

Московский физико-технический институт  
(национальный исследовательский университет)

Физтех-школа радиотехники и компьютерных технологий  
Кафедра мультимедийных технологий и телекоммуникаций

ПРАКТИЧЕСКОЕ ЗАДАНИЕ №2

## Анализ независимых компонент

Назмиев Айрат, 5 курс, группа М01-305

Москва, 2024 г.

## Цель работы

В данной работе рассматривается анализ независимых компонент (АНК, англ. Independent Component Analysis, ICA) — метод слепого разделения многомерного сигнала на независимые аддитивные компоненты. Получены аналитические выражения для вычисления энтропии и ее градиента. Проведен ряд вычислительных экспериментов для различного числа смешанных сигналов, также проведен эксперимент с записью на диктофон.

## Введение

Рассмотрим простейшую модель смешивания сигналов. Пусть имеется  $N$  независимых источников сигнала и  $M$  сенсоров, для простоты рассмотрим случай  $N = M$  (в общем случае необходимо  $M \geq N$ ). Распространяясь в линейной инвариантной во времени среде без памяти, сигналы накладываются друг на друга, поэтому на сенсорах оказываются видны линейные комбинации сигналов с некоторыми неизвестными коэффициентами. Пусть каждый сигнал после дискретизации имеет длительность  $T$  отсчетов, тогда можно записать формулу:

$$\mathbf{X} = \mathbf{A}\mathbf{S}, \quad (1)$$

где  $\mathbf{S} \in \mathbb{R}^{N \times T}$  — исходные отсчеты сигналов,  $\mathbf{X} \in \mathbb{R}^{M \times T}$  — отсчеты сигналов, полученные на сенсорах,  $\mathbf{A} \in \mathbb{R}^{M \times N}$  — матрица смешения (англ. mixing matrix). Заметим, что в данной модели мы также пренебрегаем временем распространения сигнала до каждого из сенсоров. В приведенной модели задачей линейного АНК является нахождения такой матрицы «размешивания» (англ. unmixing matrix)  $\mathbf{W} \in \mathbb{R}^{N \times M}$ , чтобы данные на сенсорах  $\mathbf{Y} \in \mathbb{R}^{M \times T}$  представляли собой исходные сигналы:

$$\mathbf{Y} = \mathbf{W}\mathbf{X}. \quad (2)$$

При условии того, что  $\mathbf{A}$  имеет полный ранг, данная задача имеет точное решение (псевдо-обратная матрица  $\mathbf{A}^+$  для прямоугольной или обратная матрица  $\mathbf{A}^+ = \mathbf{A}^{-1}$  для квадратной). Однако мы считаем, что  $\mathbf{A}$  нам неизвестна, более того, в данной модели не рассматривается возможность непосредственной оценки  $\mathbf{A}$  (например, с помощью пилотов). В данном случае может быть полезным АНК, который является слепым методом и не требует знания матрицы смешивания.

АНК предполагает, что сигналы в смеси статистически независимы, кроме того, функции распределения (англ. cumulative distribution function, cdf) каждого из сигналов считаются известными. Для простоты компактности изложения предположим, на данный момент, что функции распределения у всех исходных сигналов  $\mathbf{S}_i \in \mathbb{C}^T, i = \overline{1, T}$  одинаковы и равны  $F_s$  (в конце будет дано обобщение на случай разных cdf, которое будет необходимо для последнего эксперимента с диктофоном). Применим к  $\mathbf{Y}$  функцию распределения, получим  $\mathbf{Z} = F_s(\mathbf{Y}) \in \mathbb{C}^M$ . Так, если удастся идеально восстановить исходные сигналы, то есть  $\mathbf{Y} = \mathbf{W}\mathbf{A}\mathbf{S} = \mathbf{A}^+\mathbf{A}\mathbf{S} = \mathbf{S}$ , то получим  $F_s(\mathbf{S})$ , при этом  $F_s(\mathbf{S}) \sim \mathcal{U}([0; 1]^N)$ , здесь использованы интегральное преобразование (универсальность равномерного распределения) и независимость сигналов [2]. Обозначим  $F_s = g$

Важным свойством многомерного равномерного распределения независимых величин является то, что оно имеет максимальную энтропию среди распределений с компактным носителем [1]. Воспользуемся этим свойством для нахождения unmixing matrix  $\mathbf{W}$ . Рассмотрим энтропию сигнала

