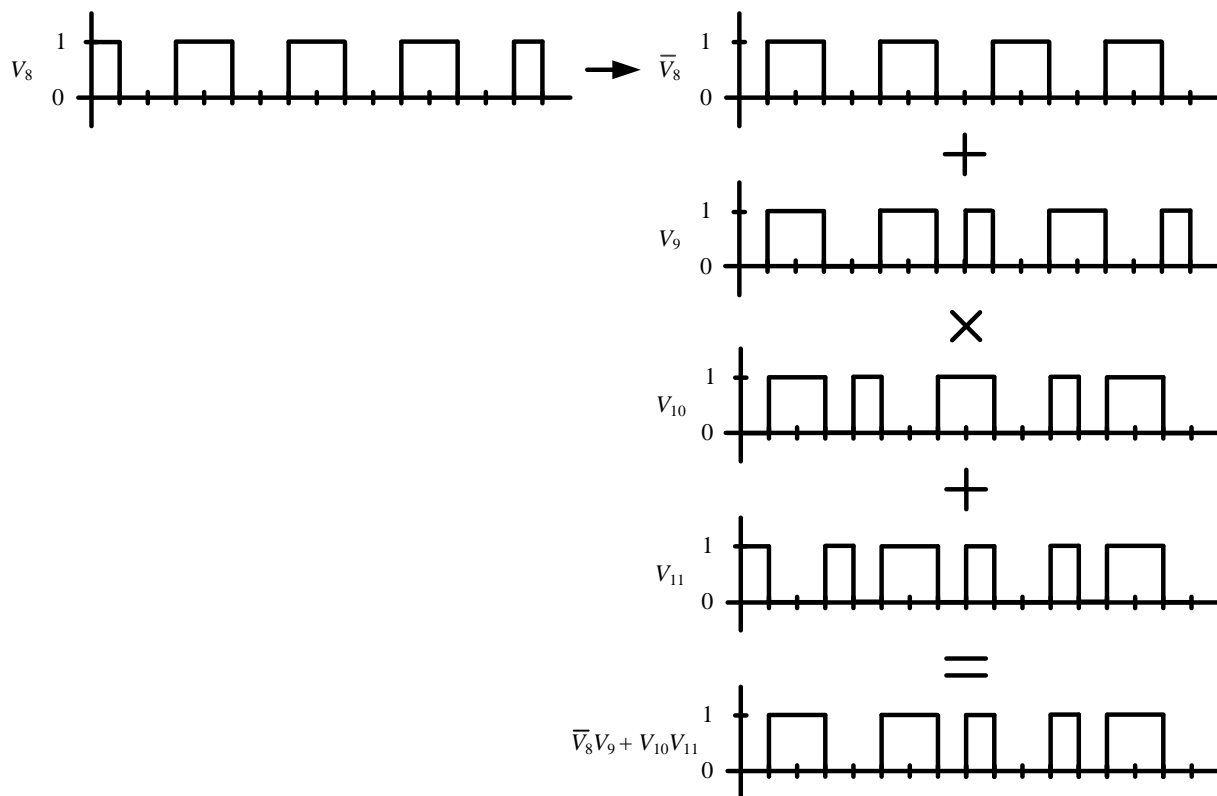


Рисунок 2.17. Вычисление образа замкнутой группы

На рис. 2.19 и рис. 2.20 схематично показаны типы сигналов, по которым нельзя вычислить одновременно полную группу (на операции сложения и умножения) и замкнутую группу. Рядом со схематичным изображением приведён пример сегмента реального сигнала.

По сигналу первого типа можно вычислить замкнутую группу, но нельзя вычислить полную (ни на операции сложения, ни на операции умножения), по сигналу второго типа можно вычислить полную группу на операции сложения, но нельзя вычислить полную группу на операции умножения и замкнутую группу.

Доля указанных сегментов сигнала составляет около 1% от доли всех сегментов сигнала. Также можно отметить, что чем больше длина сегментов, на которые разбивается сигнал, тем меньше спектров указанных типов.

Данное определение справедливо как для полных, так и для замкнутых групп.

2.4. Реализация системы монауральной локализации направления на источник звука

2.4.1. Предварительная обработка

Предварительная обработка сигнала, с позиций теории активного восприятия, заключается в выполнении операции интегрирования (см. рис. 2.21). На данном этапе обработки анализируемый сигнал f разбивается на M сегментов, по каждому из которых вычисляется Q -преобразование:

$$g(i) = Q[h_i], g(i) = \sum_{k=1}^L h_i(k), i = \overline{1, N}, k = \overline{1, M},$$

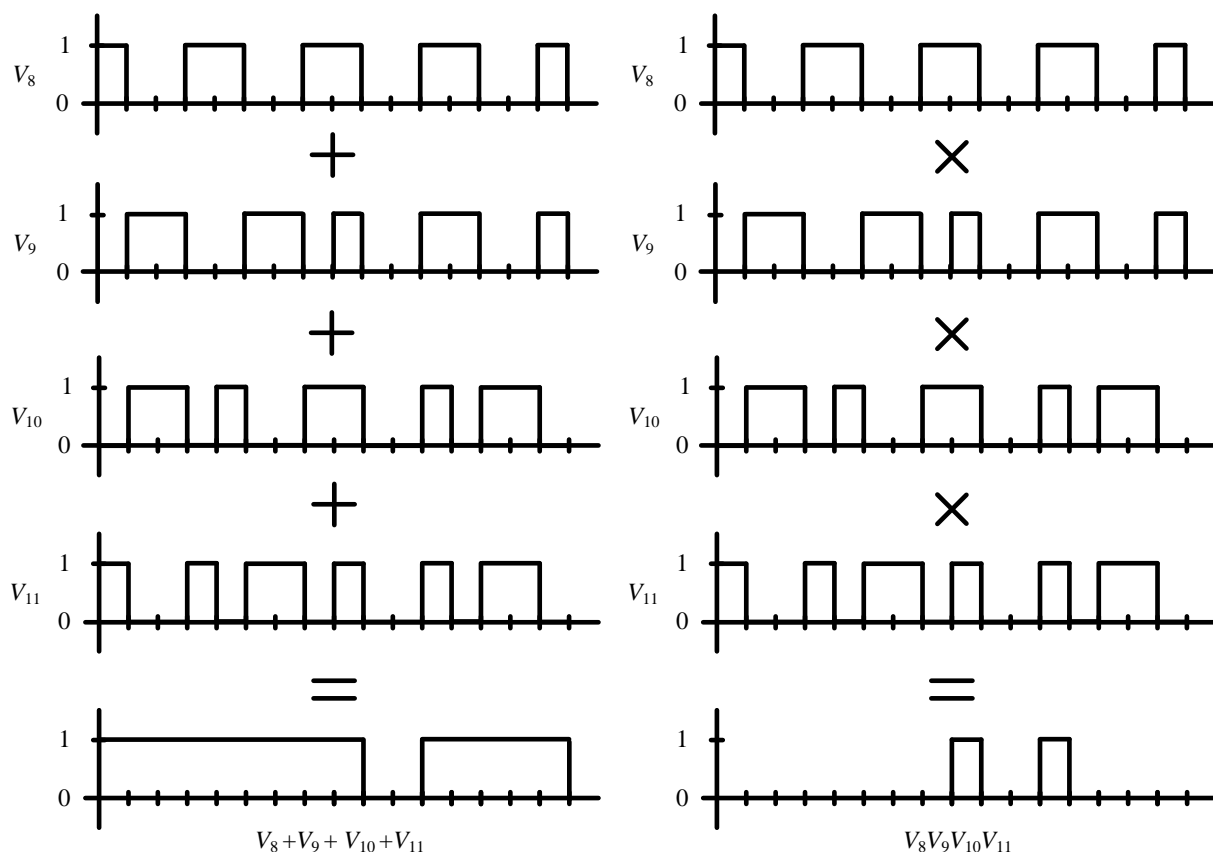


Рисунок 2.18. Вычисление образа замкнутой группы

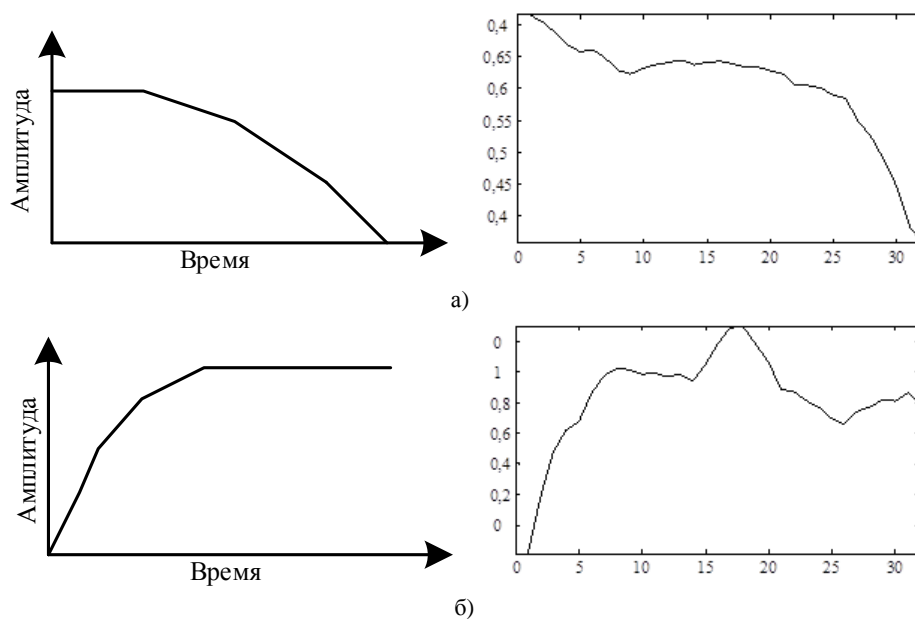
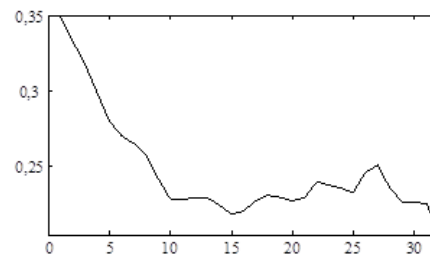
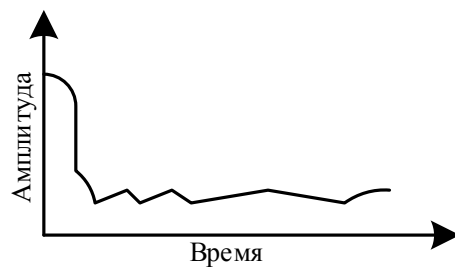
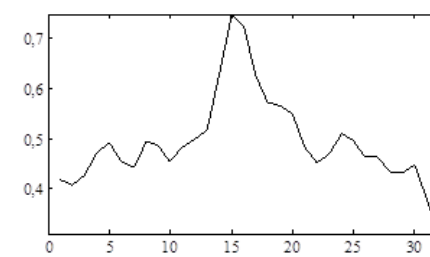
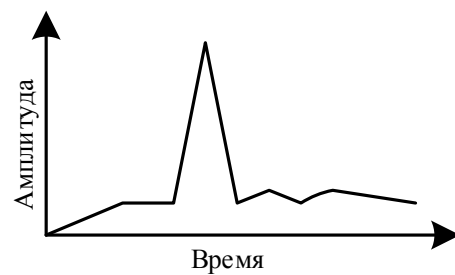


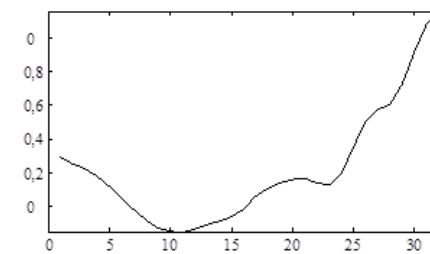
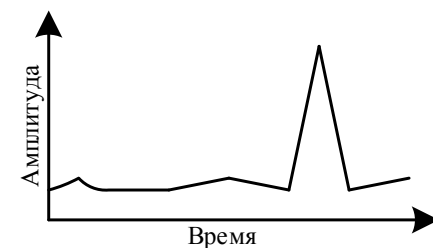
Рисунок 2.19. Первый тип сигналов



а)



б)



в)

Рисунок 2.20. Второй тип сигналов

Таблица 2.1. Возможные сочетания групп

№	Полная группа на операции умножения	Полная группа на операции сложения	Замкнутая группа	Частота появления	Рисунок
1	-	-	-	0	-
2	+	-	-	0	-
3	-	+	-	1	рис. 2.19
4	+	+	-	0	-
5	-	-	+	1	рис. 2.20
6	+	-	+	0	-
7	-	+	+	0	-
8	+	+	+	98	-

где N – число отсчётов в сигнале g , $h = \{h_i\}$, h – множество сегментов, вычисленных по сигналу f , L – количество отсчётов в сегменте. Таким образом, на следующий этап, этап вычисления признаков, передаётся сигнал g .

