

УДК 534.83

ВЫДЕЛЕНИЕ РЕЧЕВОГО СООБЩЕНИЯ ИЗ ПОМЕХ, ВНОСИМЫХ СТОРОННИМИ РАСПРЕДЕЛЁННЫМИ ИСТОЧНИКАМИ

*В. А. Канаков, Н. А. Миронов**

Нижегородский госуниверситет им. Н. И. Лобачевского, г. Нижний Новгород, Россия

Задачей исследования является выделение речевого сигнала, исходящего из определённой точки пространства, и расчёт разборчивости выделенного голосового сообщения. Она решается методом уменьшения влияния помех от источников речевых сообщений на выделяемый сигнал. Этот метод основан на введении в регистрирующие каналы временных задержек, зависящих от пространственных координат. Тестовыми объектами исследования являются аудиозаписи голосов восьми разных человек. Доказано, что при увеличении количества микрофонов разборчивость выделенного из помех речевого сообщения улучшается.

ВВЕДЕНИЕ

Выделение голоса конкретного человека из совокупности голосов является актуальной задачей при обработке сигналов. Если в некотором объёме пространства одновременно разговаривают несколько человек, то при записи сигнала с одного микрофона будет слышен только шум, возникающий при наложении одного голоса на другой. Поэтому в данной работе мы рассмотрим применение многопозиционной системы для выделения речевого сообщения из помех, вносимых сторонними распределёнными в пространстве источниками речи.

Целью исследования является разработка и экспериментальная проверка алгоритма уменьшения влияния помех от источников речевых сообщений на полезный сигнал за счёт введения в регистрирующие каналы временных задержек. Этот алгоритм позволяет одновременно выделять речевые сообщения, исходящие из любых точек пространства, при отсутствии априорной информации о координатах источников.

1. ПОСТАНОВКА ЗАДАЧИ

Сформулируем задачу исследования следующим образом. Пусть на акустической сцене имеется несколько источников речевых сигналов с одинаковой мощностью. Требуется выделить сигнал, исходящий из определённой точки пространства, с помощью системы микрофонов, расположенных по периметру сцены, и оценить понятность выделенного речевого сообщения. Для этого требуется выполнить расчёт разборчивости последнего. Тестовыми объектами в данной работе являются аудиозаписи голосов восьми разных человек.

Поставленная задача решается при определённых ограничениях.

Во-первых, каждый источник звука считается всенаправленным, т. е. звуковая энергия распространяется равномерно во всех направлениях и звуковое давление уменьшается обратно пропорционально расстоянию r от источника звука.

Во-вторых, рассматривается распространение звука в свободном пространстве. Целью исследования является изучение принципиальной возможности выделения голосового сообщения на фоне совокупности большого числа источников широкополосных помех, поэтому, для упрощения

* electro112@mail.ru

расчётов, мы пренебрегаем реверберацией звука в помещении и считаем, что звуковой сигнал достигает микрофонов кратчайшим путём по прямой, т. е. не отражаясь от поверхностей стен, пола и потолка помещения.

В-третьих, полагаем, что рост людей в помещении примерно одинаков, поэтому без потери общности рассуждений считаем, что все источники речевых сообщений расположены в горизонтальной плоскости на высоте примерно 170 см от пола помещения. Размещение микрофонов на той же высоте позволяет не учитывать в расчётах координату z .

2. ТЕОРИЯ

Измерительная система представляет собой N микрофонов, равномерно распределённых по периметру сцены. Каждый из микрофонов регистрирует сумму всех сигналов, один из которых определён как полезный сигнал $S(t)$, а остальные — как помехи $G_k(t)$, $k = 1, \dots, m$. При разных расстояниях r_i от источника до приёмного устройства с номером i , $i = 1, \dots, N$, сигналу требуется разное время, чтобы достичь датчика. Сигнал, регистрируемый каждым из микрофонов, можно представить следующим образом (c — скорость звука):

$$Q_i(t) = \frac{A}{r_i} S\left(t - \frac{r_i}{c}\right) + \sum_{k=1}^m \frac{1}{r_{ki}} G_k\left(t - \frac{r_{ki}}{c}\right). \quad (1)$$