ла после обработки на сенсорах, матрица  $\mathbf{Z}$  представляет собой  $T$  последовательных реализаций  $\mathbf{z}_t$ ,  $t = \overline{1, T}$  случайной величины  $\mathbf{z}$ . Энтропия  $\mathbf{z}$  равна:

$$H(\mathbf{z}) = - \int_{\mathbb{R}^M} f_z(\mathbf{t}) \log(f_z(\mathbf{t})) d\mathbf{t} = -\mathbb{E} [\log(f_z(\mathbf{t}))], \quad (3)$$

где  $f_z$  — плотность распределения  $\mathbf{z}$ . Так как мы считаем, что сигналы независимы (после идеального разделения смеси), можно расписать многомерную плотность как произведение одномерных:  $f_z(\mathbf{z}) = \prod_{i=1}^M f_z(z_i)$ , поэтому  $\log(f_z(\mathbf{z})) = \sum_{i=1}^M \log f_z(z_i)$ . Найдем оценку величины энтропии как математическое ожидание собственной информации по выборке  $\mathbf{Z}$  (второе равенство в формуле (3)):

$$H(\mathbf{Z}) = -\frac{1}{T} \sum_{t=1}^T \log(f_z(\mathbf{z}_t)) = -\frac{1}{T} \sum_{t=1}^T \sum_{i=1}^M \log(f_z(z_{t,i})). \quad (4)$$

Известно, что  $\mathbf{z} = F_s(\mathbf{y}) = g(\mathbf{y}_t)$ . Поэтому плотности вероятностей  $f_z$  и  $f_y$  случайных величин  $\mathbf{z}$  и  $\mathbf{y}$  связаны соотношением [4]:

$$f_z(\mathbf{z}) = \frac{f_y(\mathbf{y})}{\left| \frac{\partial \mathbf{z}}{\partial \mathbf{y}} \right|}, \quad (5)$$

где  $\frac{\partial \mathbf{z}}{\partial \mathbf{y}}$  — матрица Якоби преобразования, а  $\left| \frac{\partial \mathbf{z}}{\partial \mathbf{y}} \right|$ , модуль определителя данной матрицы — модуль Якобиана преобразования. Так как величины  $z_i$  независимы, то матрица Якоби диагональна, получим:

$$f_z(\mathbf{z}) = \frac{f_y(\mathbf{y})}{\prod_{i=1}^M g'(y_i)}. \quad (6)$$

Выразим значение  $f_y(\mathbf{y})$  через плотность  $f_x(\mathbf{y})$ . Так как  $\mathbf{y} = \mathbf{W}\mathbf{x}$ , то  $\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = |\det \mathbf{W}| = |\mathbf{W}|$ . Аналогично (6) можем записать:

$$f_y(\mathbf{y}) = \frac{f_x(\mathbf{x})}{|\mathbf{W}|}. \quad (7)$$

Объединяя формулы (4, 6, 7), получаем итоговое выражения для энтропии сигнала:

$$H(\mathbf{Z}) = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^M \log(g'(y_{t,i})) + \log |\mathbf{W}| - H(\mathbf{X}), \quad (8)$$

где мы заметили, что последнее слагаемое получилось равным по определению  $H(\mathbf{X})$ . Так как в модели  $H(\mathbf{X})$  не зависит от выбора матрицы  $\mathbf{W}$ , мы можем убрать его из рассмотрения как постоянное слагаемое в энтропии. Будем рассматривать следующую величину:

$$h_{\mathbf{W}}(\mathbf{Z}) = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^M \log(g'(y_{t,i})) + \log |\mathbf{W}| = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^M \log(g'((\mathbf{W}\mathbf{y})_{t,i})) + \log |\mathbf{W}|. \quad (9)$$

Ниже именно (9) будем называть энтропией. Теперь перед нами стоит задача максимизации (9), что будет означать, что  $g'(y_{t,i})$  распределены равномерно и независимы, это будет означать, что

удалось разделить смесь сигналов  $\mathbf{X}$  с помощью матрицы  $\mathbf{W}_*$  и получить исходные сигналы  $\mathbf{S}$ . Задача АНК ставится следующим образом:

$$\mathbf{W}_* = \arg \max_{\mathbf{W}} h_{\mathbf{W}}(\mathbf{Z}). \quad (10)$$

Решим данную задачу с использованием градиентного метода. Можно легко показать, что градиент (9) по unmixing matrix  $\mathbf{W}$  имеет вид [1]:

$$\frac{\partial h_{\mathbf{W}}}{\partial \mathbf{W}_{i,j}} = (\mathbf{W}^{-\top})_{i,j} + \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^M \frac{g''(y_{i,t})}{g'(y_{i,t})} x_{j,t}. \quad (11)$$