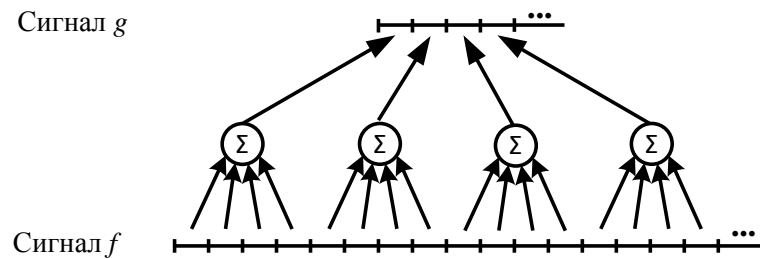


Рисунок 2.21. Предварительная обработка сигнала

2.4.2. Формирование системы признаков

Формирование описания звукового сигнала с помощью алгебры изображения и алгебры групп

Исходными данными для формирования описания звукового сигнала f с помощью алгебры изображения и алгебры групп является спектральное представление D данного сигнала:

$$\mathbf{V} = GV[D],$$

$$\mathbf{P}_{na} = GP_{na}[D, \mathbf{V}],$$

$$\mathbf{P}_{nm} = GP_{nm}[D, \mathbf{V}],$$

$$\mathbf{P}_s = GP_s[D, \mathbf{V}, \mathbf{P}_{na}, \mathbf{P}_{nm}],$$

$$\mathbf{P}_c = GP_c[D, \mathbf{V}, \mathbf{P}_{na}, \mathbf{P}_{nm}],$$

где GV – оператор вычисления операторов, GP_{na} – оператор вычисления полных групп на операции сложения, GP_{nm} – оператор вычисления полных групп на операции умножения, GP_s – оператор вычисления замкнутых групп, GP_c – оператор вычисления замкнутых множеств, \mathbf{V} – множество значений операторов, вычисленных по спектральному представлению D , \mathbf{P}_{na} – множество значений полных групп на операции сложения, \mathbf{P}_{nm} – множество значений полных групп на операции умножения, \mathbf{P}_s – множество значений замкнутых групп, \mathbf{P}_c – множество значений замкнутых множеств.

Признаковое описание сигнала в виде гистограммы

Рассмотрим метод, предлагаемый для создания признакового описания в виде гистограммы:

1. отсчёты сигнала g разбиваются на множество сегментов $g = \{g_k\}$, длиной 16 отсчётов, со смещением в S отсчётов;

Инв. № подл.	Подп. и дата					
	Взам. инв. №					
	Инв. № дубл.					
	Подп. и дата					
	Инв. № подл.					
Ли	Изм.	№ докум.	Подп.	Дат	ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)	Лист
						37

2. к каждому сегменту g_k применяется преобразование дифференцирования, реализованное с помощью системы фильтров, в результате формируется спектральное представление каждого сегмента: $u_k = dU[g_k]$, $u = \{u_k\}$, где dU – оператор вычисления преобразования дифференцирования;
3. по вычисленному спектральному представлению u_k сегмента g_k формируется описание с помощью одной или нескольких структур, входящих в алгебру групп. В алгебре групп существуют следующие структуры, которые могут использоваться для создания описания:
 - (а) операторы (V , значимыми являются 15 элементов из 16, зная, что операторы могут принимать значения -1, 0 и 1, пространство признаков на основе операторов включает 45 элементов; учитывая, что нулевые значения операторов встречаются не часто, можно учитывать только значения -1 и 1, и следовательно, сократить пространство признаков до 30 элементов);
 - (б) полные группы (P_n), известны полные группы на операции умножения (P_{nm} , 140 элементов), полные группы на операции сложения (P_{na} , 140 элементов);
 - (с) замкнутые группы (P_s , 840 элементов), замкнутые множества (P_c , 840 элементов). Допустимо использование сочетаний данных структур;
4. для объединения данных, полученных от разных сегментов анализируемого сигнала, вычисляется гистограмма элементов структур, использованных при создании описания сегмента (см. рис. 2.22):

$$h_V = H[V, \Gamma],$$

$$h_{na} = H[P_{na}, \Gamma],$$

$$h_{nm} = H[P_{nm}, \Gamma],$$

$$h_s = H[P_s, \Gamma],$$

$$h_c = H[P_c, \Gamma],$$

$$h_{nam} = H[P_{na}, P_{nm}, \Gamma],$$

$$h_{sc} = H[P_s, P_c, \Gamma],$$

где h_V – гистограмма операторов, h_{na} – гистограмма полных групп на операции сложения, h_{nm} – гистограмма полных групп на операции умножения, h_s – гистограмма замкнутых групп, h_c – гистограмма замкнутых множеств, h_{nam} – гистограмма полных групп на операции сложения и умножения, H – оператор вычисления гистограммы заданной размерности, Γ – размерность гистограммы: $1d$ – одномерная гистограмма, $2d$ – двумерная гистограмма, $3d$ – трёхмерная гистограмма. В двумерной гистограмме учитываются возможные появления пар групп в описании одного сегмента сигнала, в трёхмерной – троек.

Ниже приведены алгоритмы формирования гистограмм различных размерностей. Признаковое описание, на основе которого формируется гистограмма h_{Desc} обозначается как $Desc$. В качестве данного описания могут использоваться операторы, полные и замкнутые группы.

Алгоритм формирования одномерной гистограммы признаков:

Инв. № подл.	Подп. и дата	Взам. инв. №	Подп. и дата	Инв. № дубл.	Подп. и дата	<div style="text-align: left; padding-left: 10px;"> <p>ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)</p> </div>	Лист
							38
Ли	Изм.	№ докум.	Подп.	Дат			

$$\forall i = \overline{1, M}$$

$$\forall j = \overline{1, |Desc_i|}$$

$$h_{Desc}(Desc_i(j)) = h_{Desc}(Desc_i(j)) + 1.$$

Алгоритм формирования двумерной гистограммы признаков:

$$\forall i = \overline{1, M}$$

$$\forall j = \overline{1, |Desc_i|}$$

$$\forall k = \overline{j + 1, |Desc_i|}$$

$$h_{Desc}(Desc_i(j), Desc_i(k)) = h_{Desc}(Desc_i(j), Desc_i(k)) + 1.$$

Алгоритм формирования трёхмерной гистограммы признаков:

$$\forall i = \overline{1, M}$$

$$\forall j = \overline{1, |Desc_i|}$$

$$\forall k = \overline{j + 1, |Desc_i|}$$

$$\forall l = \overline{k + 1, |Desc_i|}$$

$$\begin{aligned} h_{Desc}(Desc_i(j), Desc_i(k), Desc_i(l)) = \\ = h_{Desc}(Desc_i(j), Desc_i(k), Desc_i(l)) + 1. \end{aligned}$$

После формирования гистограммы можно выполнить нормализацию её значений к отрезку $[0; 1]$:

$$h' = NORM[h],$$

где $NORM$ – оператор нормализации гистограммы.

Таким образом, гистограмма используется для объединения результатов вычисления признаков от нескольких сегментов, что позволяет получить представление о сигнале в целом.

Ранее в рамках работ [35] и [36] для решения задачи идентификации диктора также были предложены системы признаков. В [35] для создания описания сигнала по каждому его сегменту вычислялось три группы, максимальные по массе (полная группа на операции сложения, полная группа на операции умножения и замкнутая группа). Далее, выбранные группы по всем сегментам объединялись в гистограмму, которая и представляла собой описание сигнала. На этапе классификации использовался метод сравнения с эталоном, мера близости – расстояние Евклида. В [36] для создания описания сигнала также вычислялись три группы максимальные по массе, затем найденные группы сохранялись с учётом временного порядка. На этапе классификации выполнялся поиск похожих цепочек групп между классифицируемым сигналом и эталоном. Максимальное число совпадений указывало на класс анализируемого сигнала. Точность классификации на основе описанных методов была ниже, чем на основе предложенного в настоящей работе. Это связано с тем, что указанные методы не учитывали большое количество информации (из-за учёта только одной группы с максимальной массой).

Можно провести аналогию между предложенной системой признаков и гистограмма ориентированных градиентов, системой признаков, используемой в обработке изображений (histogram of

Подп. и дата	
Взам. инв. №	
Инв. № дубл.	
Подп. и дата	
Инв. № подл.	

Ли	Изм.	№ докум.	Подп.	Дат

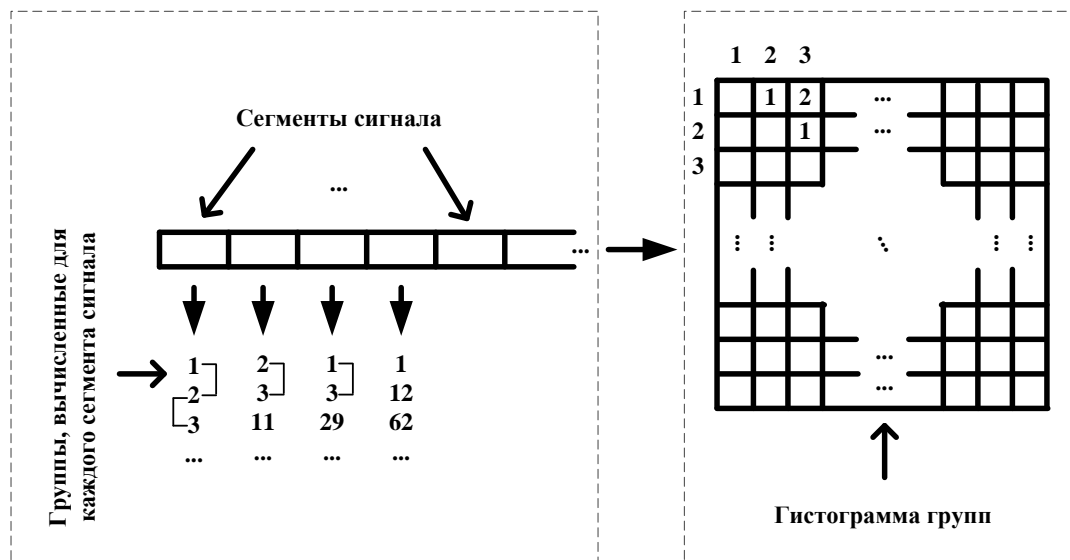


Рисунок 2.22. Формирование двумерной гистограммы

oriented gradients) [37–39]. В области обработки звуковых сигналов аналогом предложенного подхода является использование для описания сигнала мел-частотных кепстральных коэффициентов совместно с моделью гауссовых смесей.

В табл. 2.2 приведена оценка размерности предлагаемых систем признаков.

