

Санкт-Петербургский государственный университет

Прикладная математика и информатика  
Кафедра теоретической кибернетики

Жеребцов Александр Юрьевич

# Распознавание звуковых сигналов в микрофонном массиве

Бакалаврская работа

Научный руководитель:  
д. ф.-м. н., профессор Барабанов А. Е.

Рецензент:  
к. ф.-м. н., доцент. Бондарко В. А.

Санкт-Петербург  
2016

SAINT-PETERSBURG STATE UNIVERSITY

Applied Mathematics and Informatics  
Chair of theoretical cybernetics

Alexander Zherebtsov

# Recognition of sound signals in microphone array

Bachelor's Thesis

Scientific supervisor:  
professor Andrey Barabanov

Reviewer:  
docent Vladimir Bondarko

Saint-Petersburg  
2016

# Оглавление

Введение	4
1. Постановка задачи	5
2. Звуковая волна	6
3. Определение расположения по зашумленному сигналу	7
4. Решение системы	9
5. Матрица ковариации ошибки	10
6. Свойства микрофонов	12
7. Алгоритм	16
Заключение	19
Список литературы	20

# Введение

С развитием робототехники всё чаще встают задачи локализации и распознавания объектов по данным, полученным при помощи различного оборудования и датчиков. В данной работе рассматривается задача локализации источника звука, в частности, определение расположения говорящего человека. Подобная задача решается в человеческом мозгу при определении направления на источник звука. Используя больше чем два устройства для приема и обработки звука, имеется возможность получить больше информации о расположении. В данном случае, в качестве оборудования для проведения экспериментов используется массив из четырех микрофонов, а также внешняя звуковая карта с возможностью записи одновременно необходимого количества аудио потоков. При этом сам звуковой сигнал проходит стадию дискретизации и оцифровки, и в конечном счете представляется в виде массива числовых данных с плавающей запятой. Полученный массив обрабатывается при помощи специально написанной программы для локализации источника. Ключевой особенностью этой работы является то, что при решении задачи принимаются во внимание искажения, вносимые при записи, связанные с характеристиками используемых микрофонов.

# 1. Постановка задачи

Имеется 4 конечные последовательности чисел  $S^{(1)}, S^{(2)}, S^{(3)}, S^{(4)}$  длины  $n$ . Все числа в последовательностях не превышают по модулю единицы. Каждая из последовательностей представима в виде:

$$S^{(i)} = \text{Mic}(U^{(i)})$$

$$U^{(i)}(k) = V(k - \delta_i) + \rho_i(k),$$

где преобразование  $\text{Mic}(U)$  соответствует выходу некоторой дискретной линейной стационарной системы  $H$  при подаче на вход сигнала  $U$ . Кроме того известны 4 вектора  $p_1, p_2, p_3, p_4$ . Требуется найти неизвестные параметры  $\delta_i$  и найти вектор  $X$ , минимизирующий норму невязки для системы

$$\begin{cases} \|X - p_1\| - \|X - p_2\| = \delta_1 - \delta_2 \\ \|X - p_1\| - \|X - p_3\| = \delta_1 - \delta_3 \\ \|X - p_1\| - \|X - p_4\| = \delta_1 - \delta_4 \\ \|X - p_2\| - \|X - p_3\| = \delta_2 - \delta_3 \\ \|X - p_2\| - \|X - p_4\| = \delta_2 - \delta_4 \\ \|X - p_3\| - \|X - p_4\| = \delta_3 - \delta_4 \end{cases}$$

## 2. Звуковая волна

Звуковую волну можно рассматривать как пример колебательного процесса. В воздухе скорость распространения звуковых колебаний варьируется от 320 до 350 метров в секунду в зависимости от температуры и давления. Одной из характеристик звуковой волны является звуковое давление, определяемое как избыточное давление, возникающее в среде при прохождении через неё звуковой волны. При отдалении от источника звука один и тот же сигнал будет воспринят микрофонами по-разному. В идеальном случае сигнал придет с задержкой, связанной с конечностью скорости распространения волны.