В приведенной выше формуле (11) предполагается, что все сигналы в смеси имеют одинаковую функцию вероятности  $F_s = g$ . Однако как упоминалось выше, данная формула легко обобщается на случай различных cdf  $F_{s_i} = g_i$  для каждого из исходных сигналов  $s_i$ . Повторив те же рассуждения, можно показать, что в этом сценарии выражение для энтропии будет иметь вид:

$$\frac{\partial h_{\mathbf{W}}}{\partial \mathbf{W}_{i,j}} = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^M \frac{g_i''(y_{i,t})}{g_i'(y_{i,t})} x_{j,t}. \quad (12)$$

Формула именно в таком виде будет использована в последнем эксперименте с записью на диктофоне. Градиентный шаг оптимизации (10) на итерации  $k$  будет иметь вид:

$$\mathbf{W}_k = \mathbf{W}_{k-1} + \alpha \frac{\partial h_{\mathbf{W}}}{\partial \mathbf{W}}, \quad (13)$$

где  $\alpha$  — коэффициент, отвечающий за величину градиентного шага. В данной работе в качестве функции распределения  $g$  используются гиперболический тангенс, смесь распределений Лапласа (приблизительно описывает cdf речи человека [5]) и нормальное распределение. Все формулы для  $g$  и производных приведены в приложениях, здесь укажем формулы для  $\tanh$ :

$$g(x) = \tanh(x), \quad g'(x) = 1 - \tanh^2(x) = 1 - g^2(x), \quad g''(x) = -2 \tanh(x)(1 - \tanh^2(x)) = -2g(x)g'(x) \quad (14)$$

Выбор гиперболического тангенса связан с тем, что он описывает супер-гауссово распределение (распределение с положительным коэффициентом эксцесса, которое характерно для многих «нешумовых» сигналов, например, речи), также для него просто вычисляются производные, не требующие выполнения дополнительных сложных операций, зная только значение самой функции.

## Эксперимент

В работе было поставлено три основных эксперимента: разделение смеси из двух сигналов, разделение смеси из трех сигналов и белого гауссова шума и эксперимент с записью своего голоса на фоне сильного шума на двухканальный микрофон. Важно заметить, что первые два эксперимента синтетические, данные звуковые дорожки искусственно смешиваются при помощи случайной матрицы  $\mathbf{A}$ , все сигналы в каждой смеси идеально выровнены друг относительно друга. С третьим экспериментом возникло множество сложностей, связанных с нереалистичностью модели смешения сигнала. Подробнее об этом написано в соответствующем разделе.

Код, содержащий описание алгоритма с иллюстрациями, проигрыванием аудиодорожек и комментариями, находятся в приложении к отчету, здесь приводятся основные графики и некоторые рассуждения.

## Первый эксперимент

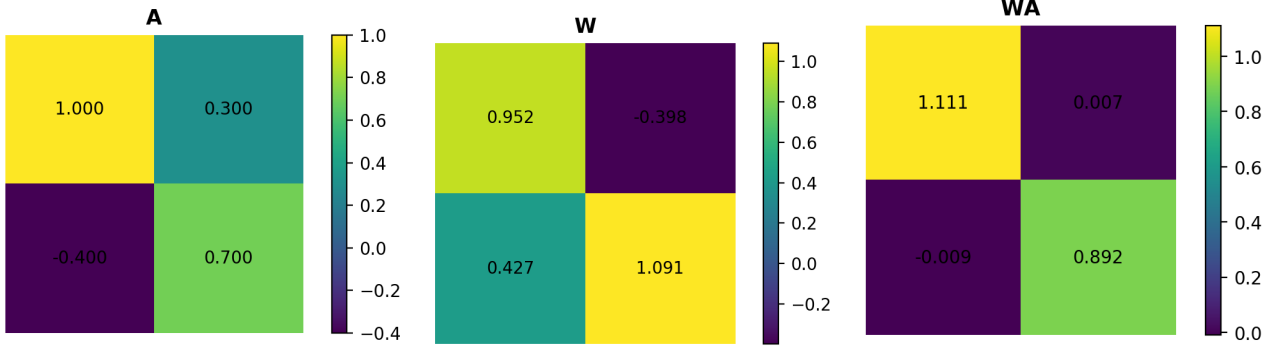


Рис. 1. Визуализация mixing matrix  $\mathbf{A}$ , unmixing matrix  $\mathbf{W}$  и их композиции  $\mathbf{WA}$