Таблица 2.2. Оценка размерности систем признаков

Система признаков / Размерность	1d	2d	3d
h_V	1×30	30×30	$30 \times 30 \times 30$
h_{nia}	1×140	140×140	$140 \times 140 \times 140$
h_{nim}	1×140	140×140	$140 \times 140 \times 140$
h_{si}	1×840	840×840	$840 \times 840 \times 840$
h_{ci}	1×840	840×840	$840 \times 840 \times 840$
h_{niam}	1×280	280×280	$280 \times 280 \times 280$
h_{sci}	1×1680	1680×1680	$1680 \times 1680 \times 1680$

Обобщённая схема формирования системы признаков показана на рис. 2.23.

Пример формирования признакового описания сигнала в виде одномерной гистограммы для полных групп размером 1×140 элементов показан на рис. 2.24.

Пример формирования признакового описания сигнала в виде двумерной гистограммы размером 140×140 элементов для полных групп показан на рис. 2.25.

Признаковое описание сигнала на основе вероятностей переходов

Описание звукового сигнала, представленное в виде гистограммы (см. пункт 2.4.2) рассматривает соседние сегменты h_i и h_{i+1} как не связанные друг с другом. В связи с этим предлагается



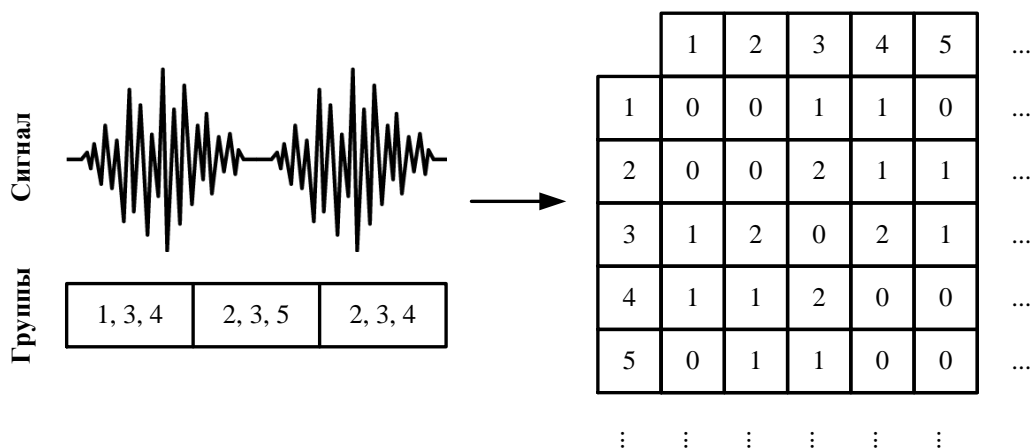


Рисунок 2.25. Пример формирования двумерной гистограммы

модель признакового описания, в которой учитываются связи между соседними сегментами звукового сигнала. Метод вычисления признакового описания в таком случае состоит в формировании матрицы вероятностей переходов между описаниями соседних сегментов.

Предлагаются следующие системы признаков, основанные на матрицах вероятностей переходов:

1. система признаков, описывающая вероятности переходов между значениями операторов (оператор может принимать три возможных значения: прямое, инверсное и равное нулю), вычисленными по соседним сегментам сигнала, без учёта связей между различными операторами, размерность пространства признаков – $3 \times 3 \times 15$ (см. рис. 2.26);
2. система признаков, описывающая вероятности переходов между значениями операторов, вычисленными по соседним сегментам сигнала, с учётом связей между операторами, размерность пространства признаков – 45×45 (см. рис. 2.27);
3. система признаков, описывающая вероятности переходов между описаниями сегментов, представленных в виде полных групп, размерность пространства признаков – 140×140 ; при использовании полных групп допустимо использовать только несколько максимальных по массе групп (граф переходов для полных групп подобен графу переходов для операторов, см. рис. 2.27);
4. система признаков, описывающая вероятности переходов между описаниями сегментов, представленных в виде замкнутых групп, размерность пространства признаков – 840×840 ; при использовании замкнутых групп допустимо использовать только несколько максимальных по массе групп.

Алгоритм формирования системы признаков на основе операторов без учёта связей ($Desc_i$ – описание i -го сегмента сигнала в виде операторов):

Инв. № подл.						Подп. и дата
Инв. № подл.						Подп. и дата
Инв. № подл.						Подп. и дата
Инв. № подл.						Подп. и дата

Ли

Изм.

№ докум.

Подп.

Дат

ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)

Лист
42

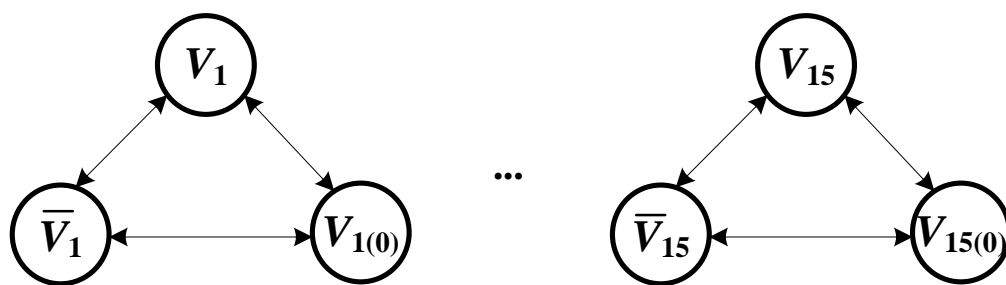


Рисунок 2.26. Графы переходов между операторами

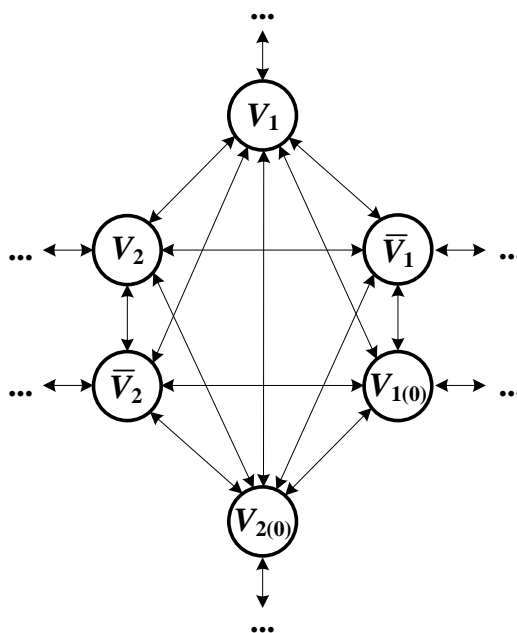


Рисунок 2.27. Графы переходов между операторами

$$M_{Desc}(p, q) = M_{Desc}(p, q) + 1.$$

Алгоритм формирования системы признаков на основе полных (замкнутых групп, $Desc_i$ – описание i -го сегмента сигнала в виде полных или замкнутых групп):

$$\forall i = \overline{1, M}$$

$$\forall k = \overline{1, |Desc_i|}$$

$$\forall l = \overline{1, |Desc_{i+1}|}$$

$$M_{Desc_i}(Desc_i(k), Desc_{i+1}(l)) = M_{Desc_i}(Desc_i(k), Desc_{i+1}(l)) + 1.$$

Примечание: при использовании в качестве систем признаков матриц вероятностей переходов между описаниями сегментов можно учитывать вероятности переходов не только между i и $(i + 1)$ сегментом сигнала, но и учитывать связи между большим числом сегментов. Для i -го сегмента возможен учёт не только $(i + 1)$, $(i + 2)$ и дальнейших сегментов, но и $(i - 1)$, $(i - 2)$ сегментов, т.е. не только «будущего», но и «прошлого».

Пример формирования признакового описания сигнала в виде матрицы вероятностей переходов размером 140×140 элементов для полных групп показан на рис. 2.28. При вычислении значений матрицы рассматривались связи между описания только пары соседних сегментов.

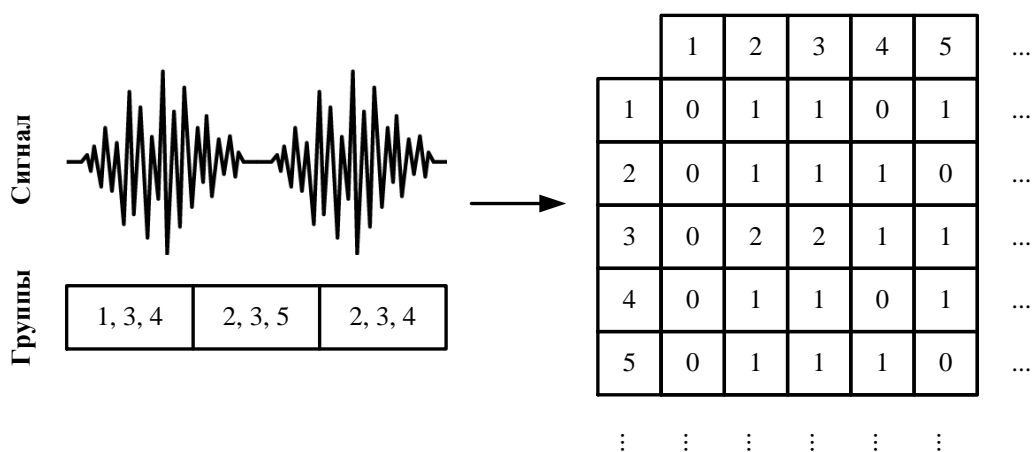


Рисунок 2.28. Пример формирования двумерной матрицы вероятностей переходов для полных групп

Пример формирования признакового описания сигнала в виде 15 независимых матриц вероятностей переходов для операторов на показан рис. 2.29. При вычислении значений матрицы рассматривались связи между описания только пары соседних сегментов и учитываются вероятности переходов только для одного оператора.

Пример формирования признакового описания сигнала в виде матрицы вероятностей переходов для операторов размером 45×45 показан на рис. 2.30. При вычислении значений матрицы рассматривались связи между описания только пары соседних сегментов. В описаниях сегментов приведены значения только операторов V_1 и V_2 .

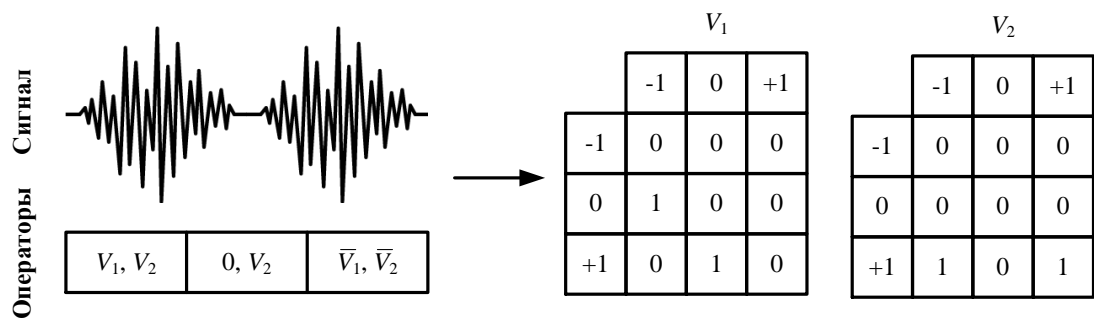


Рисунок 2.29. Пример формирования двумерной матрицы вероятностей переходов для операторов

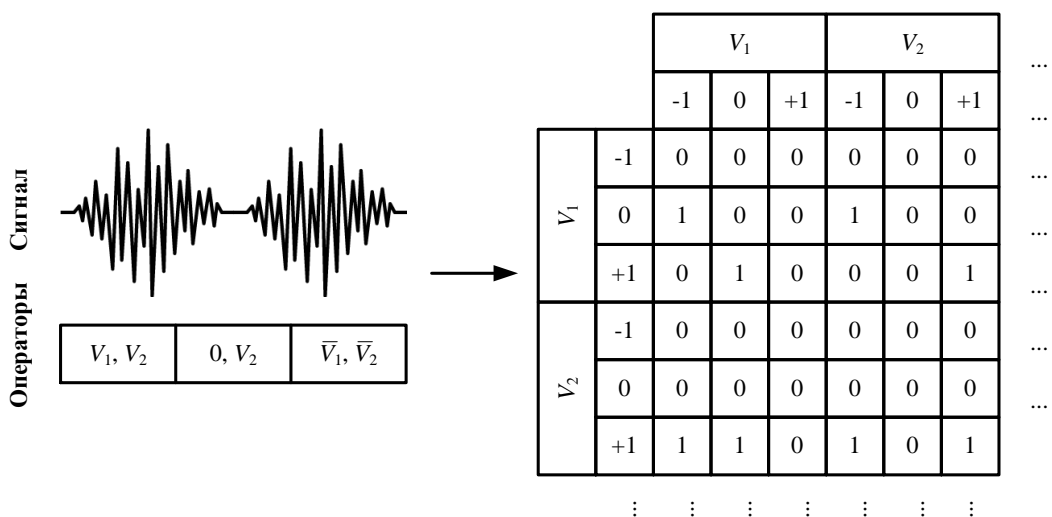


Рисунок 2.30. Пример формирования двумерной матрицы вероятностей переходов для операторов

В пункте 2.4.2 были предложены несколько систем признаков, представленных в виде гистограмм и основанных на операторах, полных и замкнутых группах. Рассматривались различные размерности гистограмм ($N = 2, 3$, где N – размерность гистограммы). При линейном увеличении N размерность пространства признаков увеличивается как показательная функция от N , в связи с этим было принято решение ограничиться только двумерными гистограммами групп и трёхмерной гистограммой операторов.

