

ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

УДК 534.87

В. Е. Гай

МЕТОД ОБРАБОТКИ БИОМЕТРИЧЕСКОЙ ИНФОРМАЦИИ В СИСТЕМАХ КОНТРОЛЯ И УПРАВЛЕНИЯ ДОСТУПОМ

Нижегородский государственный технический университет им. Р. Е. Алексева

Приводится описание метода идентификации диктора. Рассматривается вся последовательность этапов решения задачи идентификации, начиная с построения сигнатуры диктора. Разработанный алгоритм основан на использовании теории активного восприятия, адаптированной к анализу речевых сигналов. Результаты проведённых экспериментов подтверждают возможность использования предложенного алгоритма для решения поставленной задачи.

Ключевые слова: цифровая обработка сигналов, анализ речевых сигналов, идентификация диктора.

Введение

Биометрическая информация может использоваться при решении задач обеспечения безопасности и повышения качества обслуживания. Применения включают управление правами пользования материальными и информационными ресурсами. В полуавтоматическом режиме биометрические системы позволяют решать технические проблемы, связанные с поиском идентификационной информации и формированием гипотез для обоснованного принятия решения человеком.

С точки зрения практического применения биометрических технологий актуальны вопросы взаимодействия человека с системой: время считывания идентификационных признаков, время идентификации, физические и психологические препятствия для считывания признаков, удобство и естественность протоколов взаимодействия с системой (табл. 1) [1].

Таблица 1

Сравнительная характеристика биометрических систем в %

	Отпечаток пальца	Голос	Радужная оболочка	Лицо
Equal Error Rate (EER)	2-3.3	0.1-0.86	4.1-4.6	4.1
Ошибка регистрации	4	2	7	~0
Номинальное значение вероятности «допуска чужого»	2.5	0.75	6	4
Номинальное значение вероятности «отвержения своего»	0.1	0.75	0.001	10
Стоимость системы	Высокая	Низкая	Очень высокая	Высокая

Биометрические параметры для проведения идентификации можно разделить на два класса: *статические* и *динамические*. К статическим параметрам относятся изображения отпечатков пальцев, радужной оболочки и сетчатки глаза, форма лица, форма ладони, расположение вен на кисти руки и т. д. Эти параметры практически не меняются со временем. Ди-

намические параметры – параметры, изменяющиеся во времени: голос, почерк, клавиатурный почерк, личная подпись и т. п.

Идентификация диктора – процесс определения личности по образцу голоса путём сравнения данного образца с шаблонами, сохранёнными в базе [2].

Существуют различные варианты задач идентификации (верификации), например, текстозависимая и текстонезависимая идентификация, идентификация на открытом множестве пользователей и на закрытом.

В общем случае алгоритм верификации/идентификации диктора состоит из следующих шагов:

1) этап идентификации – построение базы эталонов, выполняется определение структуры сигнала (под идентификацией системы понимается определение структуры и параметров системы по наблюдениям [3, 4]).

На данном этапе, в соответствии со сложившимся подходом к процедуре обработки речевого сигнала, используется кратковременный анализ речевого сигнала [2]. Сигнал разбивается на временные окна фиксированного размера, для более точного представления сигнала между окнами делают перекрытие, равное половине длины окна.

Далее, по каждому окну вычисляются акустические признаки. Для описания признаков обычно используются различные модели [2];

2) этап распознавания – входящий сигнал сравнивается с базой эталонов (задача опознавания сводится к сравнению признаков предъявляемых объектов с заранее известными [3]), выполняется поиск соответствия вычисленного описания – описаниям, хранящимся в базе данных (1-к-1 для решения задачи верификации, 1-ко-многим для решения задачи идентификации): в результате вычисляется расстояние от анализируемого сигнала до каждого диктора, хранящегося в базе данных (для идентификации) или расстояние от анализируемого сигнала до некоторой обобщённой модели диктора (для верификации);

3) этап принятия решения: на данном этапе принимается решение о том, является ли анализируемый диктор тем, за кого он себя выдаёт или нет.

Обычно оценка качества работы алгоритма идентификации диктора выполняется на основе следующих характеристик: FAR (False Acceptance Rate, частота ложных приёмов), FRR (False Rejection Rate, частота ложных отказов), EER (Equal Error Rate, частота ошибок) [5]. Показатели FAR и FRR связаны между собой. При уменьшении одного показателя, второй – увеличивается. Значение EER указывает, что при определённых настройках метода идентификации частота ложных приёмов равна числу ложных отказов. Чем ниже значение EER показателя, тем выше точность системы идентификации.

Рассмотрим алгоритмы, предлагаемые для решения этапов идентификации и распознавания.

1. Этап идентификации (формирование модели диктора)

Формирование описания анализируемого сигнала A выполняется на основе следующего алгоритма:

1) нормировка амплитуды сигнала на отрезке $[0; 1]$;

2) формирование U -разложения D_A сигнала A [6]. Параметры разложения: длина анализируемого сегмента L , количество используемых фильтров F ;

3) для каждого спектра разложения D_A определяются три группы, которые описывают данный спектр: полная группа на операции сложения, полная группа на операции умножения и замкнутая группа [6]. В случае если для спектра нельзя определить полную или замкнутую группу, то данный спектр в дальнейшем не используется в алгоритме;

4) с каждой полной и замкнутой группой связывается число, в дальнейшем, при сравнении групп, надо будет сравнивать не операторы, входящие в состав группы (три или четыре числа), а только одно число; для этого формируется список полных и замкнутых групп с учётом инверсий элементов;

5) отбрасываются спектры с малой массой (данные спектры соответствуют участкам пауз в сигнале);

6) отбрасываются спектры, для которых по какой-то причине нельзя было вычислить полную на сложении, полную на умножении или замкнутую группу.