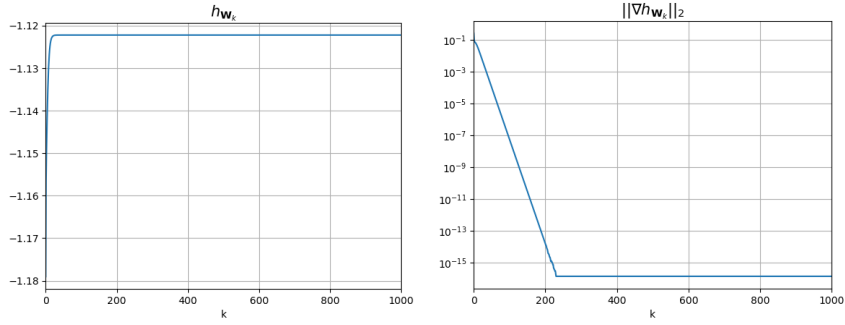


Рис. 2. Оптимизация энтропии и изменение второй нормы градиента энтропии

В данном эксперименте использовались две звуковые дорожки ( $N = M = 2$ ): пение птицы  $\mathbf{s}_1$  и гонг  $\mathbf{s}_2$ . Рассмотрим mixing matrix  $\mathbf{A}$ , приведенную на рисунке выше. Можно заметить, что в ходе оптимизации была найдена unmixing matrix  $\mathbf{W}$ , которая почти идеально разделила сигналы  $\mathbf{X}$  на двух виртуальных сенсорах (микрофонах), так что в композиции матриц  $\mathbf{WA}$  дает практически диагональную матрицу. Стоит заметить, что матрица не является единичной, так как АНК не гарантирует выравнивание сигналов по мощности, разделенные сигналы могут получиться произвольно масштабируемыми. Нормировку по уровню сигнала можно сделать уже после отработки алгоритма на полученной оценке  $\mathbf{y}_1$  и  $\mathbf{y}_2$ .

Стоит заметить, что при плохой обусловленности матрицы  $\mathbf{A}$  разделить смеси в данной простой линейной модели разделения сигналов не представляется возможным — обратная матрица будет иметь большие значения элементов, что крайне негативно скажется на процессе оптимизации, или вовсе не существовать (для вырожденной матрицы).

## Второй эксперимент

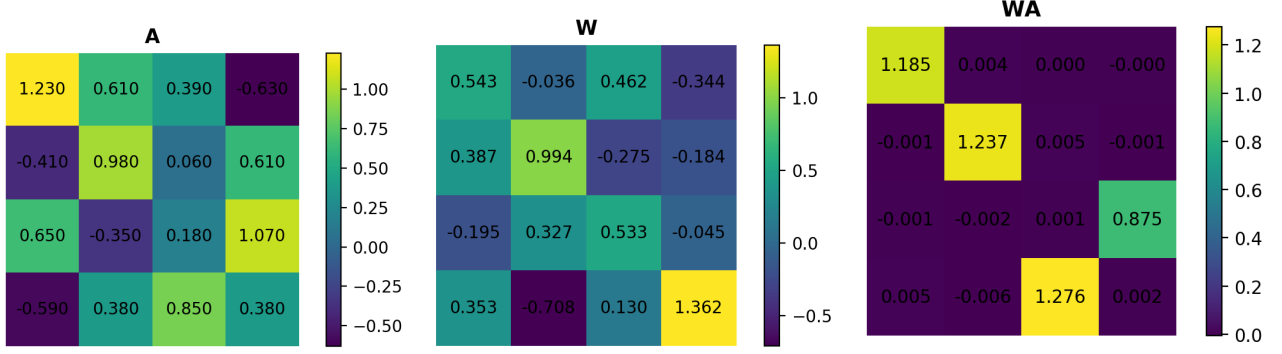


Рис. 3. Визуализация mixing matrix  $\mathbf{A}$ , unmixing matrix  $\mathbf{W}$  и их композиции  $\mathbf{WA}$