Двумерные гистограммы, представляющие собой матрицы, формируемые на основе алгоритмов, описанных в пункте 2.4.2, обладают:

1. нулевой главной диагональю;
2. симметрией относительно главной диагонали.

Таким образом, число значащих элементов в двумерной гистограмме составляет $(H_{dim} \cdot H_{dim} - H_{dim})/2$, где H_{dim} – число столбцов (строк) матрицы.

Учитывая несовместность некоторых полных и замкнутых групп возможно сокращение размерности предложенных систем признаков. Сжатие с учётом несовместности возможно только для гистограмм размерностью больших или равных двум.

Таким образом, с учётом приведённых рассуждений, размерности систем признаков на основе гистограмм показаны в табл. 2.3. В скобках указан коэффициент сжатия, рассчитываемый как отношение числа признаков до сжатия к числу признаков после сжатия.

Таблица 2.3. Оценка размерности систем признаков до и после сжатия (для гистограмм)

Система признаков / Размерность	До сжатия			После сжатия		
	$1d$	$2d$	$3d$	$1d$	$2d$	$3d$
h_V	1×30	30×30	$30 \times 30 \times 30$	–	1×435 (2.06)	–
$h_{na}(h_{nm})$	1×140	140×140	–	–	1×5050 (3.88)	–
$h_s(h_c)$	1×840	840×840	–	–	1×106030 (6.65)	–

В пункте 2.4.2 было предложено несколько систем признаков, представленных в виде матриц вероятностей переходов и основанных на операторах, полных и замкнутых группах.

Двумерные квадратные матрицы, формируемые на основе алгоритмов, описанных в пункте 2.4.2, не обладают симметрией относительно главной диагонали.

2.4.3. Принятие решения

Этап классификации может быть реализован с помощью различных классификаторов. Например, с помощью метрического метода классификации k -ближайших соседей и линейного метода опорных векторов.

Решающее правило метода k -ближайших соседей записывается следующим образом:

$$a(u; X^l, k) = \arg \max_{y \in Y} \sum_{i=1}^k [y_u^{(i)} = y],$$

где u – классифицируемый объект, k – параметр алгоритма (количество соседей), $X^m = \{(x_1, y_1), \dots, (x_m, y_m)\}$ – обучающая выборка, заданная в формате «объект-ответ», $Y = \{y_i\}$, $y = \overline{1, C}$ – множество классов, C – количество классов. Определение близости между объектами x и x' выполняется с помощью расстояния Евклида:

$$\rho(x, x') = \sum_{i=1}^M \sqrt{(x_i - x'_i)^2}.$$

Оптимальное значение параметра k определим по критерию скользящего контроля с исключением объектов по одному (leave-one-out, LOO):

$$LOO(k, X^l) = \sum_{i=1}^l [a(x_i; X^l \setminus x_i, k) \neq y_i] \rightarrow \min_k.$$

Известны различные методы вычисления расстояния между гистограммами, например, корреляционный метод, метод на основе вычисления пересечения между гистограммами, расстояние Бхатачария, метод хи-квадрат [40, 41]. Эксперименты показали, что данные методы дают результаты подобные методу на основе расстояния Евклида, которое является менее сложным, чем перечисленные.

Решающее правило метода опорных векторов выглядит следующим образом:

$$a(x) = \text{sign} \left(\sum_{j=1}^n w_j x^j - w_0 \right),$$

где $x = (x^1, \dots, x^n)$ – признаковое описание объекта x , вектор $w = (w^1, \dots, w^n)$ и скалярный порог w_0 являются параметрами алгоритма. Метод опорных векторов является бинарным классификатором. В данной работе для решения задачи мультиклассовой классификации используются два способа сведения данной задачи к бинарной [42]:

1. подход «один-против-всех» (One-vs-All) заключается в обучении N классификаторов по следующему принципу:

$$f_i(x) = \begin{cases} \geq 0, & \text{если } y(x) = i, \\ < 0, & \text{если } y(x) \neq i. \end{cases}$$

которые отделяют каждый класс от остальных. Далее, для каждого $x \in X$ вычисляются все классификаторы и выбирается класс, соответствующий классификатору с наибольшим значением:

$$a(x) = \arg \max_{i \in \overline{1, N}} f_i(x);$$

2. подход «один-против-одного» (One-vs-One) заключается в формировании $N \cdot (N - 1)$ классификаторов, которые разделяют объекты пар различных классов:

$$f_{ij}(x) = \begin{cases} +1, & \text{если } y(x) = i, \\ -1, & \text{если } y(x) = j. \end{cases}$$

Инв. № подл.	Подп. и дата					Лист 48
	Взам. инв. №					
	Инв. № дубл.					
	Подп. и дата					
Ли	Изм.	№ докум.	Подп.	Дат	ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)	

После обучения бинарных классификаторов решение принимается следующим образом:

$$a(x) = \arg \max_{i \in \overline{1, N}} \sum_{j=1, j \neq i}^N f_{ij}(x);$$

При классификации используется линейное ядро:

$$k(x, y) = x^T y + c.$$

2.5. Информационная модель локализации источника звука

Сущность информационной модели локализации источника звука поясняется графическим материалом на рисунке 2.31.

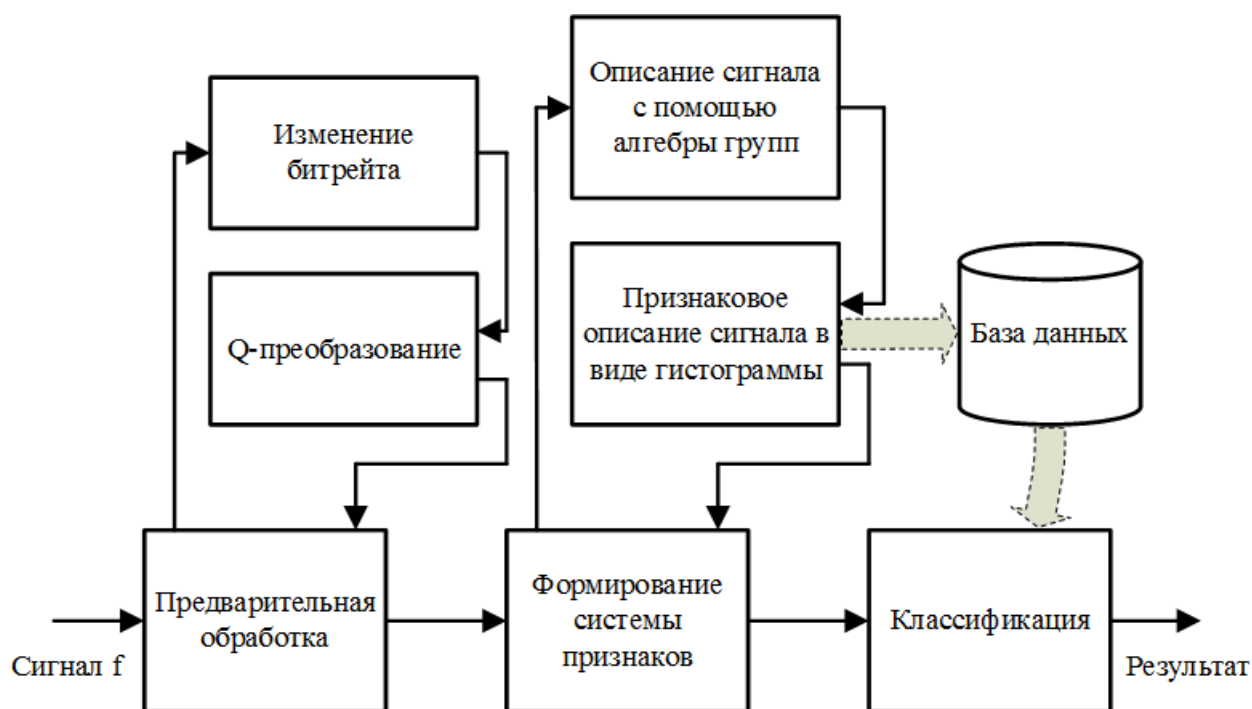


Рисунок 2.31. Сущность информационной модели локализации источника звука.

Информационная модель локализации источника звука содержит блок предварительной обработки сигнала (сигнал поставляется в программу в виде отдельных аудио файлов формата *.WAV), который выполняет роль преобразователя основных параметров звуковой записи (битрейт, частота дискретизации), в соответствии с параметрами, заданными при работе.

Далее, сигнал из блока предварительной обработки следует в блок формирования системы признаков. В этом блоке сигнал последовательно разбивается на отрезки, количество и продолжительность которых заданы в параметрах программы, проходит операцию U-преобразования, а затем для каждого из отрезков составляется признаковое описание, в соответствии с результатами, полученными на этапе U-преобразования. Составленные признаковые описания записываются в базу данных, для дальнейшего использования.

Данные с выхода блока формирования системы признаков поступают на вход блока классификации, где при помощи заранее выбранного механизма классификации происходит распределение всех объектов, описание которых было получено на этапе вычислений предыдущего блока, по

Инв. № подл	Подп. и дата	Инв. № дубл.	Взам. инв. №	Подп. и дата

Ли	Изм.	№ докум.	Подп.	Дат

- ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)

Глава 3. Вычислительный эксперимент

3.1. Описание установки для записи сигналов

Решение задачи монураальной локализации источника звука начинается с разработки макета установки, включающей в себя устройство регистрации (микрофон, подключённый к звуковой карте, смартфон), звуковой отражатель и устройство для поворота микрофона и отражателя на заданный угол (см. рис. 3.1). В качестве устройства регистрации звука используется всенаправленный микрофон, звуковой отражатель представляет собой пластину с неоднородной поверхностью. В качестве воспроизводящего устройства используется широкополосный динамик.

Во время проведения эксперимента использовались два помещения, для тестирования работы установки в различных условиях (см. рис. 3.2). Одна из комнат имела кирпичные стены, частично закрытые ковровым покрытием, другая комната имела заштукатуренные деревянные стены. Вычисления и запись базы данных выполнялись на ноутбуке Asus k42jv, процессор – Intel Core i3 350M (два ядра), SSD, 6 Гб ОП.