Для выделения полезного сигнала на фоне помех от пространственно-распределённых источников в каждый регистрируемый сигнал вводят временные задержки, вычисляемые по формуле

$$t_{ji} = \tilde{r}_{ji}/c, \quad (2)$$

где $\tilde{r}_{ji}$ ($j = 1, \dots, p$) — расстояние от произвольно выбранной точки пространства (точки фокусировки системы) до i -го микрофона. Введение временных задержек необходимо для того, чтобы приходящие на каждый микрофон сигналы из точки фокусировки системы регистрировались синхронно, т. е. для всех $i = 1, \dots, N$, и $j = 1, \dots, p$ выполнялось равенство

$$\tau_{\text{opt}} = \frac{r_i}{c} - \frac{\tilde{r}_{ji}}{c} \quad (3)$$

Тогда система задержанных сигналов для каждого микрофона определяется формулой

$$Q_i(t) = \frac{A}{r_i} S(t - \tau_{\text{opt}}) + \sum_{k=1}^m \frac{1}{r_{ki}} G_k\left(t - \frac{r_{ki}}{c} + \frac{\tilde{r}_{ji}}{c}\right). \quad (4)$$

Суммированием всех сигналов можно увеличить мощность полезного сигнала в среднем в N^2 раз и тем самым значительно подавить сигналы от источников помех:

$$S_{\text{ex}} = \sum_{i=1}^N Q_i = S(t - \tau_{\text{opt}}) \sum_{i=1}^N \frac{A}{r_i} + H(t), \quad (5)$$

где S_{ex} — выходной сигнал, $H(t)$ — суммарная помеха

$$H(t) = \sum_{i=1}^N \sum_{k=1}^m \frac{1}{r_{ki}} G_k\left(t - \frac{r_{ki}}{c} + \frac{\tilde{r}_{ji}}{c}\right). \quad (6)$$

Для оценки возможности выделения речевого сигнала необходимо определить разрешающую способность предложенной акустической системы по пространственным координатам. Для этого в каждой точке сцены рассчитывается двумерная пространственная автокорреляционная функция системы сигналов от N микрофонов по M отсчётам:

$$R(x', y') = \sum_{l=1}^M S_{\text{ex}}^2(x', y', x_0, y_0), \quad (7)$$

где (x', y') — координаты точек фокусировки N микрофонов, (x_0, y_0) — фиксированные координаты точечного источника речевого сигнала, $l = 1, \dots, M$ — индексы M отсчётов суммарного выходного сигнала.

Координаты максимумов функции $R(x', y')$ будут соответствовать координатам источников речевых сигналов. Полученные оценки координат источников сигналов могут быть использованы для фокусировки системы микрофонов на любой обнаруженный источник для последующего выделения полезного сигнала на фоне прочих мешающих источников. Однако знание координат полезного источника не является необходимым в случае реализации параллельной схемы наблюдения за большим числом равномерно распределённых по сцене точек фокусировки.

Для реализации параллельной схемы наблюдения, интересующая область пространства с площадью D разбивается на равные площадки с площадью Δx^2 , где Δx — разрешающая способность системы. Зная координаты микрофонов, можно задать матрицу временных задержек, необходимых для синхронного приёма сигналов от конкретной области разбиения:

$$\mathbf{T} = \begin{pmatrix} \tau_{11} & \tau_{21} & \cdots & \tau_{p1} \\ \tau_{12} & \tau_{22} & \cdots & \tau_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{1N} & \tau_{2N} & \cdots & \tau_{pN} \end{pmatrix}, \quad (8)$$

где N — количество микрофонов, p — количество точек фокусировки, расположенных, вообще говоря, на произвольно заданной сетке узлов на расстоянии не менее Δx друг от друга.

Мощность сигнала из точки фокусировки системы увеличится, и сигнал можно будет распознать. В итоге мы будем иметь $p = D/\Delta x^2$ одновременно работающих выходных каналов, позволяющих осуществлять автоматический контроль за всем пространством наблюдения:

$$S_{\text{ex}j} = S(t - \tau_{\text{opt}}) \sum_{i=1}^N \frac{A}{r_i} + \sum_{i=1}^N \sum_{k=1}^m \frac{1}{r_{ki}} G_k(t - \frac{r_{ki}}{c} + \tau_{ji}), \quad (9)$$

где $j = 1, \dots, p$.

Заключение о точности восстановления речевого сообщения, исходящего из определённой точки пространства, можно сделать, рассчитав разборчивость выделенного речевого сообщения [1].