Таким образом, если значение звукового давления в точке на расстоянии  $d_1$  от источника, как функцию от времени, определить за  $S(t)$ , то значение давления в точке на расстоянии  $d_2$  можно записать в виде  $\alpha S(t + (d_1 - d_2)v)$ , где  $v$  - скорость звука в воздухе.

В реальности на вид звукового сигнала влияют волны, отраженные от окружающих предметов и стен, а также внешние шумы. Величина  $\alpha$  не является постоянной, а зависит от спектра сигнала и от свойств окружающей среды. Помимо того, при записи сигнала при помощи микрофона на результат будут влиять характеристики микрофона и звукозаписывающего оборудования.

### 3. Определение расположения по зашумленному сигналу

В данной работе рассматривается метод определения положения источника звука, основанный на расчете разницы расстояний от источника до каждого из микрофонов. Вначале рассмотрим ситуацию, в которой микрофоны не вносят никаких искажений в полученный сигнал.

Пусть  $U^{(1)}$  - сигнал полученный от первого микрофона, а  $U^{(2)}$  - от второго микрофона. Разобьем сигналы на составляющие:

$$U^{(i)}(k) = \alpha^{(i)}V(k - \delta_i) + \rho^{(i)}(k),$$

$V$  - Исходный сигнал

$\rho^{(i)}$  — внешние шумы, предполагаемые малыми по амплитуде и слабо коррелированными с основным сигналом.

В этом случае взаимнокорреляционная функция определяется как

$$\begin{aligned} R_{12}(\Delta_{12}) &= \sum_k U^{(1)}(k) \cdot U^{(2)}(k + \Delta_{12}) \\ R_{12}(\Delta_{12}) &= \alpha^{(1)}\alpha^{(2)} \sum_k V(k - \delta_1) \cdot V(k - \delta_2 + \Delta_{12}) + \\ &+ \sum_k \rho^{(1)}(k) (\alpha^{(2)}V(k - \delta_2 + \Delta_{12}) + \rho^{(2)}(k + \Delta_{12})) + \\ &+ \sum_k \rho^{(2)}(k + \Delta_{12})\alpha^{(1)}V(k - \delta_1) \end{aligned} \quad (1)$$

При данных предположениях взаимнокорреляционная функция от двух различных сигналов будет достигать своего максимального значения при сдвигах близких к сдвигам, которые соответствуют задержке в приеме оригинальных сигналов между микрофонами. А именно, при

$$\Delta_{ij} = \delta_j - \delta_i$$

Отклонения при определении сдвига будут зависеть от характера шума. Так, например, в случае с белым шумом отклонение сдвига можно рассматривать как случайную величину.

Зная сдвиг легко посчитать разницу в расстоянии соответствующую этому сдвигу:

$$\delta_{ij} = \frac{\Delta_{ij}}{F}v,$$

$\Delta_{ij}$  - сдвиг

$v$  - скорость звука в воздухе. (340 м/с)

$F$  - частота дискретизации звукового сигнала.

Распишем систему уравнений, связывающую разницу в расстоянии между микрофонами с положением источника:

$$\begin{cases} \|X - p_1\| - \|X - p_2\| = \delta_{12} \\ \|X - p_1\| - \|X - p_3\| = \delta_{13} \\ \|X - p_1\| - \|X - p_4\| = \delta_{14} \\ \|X - p_2\| - \|X - p_3\| = \delta_{23} \\ \|X - p_2\| - \|X - p_4\| = \delta_{24} \\ \|X - p_3\| - \|X - p_4\| = \delta_{34} \end{cases} \quad (2)$$

$p_i$  - положение  $i$ -го микрофона в пространстве.

$X$  - искомое положение источника.

Геометрическое место точек каждого из уравнений представляет собой двуполосный гиперboloид в трехмерном пространстве. Точка, удовлетворяющая системе уравнений, должна лежать на пересечении этих гиперboloидов. Таким образом, данная система переопределена, нелинейна и более того, линеаризованная система может быть плохо обусловленной. В связи с этим нет смысла искать точное решение. Вместо этого предлагается рассматривать задачу минимизации квадрата нормы вектора невязки.