Эксперимент проводился со следующими типами звуковых сигналов:

1. аудиозаписи устной речи
2. аудиозаписи устной речи среднего качества
3. аудиозаписи с низким и нерегулярным шумом (звук двигателя мотоцикла)

Первый и основной блок программы - подготовка исходных данных для дальнейшей обработки их с помощью известных алгоритмов тестирования, основанных на машинном обучении. Процесс подготовки данных, по замыслу, должен проводиться в соответствии с положениями теории активного восприятия, соответственно, в первом блоке необходимо программно реализовать всю последовательность приведения обычных данных формата музыкальной записи к гистограммам, наиболее ярко отображающим особенности различных записей и их частей. После обработки данных описанной частью программы, в дело вступают алгоритмы тестирования данных. Так как в готовом виде данные на выходе созданной программы представляют собой массивы с большим количеством близких по структуре элементов, для их классифицирования было решено использовать методы машинного обучения. В данном случае - метод SVM и метод опорных векторов. В готовом виде программа принимает данные в виде аудио записей звукового потока каждого из угловых секторов исследуемой области в горизонтальной плоскости формата *.wav, предоставляя на выходе таблицу неточностей, на которой можно оценить на сколько запись, пришедшая с каждого из исследуемых секторов похожа на записи с соседних секторов или на запись с этого же сектора, но пришедшую в другое время.

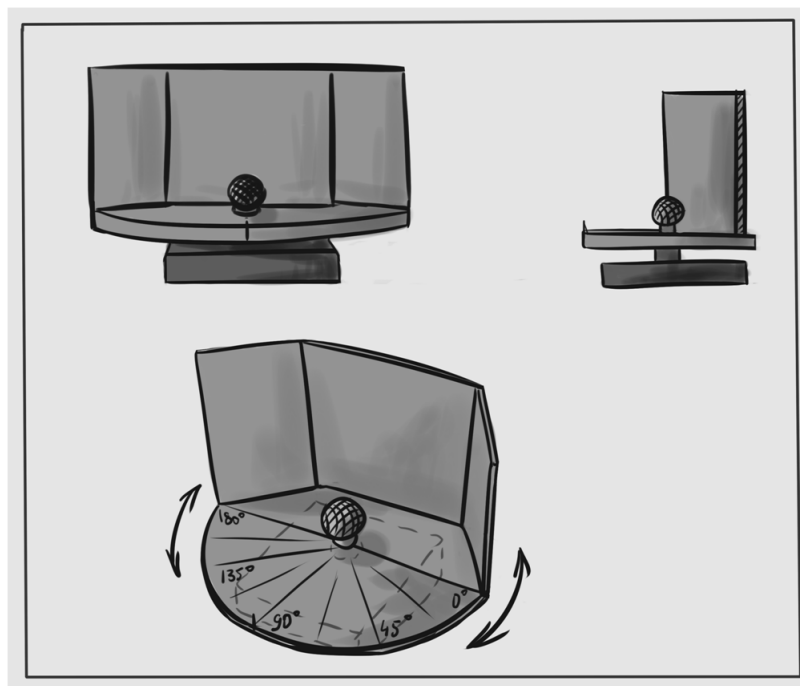


Рисунок 3.1. Макет установки №1.

Этап классификации проводился с помощью нескольких методов: метод k-ближайших соседей и метод опорных векторов.

Рассматривая особенности данных способов классификации данных, можно убедиться, что в условиях исследуемой задачи наилучшим будет применение метода SVM, так как вычисления при классификации ведутся в многомерном пространстве (размерность - несколько сотен). На практике, используя этот метод, после подбора соответствующего ядра удалось достичь хороших результатов в работе.

3.2. Условия проведения эксперимента

Цель, поставленная при проведении эксперимента: исследование возможностей иерархической системы признаков, сформулированной в рамках теории активного восприятия, при решении задач обработки и распознавания звуковых сигналов. Известна модель глубокого обучения, в которой, в качестве признакового описания, используется иерархическая система признаков. По сравнению с ней, предлагаемая иерархическая система признаков предъявляет значительно меньшие требования к аппаратному обеспечению для вычисления признаков. На основе иерархической системы признаков в работе планируется решение задачи моноауральной локализации источника звука.

Достижение поставленной цели проходит с помощью решения задачи разработки, исследования и программной реализации, на основе иерархической системы признаков, метода моноауральной локализации источника звука. Этот метод отличается использованием, в качестве признакового описания сигнала, иерархической системы признаков, обеспечивает снижение материальных и энергетических затрат (по сравнению с бинауральным методом), а так же не уступает по точности известным подходам (например подходу, описанному в работе A.Saxena, A. Y. Ng [2]).

Данная задача включает решение двух подзадач: разработка системы признаков для описания сигнала и разработка макета установки. После каждого внесения изменений в систему признаков

или макет установки должен быть выполнен вычислительный эксперимент. Каждый вычислительный эксперимент предполагает следующие этапы:

1. запись с помощью установки звукового сигнала под разными углами;
2. вычисление признаков записанных сигналов и обучение классификаторов;
3. классификация неизвестных сигналов и определение точности определения направления на источник звука.

На заданных этапах по плану используются следующие классификаторы: метод опорных векторов, деревья решений, метод k-ближайших соседей, нейронные сети. При оценке точности классификации используется метод перекрёстной проверки (cross-validation) и матрицы ошибок (confusion matrix). В ходе выполнения экспериментов используются следующие иерархические системы признаков: одномерные, двумерные и трёхмерные гистограммы полных и замкнутых групп, а также операторов. Возможно, при решении поставленной задачи использование указанных систем признаков не позволит достигнуть заданной точности. В этом случае планируется разработка других иерархических систем признаков, основанных на операторах, полных и замкнутых группах. После завершения тестирования планируется разработка программного обеспечения.

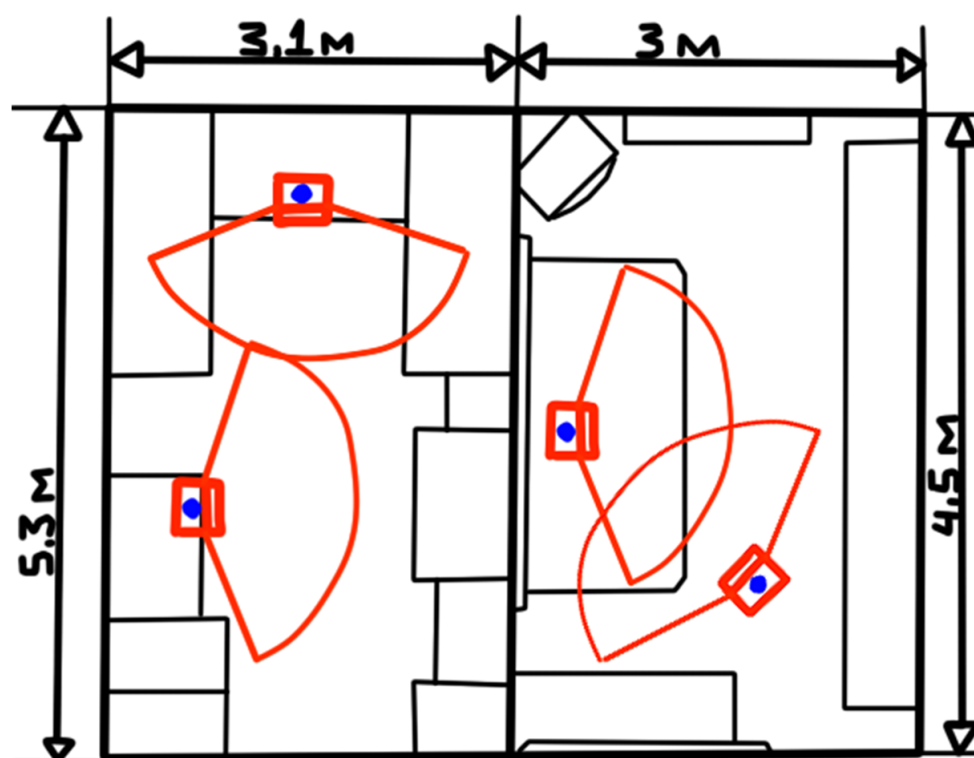


Рисунок 3.2. Помещения для проведения экспериментов с обозначением расположения объектов интерьера и записывающей установки.

После выполнения вычислительного эксперимента проводится анализ полученных результатов с последующим внесением изменений в используемую систему признаков и план экспериментов. Проводимые эксперименты (эксперименты перечисляются в порядке увеличения сложности, проведение каждого эксперимента предусматривает модификацию структуры установки):

Эксперимент № 1.1.

Эксперимент, в котором исследуемый диапазон углов составляет $[0; 180]$ с шагом в 15 градусов. Оценивается эффективность алгоритма, при его применении на различных длинах тестируемых сигналов.

Эксперимент № 1.2.

Исследование зависимости точности результата от длительности сигнала. Нахождение минимальной длительности тестируемого сигнала, при которой система работает с минимальным количеством ошибок. Нахождение минимальной длительностей тестируемого сигнала, при которой сохраняется работоспособность системы.

Эксперимент № 1.3.

Эксперимент с макетной установкой в виде упрощенной модели внешнего уха человека, как наиболее удачного природного улавливателя звуков речевого диапазона (см. рис. 3.3).

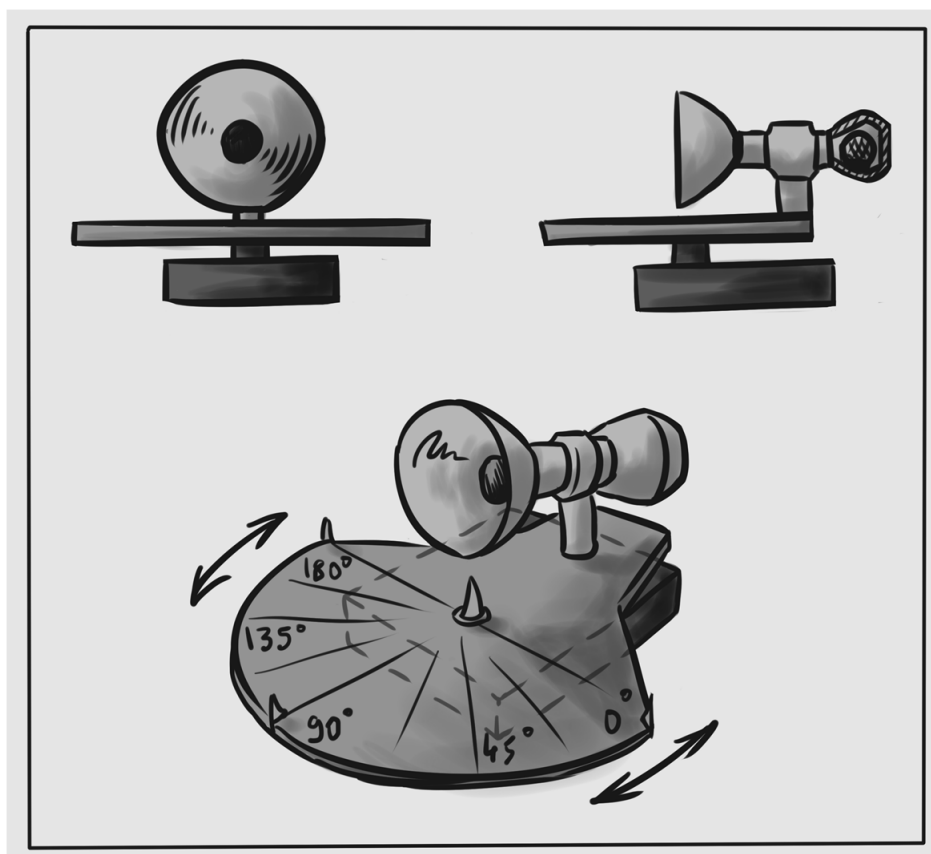


Рисунок 3.3. Макет установки №2. Макет имитирует человеческое ухо, включает в себя небольшой эллиптический экран и канал ≈ 3 сантиметра до микрофона.