3. ЧИСЛЕННАЯ МОДЕЛЬ И ПОЛУЧЕННЫЕ С ЕЁ ПОМОЩЬЮ РЕЗУЛЬТАТЫ

Для расчёта разрешающей способности акустической системы в качестве модели сигнала использовался записанный речевой фрагмент с длительностью около 4 с. После фильтрации собственных шумов приёмных устройств в полосе от частот 0 до 100 Гц (рабочий диапазон 100÷8000 Гц) и использования алгоритма введения задержек для фокусировки системы из 4 микрофонов была рассчитана двумерная пространственная автокорреляционная функция (см. рис. 1).

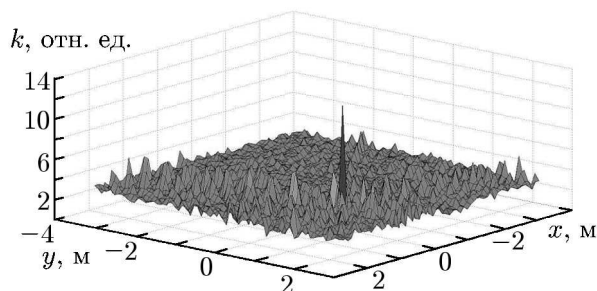


Рис. 1. Двумерная пространственная автокорреляционная функция k системы сигналов от 4-х микрофонов

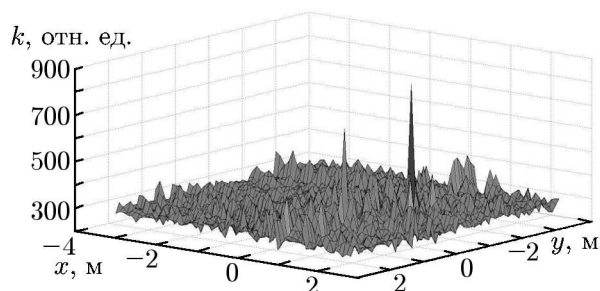


Рис. 2. Двумерная пространственная автокорреляционная функция k системы сигналов от 4-х микрофонов при наличии двух речевых источников

По полученному графику можно оценить разрешающую способность акустической системы как ширину главного максимума автокорреляционной функции в точке с координатами источника речевого сигнала по уровню $1/2$ от её максимума [2]. Разрешающая способность предложенной измерительной системы составила 8 см. Это означает, что система из четырёх микрофонов, установленных по периметру сцены, способна однозначно разделить два одновременно действующих источника речевого сигнала и определить их координаты при условии, что расстояние между ними больше 8 см. Данный вывод подтверждается расчётом двумерной пространственной автокорреляционной функции для двух одновременно действующих источников речевого сигнала (см. рис. 2).

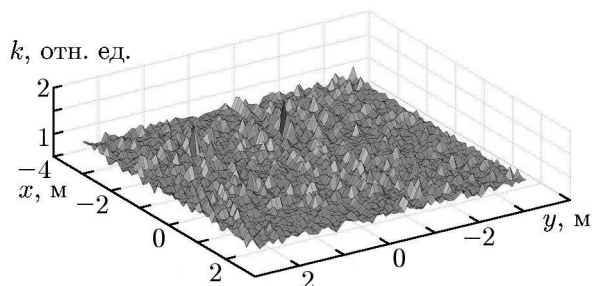


Рис. 3. Двумерная пространственная автокорреляционная функция k системы сигналов от 4-х микрофонов при наличии восьми речевых источников

Если число источников сигналов больше четырёх, то обнаружить и однозначно оценить их координаты на фоне боковых лепестков двумерной автокорреляционной функции затруднительно (см. рис. 3).

Для оценки координат большого числа источников широкополосных шумовых сигналов можно использовать трёхмерную автокорреляционную функцию совокупности сигналов от четырёх микрофонов [3].

Однако, как отмечалось выше, знание координат полезного сигнала становится необязательным, если реализовать параллельную схему наблюдения.

В данной работе компьютерный эксперимент был проведён для восьми одновременно действующих источников широкополосного сигнала. По периметру сцены равномерно были распределены $N=4$ микрофона, каждый из которых регистрировал сумму всех акустических сигналов. Для примера была поставлена задача о выделении сигнала от источника № 5; реализация сигнала представлена на рис. 4а.