## 4. Решение системы

Для решения задачи определения положения, требуется найти решение системы (2). Данная система переопределена и может не иметь точного решения. В связи с этим приходится решать задачу минимизации функционала от вектора невязки предыдущей системы, а именно[1]:

$$F = \sum_{i=1}^3 \sum_{j=i+1}^4 (\|X - p^{(i)}\| - \|X - p^{(j)}\| - \delta_{ij})^2$$

Градиент функционала  $F$  равен

$$F_x = 2 \sum_{i=1}^3 \sum_{j=i+1}^4 \left( \frac{X_x - p_x^{(i)}}{\|X - p^{(i)}\|} - \frac{X_x - p_x^{(j)}}{\|X - p^{(j)}\|} \right) \cdot (\|X - p^{(i)}\| - \|X - p^{(j)}\| - \delta_{ij})$$

$$F_y = 2 \sum_{i=1}^3 \sum_{j=i+1}^4 \left( \frac{X_y - p_y^{(i)}}{\|X - p^{(i)}\|} - \frac{X_y - p_y^{(j)}}{\|X - p^{(j)}\|} \right) \cdot (\|X - p^{(i)}\| - \|X - p^{(j)}\| - \delta_{ij})$$

$$F_z = 2 \sum_{i=1}^3 \sum_{j=i+1}^4 \left( \frac{X_z - p_z^{(i)}}{\|X - p^{(i)}\|} - \frac{X_z - p_z^{(j)}}{\|X - p^{(j)}\|} \right) \cdot (\|X - p^{(i)}\| - \|X - p^{(j)}\| - \delta_{ij})$$

Приравнивая к нулю градиент, получим необходимое условие для локального минимума функционала. У данной системы уравнений также нет аналитического решения, однако мы можем использовать градиент при решении задачи численными методами.

При решении задачи минимизации функционала предлагается использовать градиентные методы. Стоит учитывать, что данный функционал может иметь несколько локальных минимумов, а линеаризованная система может оказаться плохо обусловленной(в чем легко убедиться, рассмотрев систему в точке отдаленной от точек  $p^{(i)}$ ).

Для сведения к минимуму влияния данных факторов, при практическом решении данной задачи градиентным методом начальные данные были выбраны из сетки с целью минимизации начального значения функционала. А именно, в качестве начальных данных последовательно выбирались точки равномерно расположенные в узлах сетки. Размеры сетки связаны лишь с ограничением на расположение источника.

## 5. Матрица ковариации ошибки

Введем обозначения

$$G(X) = \begin{cases} \|X - p_1\| - \|X - p_2\| \\ \|X - p_1\| - \|X - p_3\| \\ \|X - p_1\| - \|X - p_4\| \end{cases}, \quad \delta = \begin{pmatrix} \delta_{12} \\ \delta_{13} \\ \delta_{14} \end{pmatrix} \quad (3)$$

Тогда подсистему системы (2) можно записать в виде

$$F(X) = \delta. \quad (4)$$

Пусть  $X_0$  - точное решение системы (4) при  $\delta = \delta_0$ . Пусть  $X_1$  - точное решение системы (4) при  $\delta = \delta_1$ . В таком случае выполнено равенство

$$G(X_1) - G(X_0) = \delta_1 - \delta_0.$$

Ввиду непрерывности  $G$ , в малой окрестности  $X_0$  будет выполнено

$$G'(X_0) \cdot (X_1 - X_0) = \delta_1 - \delta_0 + o(|X_1 - X_0|).$$

За исключением пограничных случаев  $F'(X_0)$  не вырождена и следовательно:

$$X_1 - X_0 \approx (G'(X_0))^{-1} \cdot (\delta_1 - \delta_0).$$

Таким образом, если рассматривать  $\delta_1 - \delta_0$  в качестве нормально распределенной случайной величины с нулевым математическим ожиданием и ковариационной матрицей  $Q$ , то отклонение  $X_1 - X_0$  будет также распределено нормально с нулевым математическим ожиданием и матрицей ковариации