Эксперимент № 1.4.

Исследование зависимости точности работы системы от выбранного типа системы признаков, которая используется при составлении описывающей сигнал гистограммы.

При проведении экспериментов № 1.1 - 1.4 условия записи выполняются с минимальным значением отношения сигнал/шум, равным 50 дБ. В результате проведения экспериментов нужно определить систему признаков и структуру макета, позволяющие получить наилучшие результаты по точности определения направления на источник звука. Затем, после отладки системы и предварительного выбора системы признаков, исследования можно будет проводить в различных

Инва. № подл.	Подп. и дата
Инва. № дубл.	Взам. инв. №
Подп. и дата	
Инва. № подл.	

Ли	Изм.	№ докум.	Подп.	Дат	ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)	Лист
						54

Инв. № подл	Подп. и дата	Инв. № дубл.	Взам. инв. №	Подп. и дата

3.3.1. Эксперимент № 1.1.

Каждая запись делится на отрезки продолжительностью 2 секунды.

В данном случае, матрица показывает, сколько двухсекундных отрезков из тестовой записи (например, 1 минута) успешно классифицируются. По такой матрице составляется вывод о том, насколько точна классификация одной минуты записи.

Также матрица приведена к виду т.н. «теплокарты»; чем больший вес имеет ячейка, тем темнее ее цвет. В данном случае это привносит дополнительное удобство в визуальной оценке точности работы системы, становится хорошо различима главная диагональ матрицы и области, в которых произошло наибольшее количество ошибок. Значения масс в ячейках переведены в процентное отношение, относительно всей матрицы.

Приведены результаты эксперимента сигналами по одному с каждого направления, длительно-стью 2 минуты. Запись разбивается на отрезки по 2 сек, количеством по 40 сигналов на запись, обучение на 60 сек, тестирование на 20 сек(см. рис. 3.4).

На рисунке ясно прослеживается главная диагональ матрицы, это говорит о том, что направление на источник звука в большинстве случаев распознается корректно. Из-за того, что отрезок исследуемого сигнала имеет достаточную продолжительность и разбит на большое количество классифицируемых отрезков, ячейки главной диагонали имеют вес больший, чем какая-либо ячейка по вертикали или горизонтали относительно нее. В результате, при использовании совокупности полученных данных, определение угла направления на источник звука произойдет без ошибок (в пределах 15°).

Приведены результаты эксперимента с длительностью сигнала 2 сек, количеством по 35 сигналов на запись (итого 70 секунд), обучение на 60 сек, тестирование на 10 секунд (см. рис. 3.5). Заметно снижение точности корректного определения направлений. Однако, за счет того, что количество тестируемых значений теперь меньше (запись для теста в общей сложности на 100 секунд короче), удачные «попадания» теперь привносят больший коэффициент в каждую ячейку. Можно увидеть, что ошибки, ранее рассеянные по всему диапазону значений, теперь встречаются в меньшем количестве ячеек. А удачные результаты концентрируются у тех направлений на источник

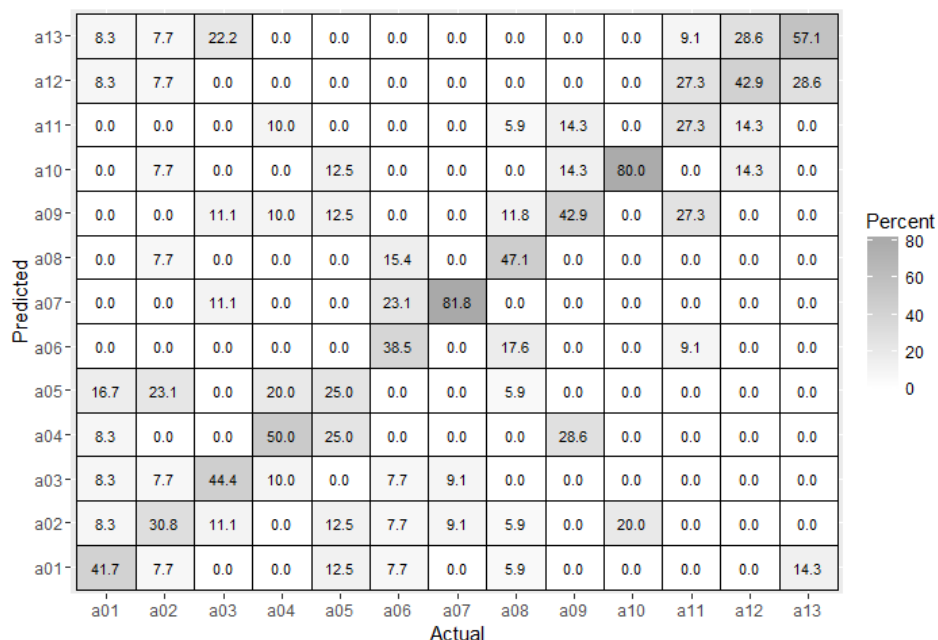


Рисунок 3.4. Эксперимент № 1.1. Матрица неточностей №1

звука, которые, очевидно, привносят наиболее выделяющиеся изменения в сигнал. Судя по ячейкам, имеющим самый большой вес, это направления, находящиеся под углами в 45° по отношению к осевой линии установки, а так же характерен сигнал, приходящий непосредственно с осевой линии установки, и в относительно меньшей степени выделяются сигналы с крайних направлений (0° и 180°). Наблюдается редкая «зеркальность» в значениях масс ячеек, относительно главной диагонали, это вызывается симметричностью макета установки относительно центральной оси.

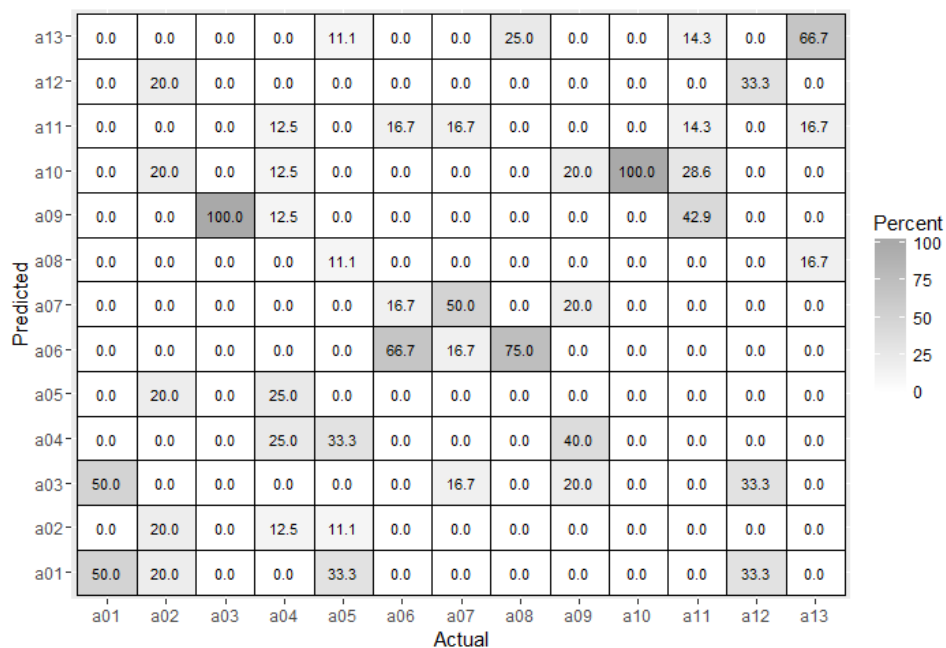


Рисунок 3.5. Эксперимент № 1.1. Матрица неточностей №2

3.3.2. Эксперимент № 1.2.

Представлен график зависимости точности результата от длительности сигнала (см.рис. 3.6).

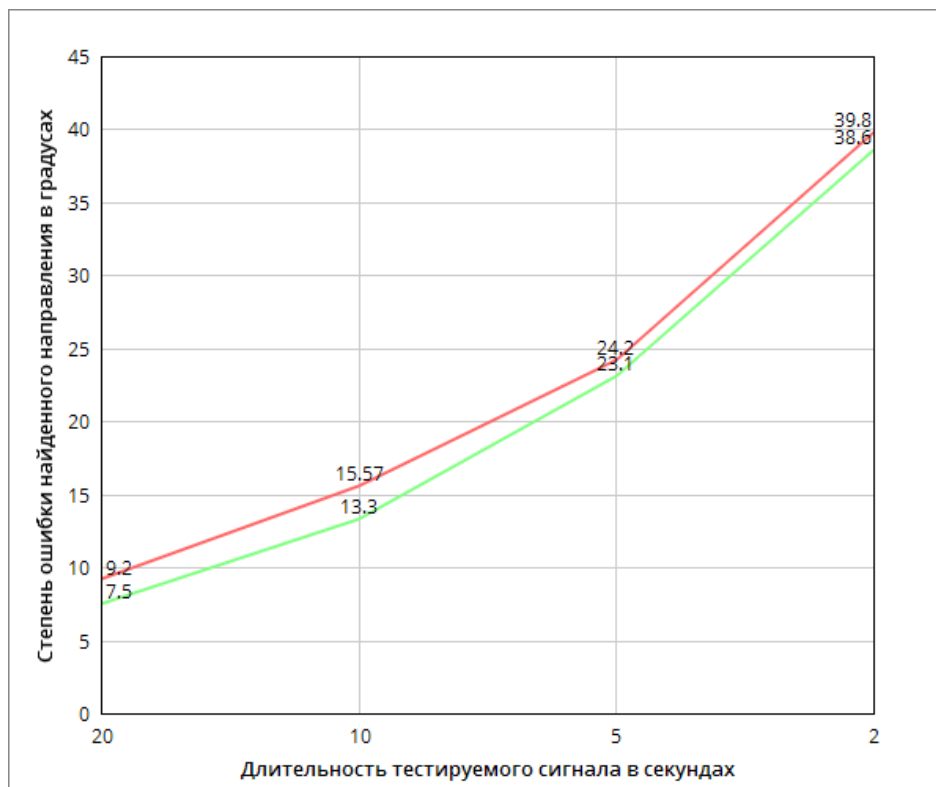


Рисунок 3.6. Эксперимент № 1.2. Зависимость точности нахождения направления на источник звука от длительности сигнала тестирования

На графике отображены величины ошибок нахождения направления. Верхняя линия соответствует макетной установке с звукоотражателем в форме стенки, нижняя линия соответствует установке с звукоотражателем, имитирующим человеческое ухо.

Наилучший результат достигается при длительности записи в 20 секунд. В этом случае ошибки нахождения направлений недостаточны, чтобы показать отклонение свыше 15° .

При длительности записи в 10 секунд, сохраняется общее расположение найденных направлений, однако встречаются ошибки определения, в размере 15° , 30° , а иногда и в 45° . Хотя такие ошибки редки и мало влияют на среднее значение ошибки по эксперименту.

Тестовая запись, продолжительностью 5 секунд, имеет достаточно большое количество различной величины ошибок, хотя на некоторые углы направление может определяться довольно точно (относительно остальных углов).

Записи продолжительностью 2 секунды обладают очень большим количеством ошибок. Хотя редкие направления могут определиться точно, общее количество ложных направлений не даст их выделить.

Таким образом, минимальная длительность сигнала для корректной работы должна быть не меньше 5 секунд. Чтобы совсем минимизировать количество ошибок, рекомендуется использовать тестовые записи продолжительностью от 20 секунд и более.