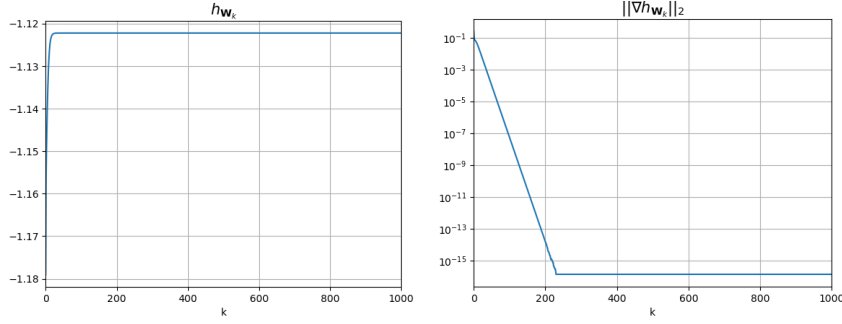


Рис. 4. Оптимизация энтропии и изменение второй нормы градиента энтропии

Во втором эксперименте использовались три записи голоса и одна дорожка с белым гауссовым шумом ( $N = M = 4$ ). Здесь также удалось полностью разделить сигналы. Однако кроме различий в нормировке уровня сигналов также можно заметить, что третий и четвертый сигнал поменялись местами. Это не проблема, ведь нам требуется просто разделить сигналы, АНК по своей постановке не дает гарантий на сохранение порядка сигналов. Таким образом, при успешной работе алгоритма из смеси будут получены исходные сигналы с точностью до нормировки по мощности и перестановки. Также стоит заметить, что АНК в такой постановке не может устойчивым образом разделить сигналы, содержащие более одной шумовой дорожки, так как любая линейная комбинация, например, двух независимых белых гауссовых случайных процессов также является белым гауссовым случайным процессом. Данный факт может привести к произвольно большим значениями коэффициентов матрицы  $\mathbf{W}$ .

## Третий эксперимент

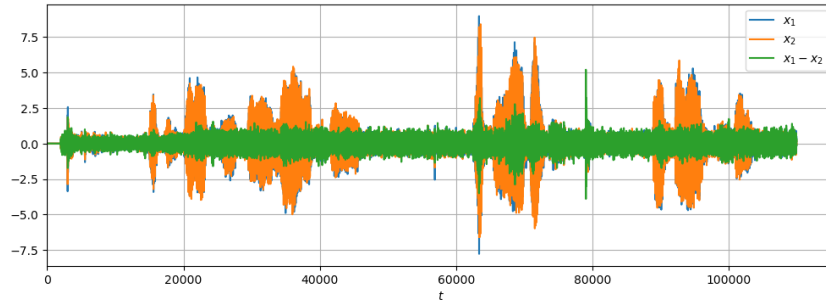


Рис. 5. Запись сигналов речи и белого гауссова шума на двухканальный микрофон

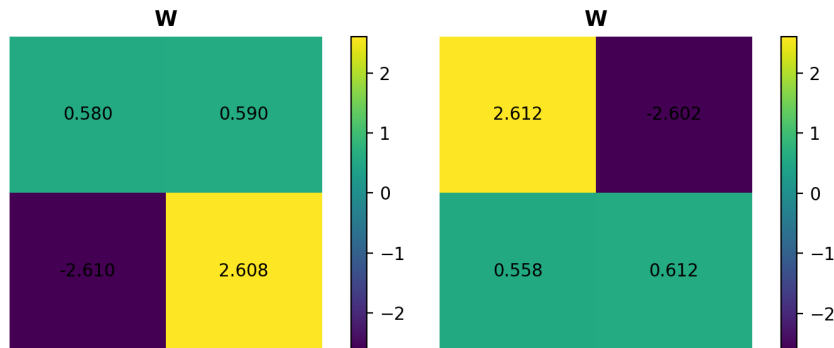


Рис. 6. Матрицы  $W$  при переставлении местами cdf

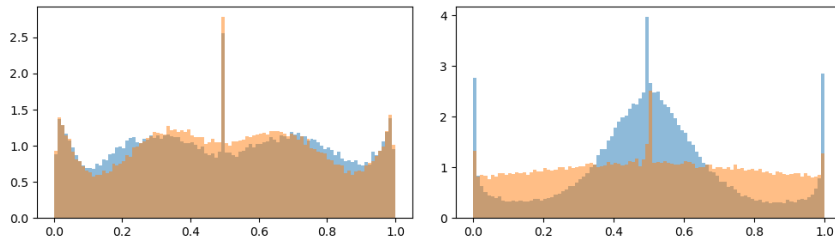


Рис. 7. Согласование с cdf для двух каналов до (синий) и после оптимизации (оранжевый)

Третий эксперимент производился с записью на диктофон голоса на фоне довольно сильного белого гауссова шума. Рассматриваемая модель смешения сигналов в реальности практически нереализуема (в особенности со звуком): есть довольно длинные импульсные характеристики канала между источниками и приемниками, которые могут значительно отличаться друг от друга, и, кроме того, изменяться во времени. Самая большая проблема заключается в том, что исходные сигналы в каждой из смесей по-разному смещены относительно друг друга, на них по-разному действовал канал. Так что модель смешения  $\mathbf{X} = \mathbf{A}\mathbf{S}$  в общем случае абсолютно не подходит для данного сценария. Однако был рассмотрен крайне выделенный случай, на котором удалось получить интерпретируемый результат.