В начале был записан суммарный (на сцене располагается 8 источников речевых сигналов) выходной сигнал системы от 4 микрофонов без введения задержек в регистрируемые каналы (см. рис. 4б). В данном случае сигналы приходят на микрофоны в разное время, накладываются друг на друга и возникает шум.

Далее в 4-х каналах регистрации были введены временные задержки, соответствующие про-

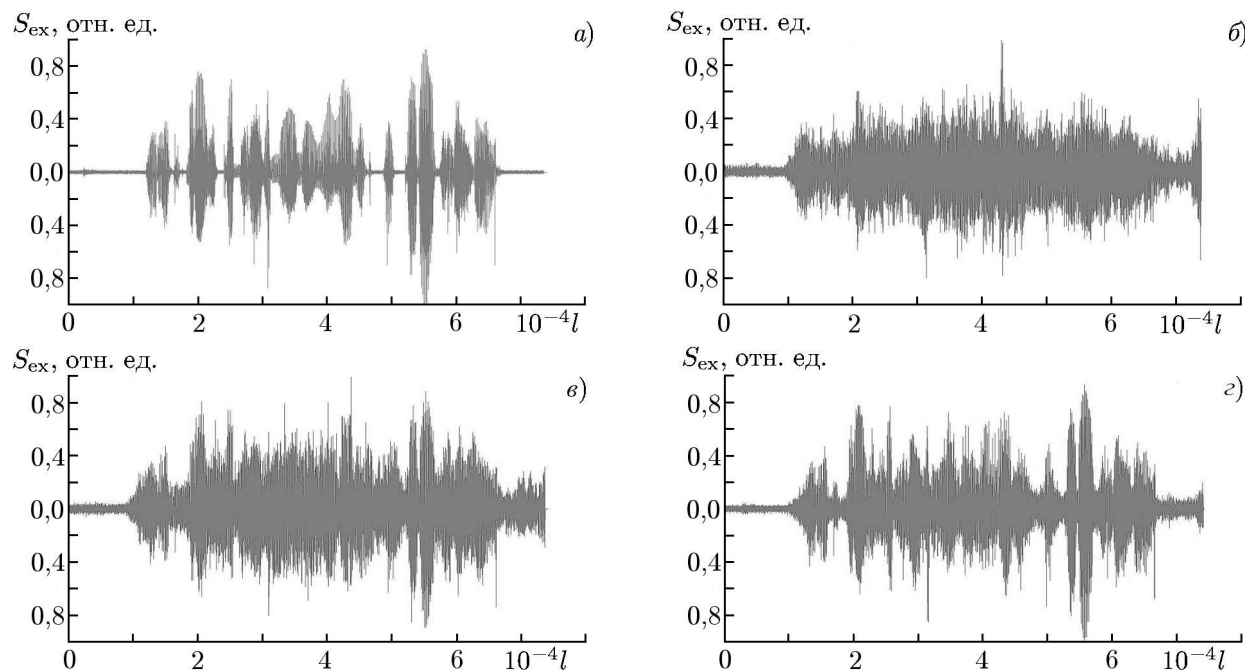


Рис. 4. Реализация исходного сигнала от источника № 5 (а); суммарный выходной сигнал от 4-х микрофонов без учёта (б) и с учётом (в) временных задержек; суммарный сигнал от 20-ти микрофонов с учётом задержек (г); l — номер отсчёта

Таблица 1. Сравнительный анализ расчёта разборчивости речи с учётом и без учёта вводимых временных задержек

	R	L , %	W , %
Без учёта задержек (см. рис. 4б)	0,159	28,74	79,19
С учётом временных задержек (см. рис. 4в)	0,264	50,14	91,38

пространственным координатам источника № 5.

На рис. 4в приведена реализация выделенного из шума суммарного сигнала. При введении задержек и фокусировке системы в точке с координатами источника полезного сигнала мощность последнего относительно помех возрастает. На рис. 4г приведена реализация выделенного из шума суммарного сигнала системой из 20-ти микрофонов. Для этой системы при использовании алгоритма введения задержек отношение сигнал/шум увеличилось на 11,5 дБ.

Для реализаций сигнала, показанных на рис. 4б, в была рассчитана разборчивость речи по методу, представленному в работе [4]. В табл. 1 приведены результаты расчёта интегрального индекса артикуляции речи R , её слоговой L и словесной W разборчивости с учётом и без учёта временных задержек (2). В табл. 2 отражены показатели понятности речи для словесной и слоговой разборчивости.