Инв. № подл	Подп. и дата	Инв. № дубл.	Взам. инв. №	Подп. и дата	<p>На графике отображены величины ошибок нахождения направления. Верхняя линия соответствует макетной установке с звукоотражателем в форме стенки, нижняя линия соответствует установке с звукоотражателем, имитирующим человеческое ухо.</p> <p>Наилучший результат достигается при длительности записи в 20 секунд. В этом случае ошибки нахождения направлений недостаточны, чтобы показать отклонение свыше 15°.</p> <p>При длительности записи в 10 секунд, сохраняется общее расположение найденных направлений, однако встречаются ошибки определения, в размере 15°, 30°, а иногда и в 45°. Хотя такие ошибки редки и мало влияют на среднее значение ошибки по эксперименту.</p> <p>Тестовая запись, продолжительностью 5 секунд, имеет достаточно большое количество различной величины ошибок, хотя на некоторые углы направление может определяться довольно точно (относительно остальных углов).</p> <p>Записи продолжительностью 2 секунды обладают очень большим количеством ошибок. Хотя редкие направления могут определиться точно, общее количество ложных направлений не даст их выделить.</p> <p>Таким образом, минимальная длительность сигнала для корректной работы должна быть не меньше 5 секунд. Чтобы совсем минимизировать количество ошибок, рекомендуется использовать тестовые записи продолжительностью от 20 секунд и более.</p>
Ли	Изм.	№ докум.	Подп.	Дат	<p>ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)</p> <p>Лист 57</p>

3.3.3. Эксперимент № 1.3.

Эксперимент со звукоулавливателем в форме уха.

Результаты работы макета установки со звукоулавливателем в форме уха и полученный при этом прирост в точности включены в сравнительный график, который можно наблюдать на рис. 3.6.

3.3.4. Эксперимент № 1.4.

Исследована точность работы системы при использовании списка полных групп в качестве системы признаков. Для операторов полных групп допустимы теоретико-множественные операции объединения (сложения) и пересечения (умножения). Проведен эксперимент, в котором сравнивается эффективность работы системы при использовании списка полных групп на сложении с эффективностью при использовании полных групп на умножении. Результаты эксперимента с использованием макета со звукоотражателем в форме стенки приведены на рис. 3.7. Результаты эксперимента с использованием макета со звукоотражателем в форме человеческого уха приведены на рис. 3.8.

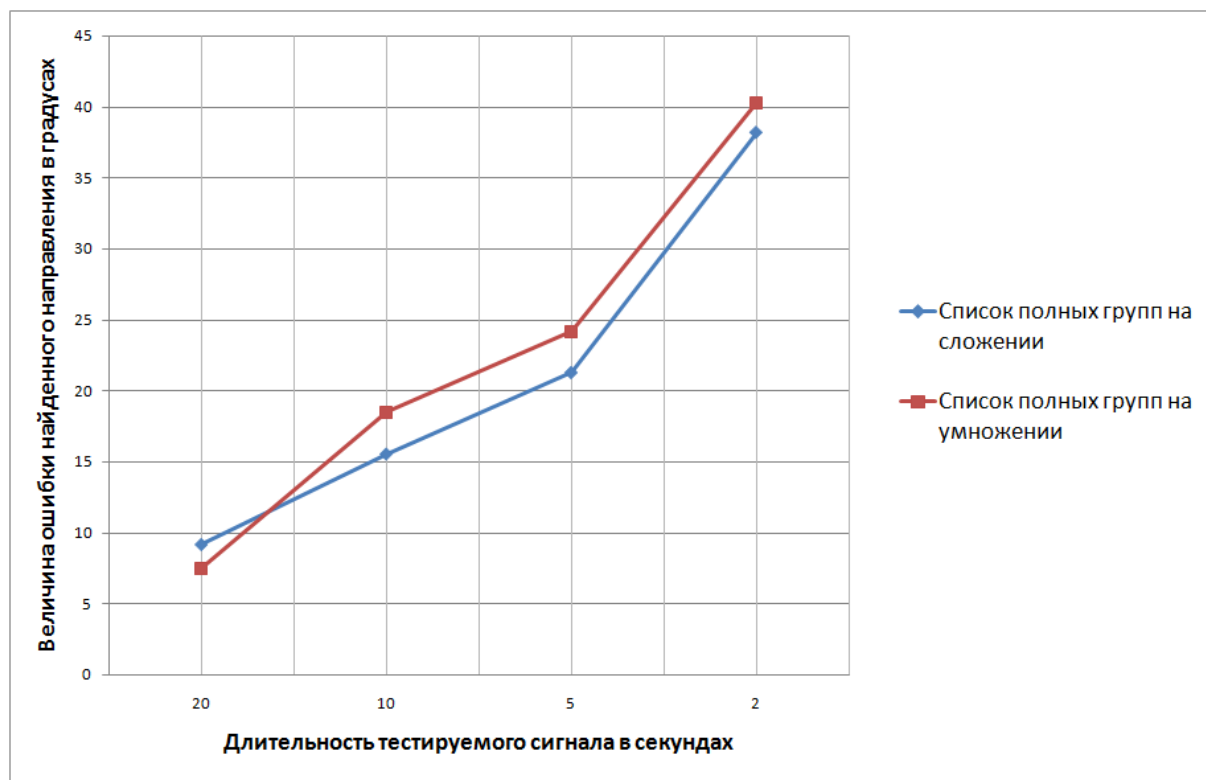


Рисунок 3.7. Эксперимент № 1.4. Зависимость точности нахождения направления на источник звука от использованной системы признаков (список полных групп на сложении).

Таким образом, можно наблюдать, что существует разница в величине ошибки найденного направления при использовании той или иной системы признаков. Судя по полученным данным, при тестировании системы с помощью сигналов продолжительностью 10 секунд и менее, преимущество за использованием списка полных групп на сложении. Однако, с увеличением продолжительности тестируемого сигнала происходит смещение точности работы в пользу использования списка полных групп на умножении.

Ив. № подл.	Подп. и дата
Ив. № дубл.	Взам. инв. №
Подп. и дата	Ив. № инв.
Ив. № подл.	Подп. и дата

Ли	Изм.	№ докум.	Подп.	Дат
----	------	----------	-------	-----

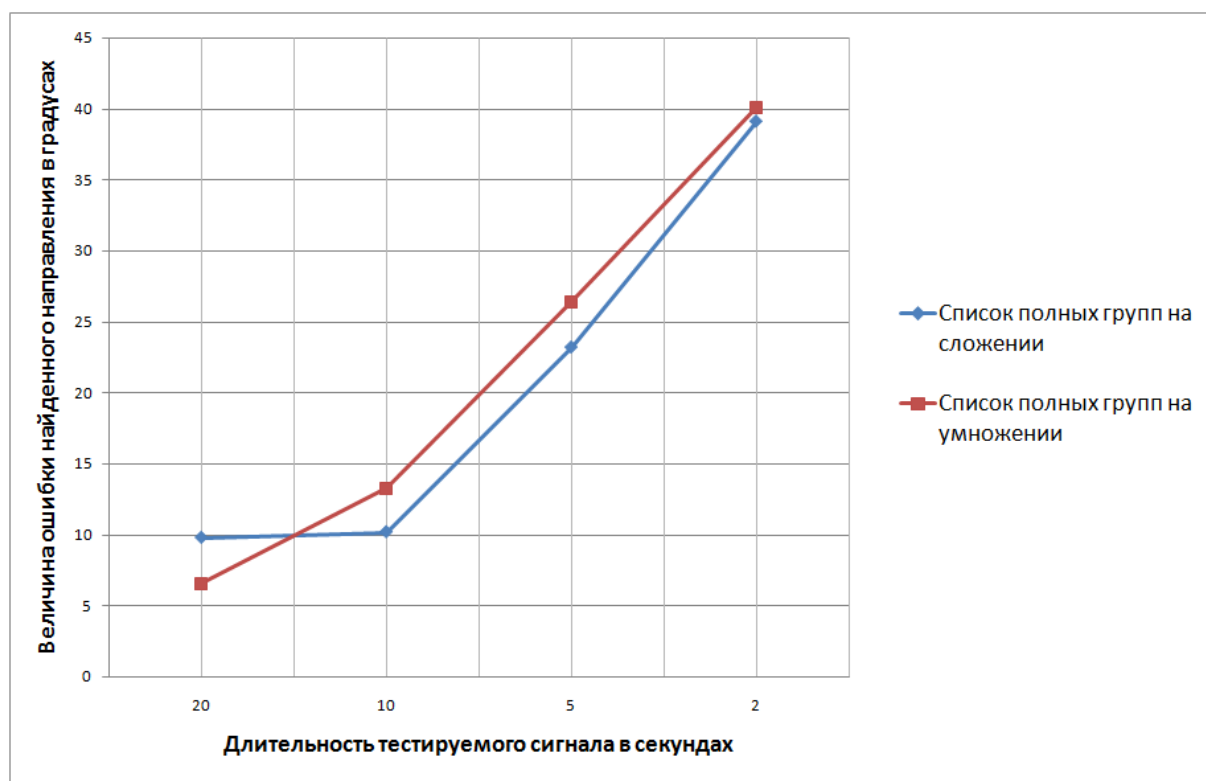


Рисунок 3.8. Эксперимент № 1.4. Зависимость точности нахождения направления на источник звука от использованной системы признаков (список полных групп на умножении).

3.3.5. Известные исследования. Сравнение результатов.

Для оценки эффективности системы путем сравнения, представлены результаты работы системы, описанной в работе Learning Sound Location from a Single Microphone [2]. Приведена таблица, в которой указаны величины ошибки найденного направления в градусах, при проведении различных экспериментов (см. рис. 3.9). Эксперименты проводились с различными типами звуков, продолжительность каждой записи составляла 5 минут. Также в составе звукозаписывающей установки были использованы четыре типа звукоотражающих стенок (А, В, С, D) с разной степенью сложности формы.

Из таблицы можно увидеть, что с использованием звукоотражающей стенки с самой сложной формой (С), удалось добиться лучших результатов. Исследования проведены на различных типах звуков, натуральных и искусственных, что позволяет как оценить эффективность работы системы, так и пронаблюдать зависимость успешности использования моноаурального метода при работе со звуками с различными свойствами.

Для сравнения результатов представлена сводная таблица эффективности работы разрабатываемой системы (см. рис. 3.10).

Как можно наблюдать, лучшие результаты, представленные в таблице сравнимы с рассмотренными результатами известного исследования [2]. Однако, величина тестируемого отрезка звука в разрабатываемой системе равна 20 секундам. Таким образом можно сказать, что эффективность разрабатываемой системы выше, за счет сокращения продолжительности тестируемого сигнала.

Подп. и дата	
Взам. инв. №	
Инв. № дубл.	
Подп. и дата	
Инв. № подл.	

Ли	Изм.	№ докум.	Подп.	Дат

Формат записывающей установки	A	B	C	D
Широкополосный шум	42,6°	8,8°	4,3°	22,3°
Человеческая речь	67,8°	19,3°	7,7°	21,35°
Собачий лай	55,7°	14,2°	18,3°	60,28°
Звук водопада	42,1°	11,8°	9,3°	42,7°
Чистый тон	88,7°	89,1°	86,4°	82,6°

Рисунок 3.9. Известные исследования. Величины ошибок найденного направления, при проведении различных экспериментов

№ Установки	Система признаков	Длительность сигнала в секундах	Средняя ошибка определения угла в градусах
1	Полные группы (сложение)	20	9,2
1	Полные группы (умножение)	20	7,5
2	Полные группы (сложение)	20	9,8
2	Полные группы (умножение)	20	6,6
1	Полные группы (сложение)	10	15,57
1	Полные группы (умножение)	10	18,46
2	Полные группы (сложение)	10	10,2
2	Полные группы (умножение)	10	13,3
1	Полные группы (сложение)	5	21,3
1	Полные группы (умножение)	5	24,2
2	Полные группы (сложение)	5	23,2
2	Полные группы (умножение)	5	26,4
1	Полные группы (сложение)	2	38,2
1	Полные группы (умножение)	2	40,3
2	Полные группы (сложение)	2	39,1
2	Полные группы (умножение)	2	40,1

Рисунок 3.10. Сводная таблица результатов экспериментов.