$$G'(X_0)^{-1} Q (G'(X_0)^{-1})^T.$$

При тех же предположениях, отклонение в расстоянии по направлению на источник будет нормально распределено с дисперсией

$$\frac{X_0^T G'(X_0)^{-1} Q (G'(X_0)^{-1})^T X_0}{|X_0|^2}.$$

Отклонение угла в плоскости  $(x, y)$  в первом приближении будет распределено как

$$2\arctg\left(\frac{v}{2|X_0|}\right),$$

где

$$v \sim N(0, X_N^T G'(X_N))^{-1} Q(G'(X_0))^{-1} X_N / |X_N|^2, \quad X_N = (-X_{0y} X_{0x} 0)^T.$$

Разница в отклонении угла перпендикулярного плоскости  $(x, y)$  будет распределена как

$$2\arctg\left(\frac{u}{2|X_0|}\right),$$

где

$$u \sim N(0, X_{N2}^T G'(X_N))^{-1} Q(G'(X_0))^{-1} X_{N2} / |X_{N2}|^2, \quad X_{N2} = X_N \times X_0.$$

Таким образом, все интересующие нас характеристики будут иметь распределения зависящие как от положения источника в пространстве, так и от ошибки в определении разницы расстояний.

## 6. Свойства микрофонов

В ходе работы была проведена серия экспериментов. На вход динамика подавался сигнал вида

$$A(t) \sin(2\pi ft)$$

где  $A(t) = 1$  в середине промежутка и линейно приближается к нулю на границе. При этом сигнал на выходе микрофона ожидаемо представим в виде

$$B(t) \sin(2\pi ft + \phi)$$

Установив соответствие между ожидаемым и реальным сигналом, находятся коэффициенты  $B$ ,  $\phi$  [2]. Зная эти коэффициенты можно выразить значение передаточной функции системы динамик - воздушная среда - микрофон в точке  $f$  следующим образом:

$$H(f) = \frac{B}{A} e^{i\phi}$$

Эксперимент был проведен для частот  $f$  от 110 Гц до 3520 Гц, для каждого из четырех микрофонов и для различных положений динамика относительно микрофонов.

Таким образом, было построено приближение передаточной функции для каждого из микрофонов на указанных промежутках частот.[3]

Стоит отметить, что при использовании качественного динамика его влияние на передаточную функцию будет не значительным и его можно исключить из рассмотрения.

Зная передаточную функцию и полученный с микрофона сигнал мы можем, воспользовавшись прямым и обратным преобразованием Фурье, получить сигнал, который был бы ближе к исходному. Преобразованный таким образом сигнал будет содержать в себе меньше отклонений связанных со свойствами микрофонов и потому может быть использован для более точного определения местоположения.