Анализируя табл. 1 и 2, можно сделать следующий вывод: с введением временных задержек, зависящих от пространственных координат точки фокусировки, результат становится удовле-

Таблица 2. Показатели понятности речи для словесной и слоговой разборчивости

	L , %	W , %
Предельно допустимая	25÷40	75÷87
Удовлетворительная	40÷50	87÷93
Хорошая	50÷80	93÷98
Отличная	более 80	более 98

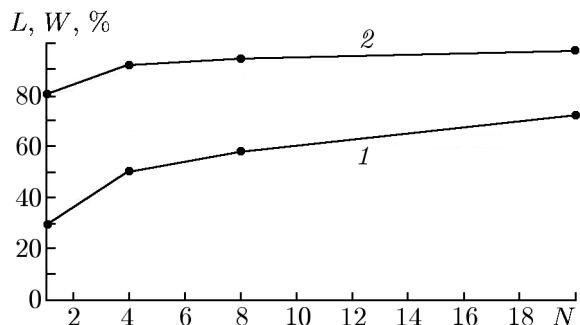


Рис. 5. Зависимость слоговой (кривая 1) и словесной (кривая 2) разборчивости от количества микрофонов N

творительным, но недостаточным для «хорошей» или «отличной» понятности речи. Необходимо повысить разборчивость выделенного сигнала. Для этого было рассмотрено влияние количества приёмных устройств на объективную оценку качества речевого сообщения. Эксперимент был проведён также для 8-ми и 20-ти микрофонов. Увеличение числа регистрирующих каналов приводит к улучшению качества выделяемого речевого сигнала при наличии помех от сторонних источников, о чем свидетельствует рис. 5.

4. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ И ВЫВОДЫ

В ходе исследования была определена разрешающая способность акустической системы, состоящей из четырёх микрофонов, которая составила 8 см. Этот результат позволяет сделать вывод о возможности однозначного разрешения двух источников речевых сигналов на акустической сцене, если расстояние между ними больше 8 см.

Далее системой из четырёх микрофонов было выделено одно речевое сообщение из помех, вносимых семью сторонними источниками. Доказано, что такое выделение возможно при введении в каждый регистрируемый канал временных задержек, которые зависят от пространственных координат точки фокусировки. Анализ разборчивости речи показал, что при использовании алгоритма введения задержек понятность речи улучшается. Это происходит за счёт роста мощности полезного сигнала, исходящего из определённой точки пространства.

Также было доказано, что при увеличении количества регистрирующих датчиков понятность выделенного из помех речевого сообщения улучшается. Так, при использовании акустической системы из четырёх микрофонов слоговая разборчивость составила всего 50 % (понятность речи удовлетворительная), в то время как при использовании 20-ти микрофонов слоговая разборчивость составила 72 % (понятность речи хорошая).

СПИСОК ЛИТЕРАТУРЫ

1. Покровский Н.Б. Расчёт и измерение разборчивости речи. М.: Гос. изд-во литературы по вопросам связи и радио, 1962. 392 с.
2. Канаков В.А., Миронов Н.А. // Современное состояние естественных и технических наук. 2014. № 17. С. 8.
3. Канаков В.А., Горда В.В. // Изв. вузов. Радиофизика. 2013. Т. 56, № 2. С. 124.
4. Железняк В.К., Макаров Ю.К., Хорев А.А. // Спецтехника. 2000. № 4. С. 39.

Поступила в редакцию 10 декабря 2015 г.; принята в печать 1 апреля 2016 г.

**THE SPEECH-MESSAGE EXTRACTION FROM INTERFERENCE INTRODUCED
BY EXTRANEEOUS DISTRIBUTED SOURCES***V. A. Kanakov and N. A. Mironov*

The problem of this study involves extraction of a speech signal originating from a certain space point and calculation of the intelligibility of the extracted voice message. It is solved by the method of decreasing the influence of interference from the speech-message sources on the extracted signal. This method is based on introducing the time delays, which depend on the space coordinates, to the recording channels. Audio records of the voices of eight different people were used as the test objects during the studies. It is proved that an increase in the number of microphones improves intelligibility of the speech message which is extracted from interference.