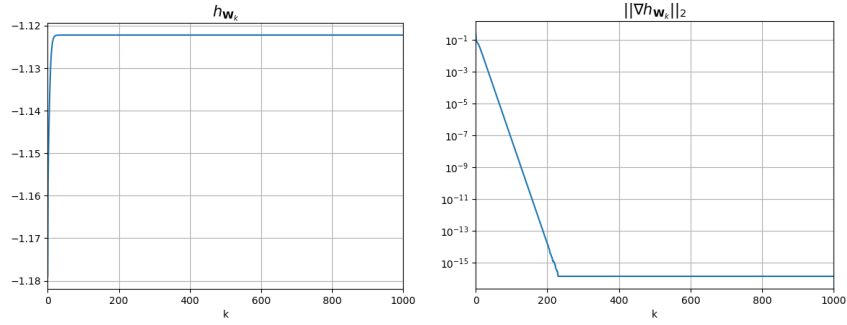


Рис. 8. Оптимизация энтропии и изменение второй нормы градиента энтропии

Для двух сигналов были выбраны различные cdf: смесь распределений Лапласа для голоса и гауссово распределение для белого гауссова шума. Диктофон был поставлен практически ровно посередине между двумя источниками так, чтобы расстояние до каждого из микрофонов для обоих из источников сигналов было близким по величине. Результат оказался довольно простым и, пожалуй, самым естественным в данном сценарии: полусумма для голоса и полуразность для шума, обе величины с некоторым коэффициентом. Интерпретация проста: усредняем шум для отделения дорожки с голосом и дифференцируем, вычитаем для шума. При реальном эксперименте в линейной модели без памяти едва ли можно было ожидать что-то лучше.

Интересно заметить, что при переставлении cdf местами также меняются местами и компоненты смеси: голос и шум согласуются со своими приближенными оценками cdf.

Сигналы также передискретизовывались, фильтровались, чтобы снизить влияние разности задержек и сильнее выделить речь на фоне шума. При необходимости сигналы незначительно выравнивались относительно друг друга. Но без описанных выше строгих условий на постановку эксперимента данные преобразования практически не имеют смысла.

## Выводы

Метод анализа независимых компонент — вычислительный метод разделения многомерного сигнала на независимые составляющие. Данный метод является слепым, то есть ему не требуется знать ни исходные сигналы, ни иметь непосредственную информацию о процессе смешивания сигналов. При достаточно слабых, естественных предположениях о независимости сигналов, при наличии некоторой общей модели смешивания есть возможность довольно эффективно разделить или «до-разделить» сигналы, что активно используется во многих приложениях как самостоятельный, так и как вспомогательный инструмент. АНК находит применение, например, в обработке изображений, звука, данных с датчиков, в беспроводной связи, медицине и других областях, работающих с физическими сигналами [3].

В данной работе была рассмотрена простейшая линейная модель без памяти, но даже она позволила в синтетических экспериментах справиться с сильно перемешанными сигналами. В реальной же задаче разделения звуковых сигналов на диктофоне рассмотренная модель смешивания имеет мало общего с физической реальностью. Существенным, как минимум, является наличие нетривиальной нестационарной импульсной характеристики каналов между источниками и микрофонами, отличие каналов друг от друга. Для решения данной задачи нужно использовать более подходящие, сверточные модели смешивания сигналов, более сложные, возможно, нелинейные модели разделения, рассматривать возможность использования полуслепых методов [3].



## Список литературы

- [1] Stone, J. V. Independent Component Analysis: A Tutorial Introduction, 2004
- [2] Casella, G. Statistical Inference, 2002
- [3] Comon, P. and Jutten, C. Handbook of Blind Source Separation: Independent Component Analysis and Applications, 2010
- [4] Кельберт М., Сухов Ю. Вероятность и статистика в примерах и задачах. Том I Основные понятия теории вероятностей и математической статистики, 2007
- [5] Радзишевский, А. Основы аналогового и цифрового звука, 2006