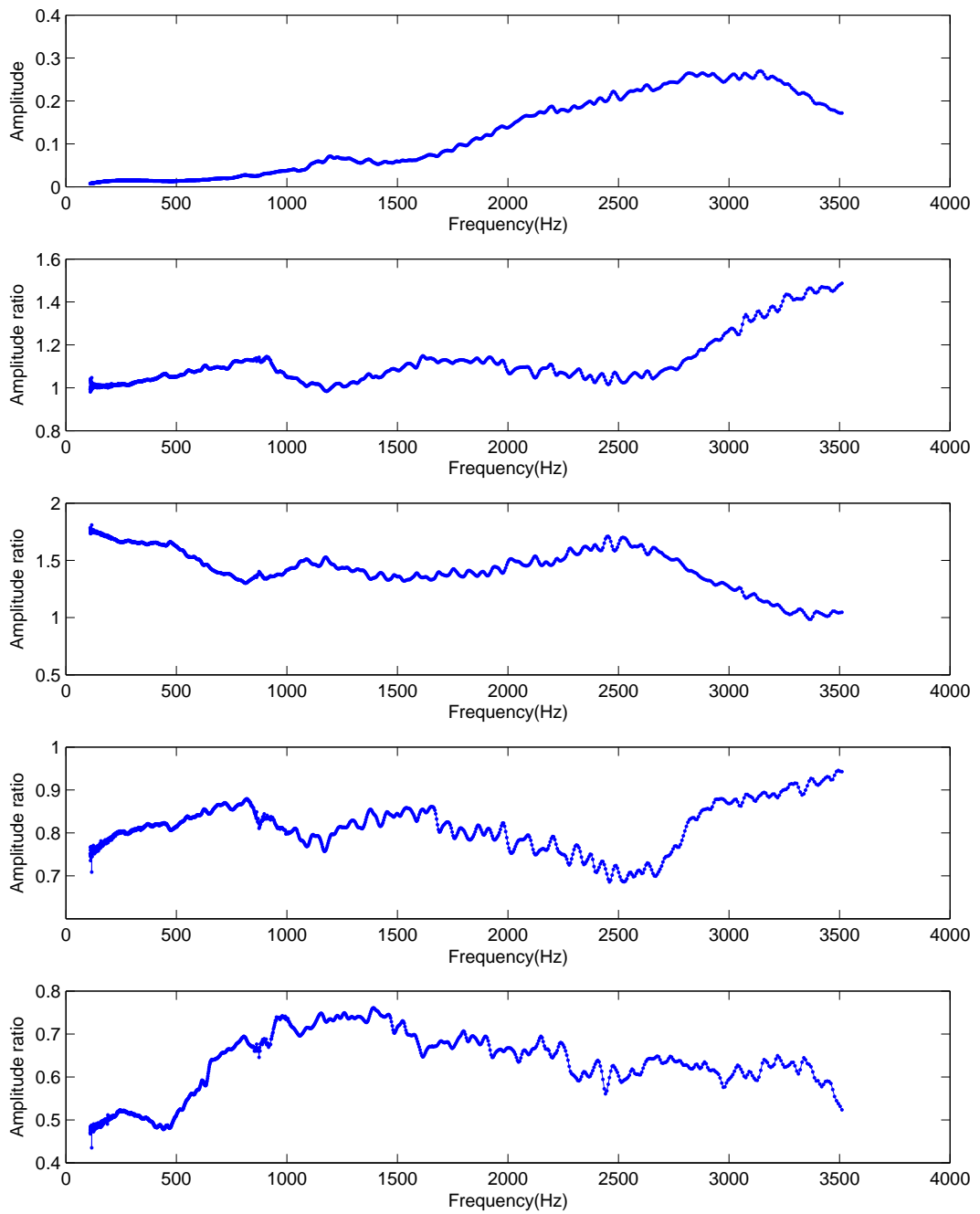


Рис. 1: Амплитудно-частотная характеристика.