Инв. № подл.	Подп. и дата	Инв. № дубл.	Взам. инв. №	Подп. и дата	Инв. № инв.	Подп. и дата	Лист
Ли	Изм.	№ докум.	Подп.	Дат	ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)		

3.4. Выводы по главе 3

1. На основе предложенных в главе 2 модели и описания, а так же исследования методов классификации массивов данных, был разработан алгоритм монаурального поиска направления на источник звука, а так же была совершена его программная реализация.
2. При анализе систем-аналогов был разработан план экспериментов для проверки адекватности работы системы, а так же для сравнения результатов с результатами известных систем.
3. Проведен ряд экспериментов, в процессе которых также проводилась настройка программной части системы для наилучшей продуктивности работы.
4. Результаты работы алгоритма поиска направления на источник звука подтвердили адекватность использования систем распознавания образов в задачах обработки звука. Так же подтвердилась возможность использования моделей обработки изображения для решения задач обработки звуковых данных.
5. Результаты исследований разработанных алгоритмов поиска направления на источник звука показывают, что разработанные алгоритмы позволяют повысить точность поиска, по сравнению с другими алгоритмами.

Инв. № подл.	Подп. и дата	Инв. № дубл.	Взам. инв. №	Подп. и дата										
Инв. № подл.	Подп. и дата	Инв. № дубл.	Взам. инв. №	Подп. и дата										
<table border="1"> <tr> <td>Ли</td> <td>Изм.</td> <td>№ докум.</td> <td>Подп.</td> <td>Дат</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table>					Ли	Изм.	№ докум.	Подп.	Дат					
Ли	Изм.	№ докум.	Подп.	Дат										
<div>ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)</div> <div>Лист 61</div>														

Заключение

Основные результаты работы заключаются в следующем.

1. В данной работе был представлен и обоснован пример разработки моделей монауральной локализации источника звука. Предварительно проведен анализ исследований и результатов работы существующих систем, выполняющих схожие задачи.
2. В диссертации исследованы следующие задачи
 - (а) Разработана система локализации источника звука, было выполнено построение ее модели и решение с помощью необходимых методов.
 - (б) Разработана программная реализация модели системы
 - (с) Выполнено тестирование, включающее в себя проведение экспериментов с реализованной моделью, поиск алгоритма ее работы, дающего наилучший результат, а так же сравнение конечных результатов с аналогами.
3. В работе адаптирована теория активного восприятия, применительно к обработке и анализу звуковых сигналов
4. Все представленные в работе методы реализованы в виде программного обеспечения, способного работать на большом количестве распространенных персональных вычислительных машин.
5. По результатам исследований опубликовано несколько статей.

В результате выполнения работы были получены знания в областях исследований, имеющих непосредственное отношение к теме магистерской работы: машинном обучении, анализу сигналов, тематическом моделировании, улучшены практические навыки разработки программного обеспечения. Получен опыт обоснованного выбора и использования программных продуктов для решения поставленных задач, самостоятельного изучения новых методов исследований и проведения научных исследований в соответствии с тематикой задания. На практике выполнена разработка программной системы, решающей поставленную задачу, а так же проведено ее тестирование с анализом полученных результатов и их сравнением с результатами известных исследований в данной области. Обозначены перспективы дальнейшего модифицирования и развития и системы.

ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)

Лист

62

Инва. № подл.	Подп. и дата
Инва. № дубл.	Взам. инв. №
Подп. и дата	Инва. № дубл.
Инва. № подл.	Подп. и дата

Ли	Изм.	№ докум.	Подп.	Дат

Литература

1. *Butler R. A. Green D. M.* Sound localization by human listeners // *Perception and Psychophysics*. — 1991. — no. 51. — Pp. 182–186.
2. *Ashutosh Saxena Andrew Y. Ng.* Learning Sound Location from a single microphone // *Proceedings of the 2009 IEEE international conference on Robotics and Automation*. — 2009.
3. *Ryoichi Takashima Tetsuya Takiguchi, Arika Yasuo.* Single-Channel Sound Source Localization Based on Discrimination of Acoustic Transfer Functions // *Advances in Sound Localization*. — 2011. — no. 51. — Pp. 182–186.
4. *Утробин В. А.* Информационные модели системы зрительного восприятия для задач компьютерной обработки изображений. — Нижний Новгород: НГТУ им. Р. Е. Алексеева, 2001. — С. 234.
5. *А. Альтман Я.* Пространственный слух. — СПб.: Ин-т физиологии им. И.П. Павлова: РАН, 2010. — С. 311.
6. *А. Гельфанд С.* Слух: введение в психологическую и физиологическую акустику. — Медицина, 1984. — С. 352.
7. *Воробьев Ярослав.* В окружении виртуального звука.
8. Звук и слух. В окружении пространственного звука.
9. Monaural Hearing and Sound Localization.
10. *Rogers M. E. Butler R. A.* The linkage between stimulus frequency and covert peak areas as it relates to monaural localization // *Perception and Psychophysics*.
11. *Wightman F. L. Kistler D. J.* Monaural sound localization revisited // *Journal of the Acoustical Society of America*. — 1997. — no. 101. — Pp. 1050–1063.
12. *Carlile S.* The auditory periphery of the ferret II: The spectral transformations of the external ear and their implications for sound localization // *Journal of the Acoustical Society of America*. — 1990. — no. 88. — Pp. 2196–2204.
13. *Middlebrooks J. C. Green D. M.* Sound localization by human listeners // *Annual Review of Psychology*. — 1991. — no. 42. — Pp. 135–159.
14. *Fuzessery Z. M. Wenstrup J. J. Pollak G. D.* Determinants of horizontal sound location selectivity of binaurally excited neurons in an isofrequency region of the mustache bat inferior colliculus // *Journal of Neurophysiology*. — 1990. — no. 63. — Pp. 1128–1147.

15. Goldstein B. E. Sensation and Perception. — Pacific Grove, CA: Brooks/Cole.
16. Wotton J. M. Haresign T. Simmons J. A. Spatially dependent acoustic cues generated by the external ear of the big brown bat, *Eptesicus fuscus* // *Journal of the Acoustical Society of America*. — 1995. — no. 98. — Pp. 1423–1445.
17. J. Sodnik S. Tomazic R. Grasset A. Duenser M. Billinghamurst. Spatial sound localization in augmented reality environment // In Proc Computer-Human Interaction special interest group (CHISIG). — 1992.
18. Knudsen E. Instructed learning in the auditory localization pathway of the barn owl // *Nature*. — 2002. — no. 417.
19. Moiseff A., Konishi. M. The owl's interaural pathway is not involved in sound localization // *Journal Comp Physiology A: Neuroethology, Sensory, Neural and Behavioral Physiology*. — 1981. — no. 144(3).
20. Takiguchi T. Nakamura S. Shikano K. HMM-separation-based speech recognition for a distant moving speaker // *Transactions on Speech and Audio Processing / IEEE*. — Vol. 9. — 2001. — Pp. 127–140.
21. Takiguchi T. Nishimura M. Acoustic Model Adaptation Using First Order Prediction for Reverberant Speech // *Proceedings of 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP04) / IEEE*. — Montreal, Quebec, Canada: 2004.
22. Oppenheim A. V., Schaffer R. W. Discrete-Time Signal Processing // *Pearson Education*. — 1999.
23. McCarthy B. Sound Systems: Design and Optimization // *Focal Press*. — 2006.
24. Picone J. Signal modeling techniques in speech recognition // *Proc of the IEEE*. — Vol. 81(8).
25. Saxena A., Singh A. A microprocessor based speech recognizer for isolated hindi digits // *In IEEE ACE*.
26. Rabiner L. R. A tutorial on hidden markov models and selected applications in speech recognition // *Proc IEEE*. — Vol. 77(2).
27. Fuchs A. K. Feldbauer C. Stark M. Monaural Sound Localization // *INTERSPEECH*. — 2011. — Pp. 2521–2524.
28. Гай В. Е. Алгоритм определения максимумов сигнала // *Вестник Саратовского государственного технического университета*. — 2012. — Т. 65, № 2. — С. 49–53.
29. Гай В. Е. Алгоритмы формирования спектрального представления звукового сигнала на основе У-преобразования // *Бизнес информатика*. — 2013. — № 1 (23). — С. 44–49.
30. Утробин В. А. Элементы теории активного восприятия изображений // *Труды Нижегородского государственного технического университета им. Р.Е. Алексеева*. — 2010. — Т. 81, № 2. — С. 61–69.

Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата	Инв. № подл.						ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)	Лист	
	Ли	Изм.	№ докум.	Подп.	Дат						64		

31. Утробин В. А. Компьютерная обработка изображений. Принятие решений в пространстве эталонов. — Нижний Новгород: НГТУ им. Р. Е. Алексеева, 2004. — С. 221.
32. Утробин В. А. Физические интерпретации элементов алгебры изображения // *Успехи физических наук*. — 2004. — Т. 174, № 10. — С. 1089–1104.
33. Утробин В. А. Компьютерная обработка изображений. Анализ и синтез. — Н. Новгород: НГТУ, 2003. — С. 228.
34. Утробин В. А. Компьютерная обработка изображений. Информационные модели этапа понимания. — Н. Новгород: НГТУ, 2006. — С. 247.
35. Гай В. Е. Метод обработки биометрической информации в системах контроля и управления доступом // *Труды Нижегородского государственного технического университет им. Р.Е. Алексеева*. — 2013. — Т. 94, № 8. — С. 65–32.
36. Гай В. Е. Идентификация диктора по спектральным характеристикам речевого сигнала // *Датчики и системы*. — 2013. — № 7. — С. 40–45.
37. Dalal Navneet, Triggs Bill. Histograms of oriented gradients for human detection // *Computer Vision and Pattern Recognition*, 2005. CVPR 2005. IEEE Computer Society Conference on / IEEE. — Vol. 1. — 2005. — Pp. 886–893.
38. Lazebnik Svetlana, Schmid Cordelia, Ponce Jean. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories // *Computer Vision and Pattern Recognition*, 2006 IEEE Computer Society Conference on / IEEE. — Vol. 2. — 2006. — Pp. 2169–2178.
39. Fast human detection using a cascade of histograms of oriented gradients / Qiang Zhu, M-C Yeh, Kwang-Ting Cheng, Shai Avidan // *Computer Vision and Pattern Recognition*, 2006 IEEE Computer Society Conference on / IEEE. — Vol. 2. — 2006. — Pp. 1491–1498.
40. Bradski Gary, Kaehler Adrian. Learning OpenCV: Computer vision with the OpenCV library. — "O'Reilly Media, Inc. 2008.
41. Empirical evaluation of dissimilarity measures for color and texture / Jan Puzicha, Joachim M Buhmann, Yossi Rubner, Carlo Tomasi // *Computer Vision*, 1999. The Proceedings of the Seventh IEEE International Conference on / IEEE. — Vol. 2. — 1999. — Pp. 1165–1172.
42. Карасиков М. Е., Максимов Ю. В. Поиск эффективных методов снижения размерности при решении задач многоклассовой классификации путем её сведения к решению бинарных задач // *Машинное обучение и анализ данных*. — 2014. — Т. 1, № 9. — С. 1273 – 1290.

Инв. № подл.	Подп. и дата							
	Взам. инв. №							
	Инв. № дубл.							
	Подп. и дата							
Инв. № подл.	Подп. и дата							
		Ли	Изм.	№ докум.	Подп.	Дат	ВКР-НГТУ-09.04.01-(М14 ИВТ-3)-006-2016 (ПЗ)	Лист
								65