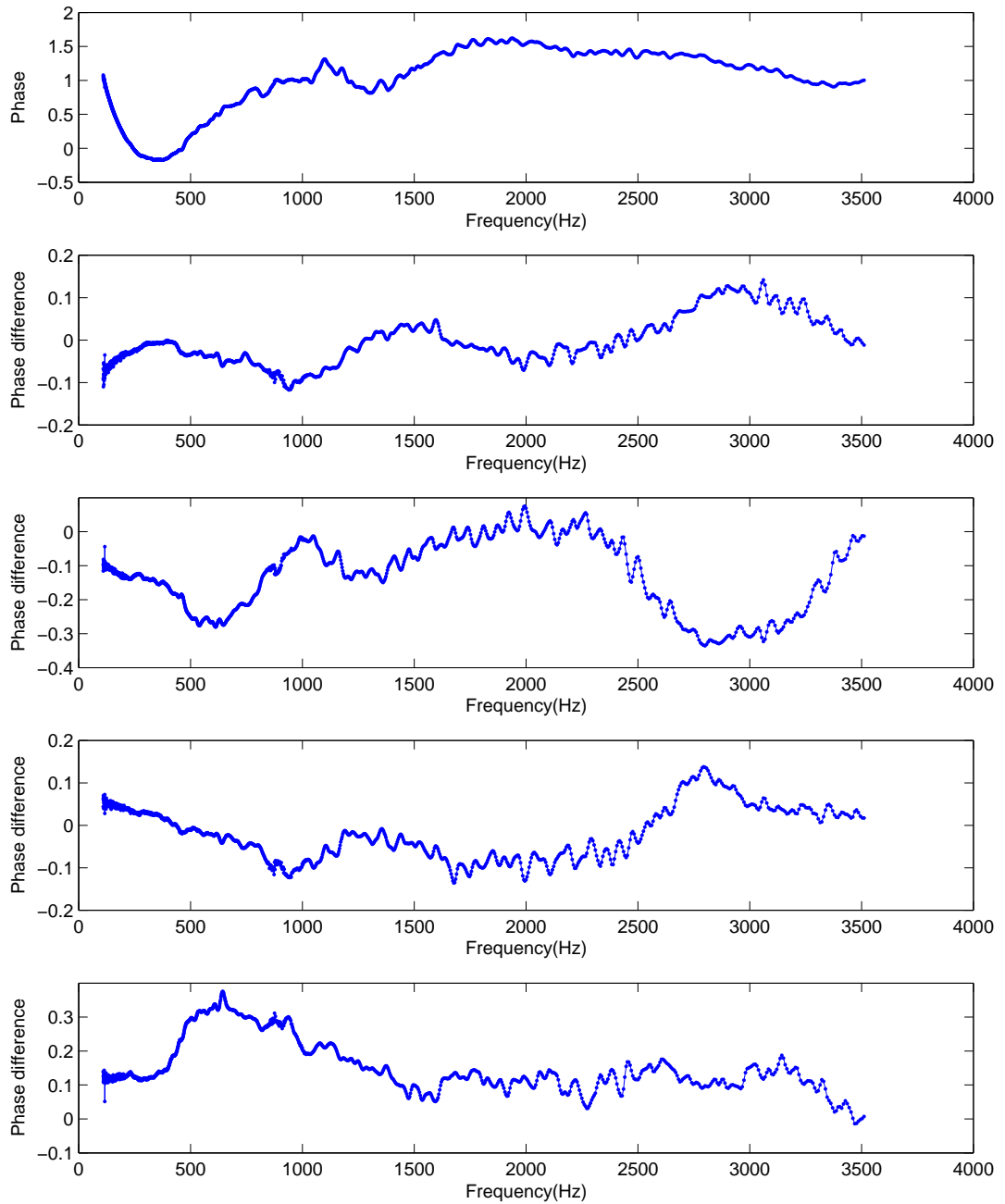


Рис. 2: Фазо-частотная характеристика.

На первых графиках на страницах изображены соответственно графики среднего, по четырем микрофонам, значения амплитудно-частотной и фазо-частотной характеристик. Ниже изображены графики отношения амплитуд для каждого из микрофонов к среднему значению для амплитудно-частотной характеристики и разницы фаз для фазо-частотной характеристики. Как видно из графиков амплитудно-частотной характеристики, сдвиг фазы для выходного сигнала может варьироваться до половины периода. Этот сдвиг может оказать существенное влияние на взаимнокорреляционную функцию сигналов. Это в свою очередь влияет на точность локализации источника звука.

Стоит отметить, что не смотря на то, что в экспериментах использовались одинаковые модели микрофонов, на графиках явно видны их различия. Так по графикам фазо-частотной характеристики можно заметить разницу до десятых долей периода, что пусть и не так сильно, но повлияет на результаты последующих вычислений. По графикам амплитуд также можно увидеть заметные различия между микрофонами, что оказывает влияние на то, какой вклад будут вносить каждый из участков спектра входного сигнала на конечный результат.

## 7. Алгоритм

- Воспроизвести специальный файл на качественном оборудовании, записав полученный результат на используемые микрофоны.
- Обработать полученные записи, получив на выходе амплитудно-фазовую частотную характеристику для записанных частот.

На этом действия необходимые для определения свойств микрофонов заканчиваются. Далее следует обработка сигнала с микрофонов и его локализация.

- Взять преобразование Фурье от входного сигнала.
- Поделить на передаточную функцию соответствующего микрофона.
- Взять обратное преобразование от преобразованного сигнала.
- Посчитать взаимнокорреляционную функцию для каждой пары микрофонов и найти сдвиг её максимизирующий.
- Посчитать разницу расстояний до источника для каждого микрофона, основываясь на сдвиге.
- Основываясь на разнице расстояний найти положение источника.



№	$x$	$y$	$z$	$\phi$	$\Delta$	$\widehat{\Delta}$	$\Delta_\phi$	$\widehat{\Delta}_\phi$
1	-100	0	20	-90	80	87	9	5
2	-85	-50	20	-60	87	91	11	15
3	-50	-85	20	-30	79	81	20	24
4	0	-100	20	0	98	99	48	21
5	-50	-85	20	30	88	87	1	5
6	85	-50	20	60	66	88	3	3
7	100	0	20	90	32	52	3	5
8	-100	0	20	-90	78	87	8	5
9	-85	-50	20	-60	88	89	10	13
10	-50	-85	20	-30	84	90	28	26
11	0	-100	20	0	51	83	3	2
12	-50	-85	20	30	77	86	6	1
13	85	-50	20	60	80	64	22	31
14	100	0	20	90	82	80	1	1
15	-100	0	20	-90	83	88	20	22
16	-85	-50	20	-60	88	92	7	23
17	-50	-85	20	-30	90	95	8	30
18	0	-100	20	0	100	100	21	46
19	-50	-85	20	30	86	88	1	10
20	85	-50	20	60	71	80	4	2
21	100	0	20	90	59	76	6	4
22	-100	0	20	-90	77	87	8	5
23	-85	-50	20	-60	86	86	1	1
24	-50	-85	20	-30	76	86	13	6
25	0	-100	20	0	77	83	1	1
26	-50	-85	20	30	90	91	1	7
27	85	-50	20	60	71	74	13	13
28	100	0	20	90	24	77	2	2
avarage					76.71	84.53	9.96	11.75

Таблица 1. Результаты экспериментов.

Здесь  $(x, y, z)$  - положение диктора.  $\Delta, \widehat{\Delta}$  - ошибки в определении положения для алгоритма учитывающего свойства микрофонов и для прямого алгоритма.  $\Delta_\phi, \widehat{\Delta}_\phi$  - ошибки в определении направления на источник звука (угловой координаты в плоскости  $(x, y)$ ) для указанных алгоритмов.

Все значения связанные с положением приведены в сантиметрах, а углы в градусах.

Как видно из таблицы, в большинстве случаев полученный алгоритм заметно точнее определяет расположение источника звука. Точность определения направления на источник в среднем также была увеличена.

Стоит отметить большое значение модуля ошибки при определении положений для обоих алгоритмов. Это можно объяснить явлением реверберации, когда отраженные от предметов и стен колебания вносят существенные изменения на полученный микрофоном сигнал. Также стоит учитывать погрешности при определении точки максимума взаимнокорреляционной функции, которые также оказывают заметное влияние на ошибку в определении положения при значительном отдалении источника от микрофонов.

## Заключение

В результате работы было проведено множество экспериментов с записью звуковых сигналов при помощи специально установленных микрофонов и дополнительного оборудования. Был реализован алгоритм для определения положения источника звука, а также разработаны способы улучшения результатов его работы. Полученный в итоге алгоритм может найти широкое применение как в системах определения источников звука, так и в системах требующих определения источников различных колебаний.

## Список литературы

- [1] Benesty, Sondhi, Huang. Springer Handbook of Speech Processing. Part 1 | 51. — Springer, 2008. — P. 1056.
- [2] А.В.Опенгейм, Р.В.Шафер. Цифровая обработка сигналов / Под ред. С.Я. Шац. — М. : Связь, 1979. — С. 67.
- [3] Рабинер Л., Гоулд Б. Теория и применение цифровой обработки сигналов / Под ред. С.Я. Шац. — М. : Мир, 1979. — С. 30.