Шаги 4 – 6 позволяют ускорить сравнение сигналов на этапе распознавания.

В результате, для сигнала A формируется модель M_A . Единственный параметр данного алгоритма – длина сегмента сигнала, по которому вычисляется U -спектр.

2. Этап распознавания (сравнение моделей дикторов)

На рис. 1 показаны гистограммы частот появления полных и замкнутых групп для двух разных дикторов, длина записи 32 с (512000 отсчётов, частота дискретизации 16 кГц).

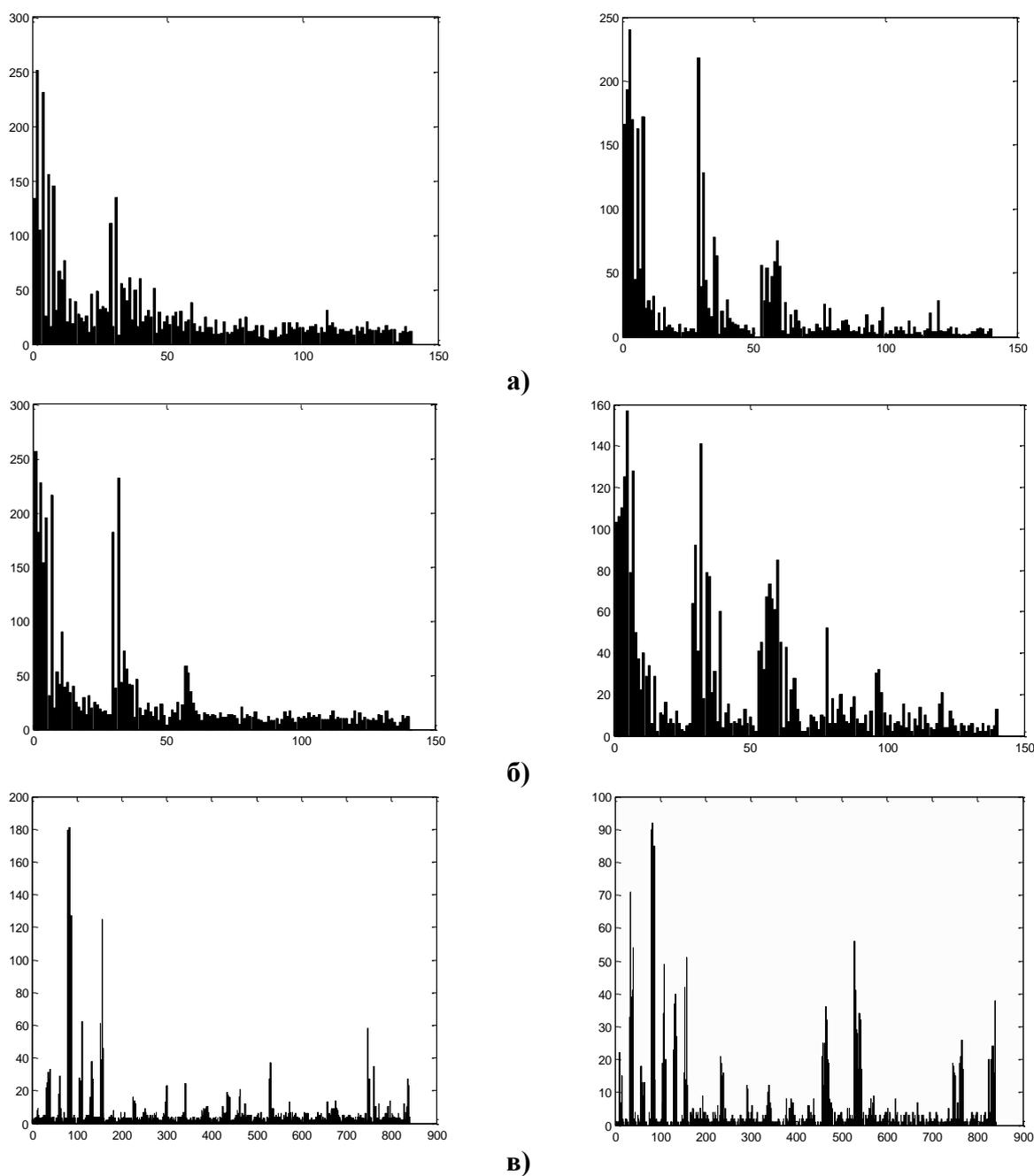


Рис. 1. Гистограммы частоты появления:

а – полных групп на операции умножения; *б* – полных групп на операции сложения;
в – замкнутых групп

Проанализировав гистограммы частот появления групп, можно отметить:

- 1) что в звуковых сигналах, содержащих записи речи дикторов, часто появляются одни и те же группы (полные и замкнутые);
- 2) частота появления разных групп – разная;
- 3) гистограммы частот появления групп для разных дикторов отличаются, указанный факт можно использовать для решения задачи идентификации дикторов.

На рис. 2 показаны гистограммы частоты появления групп, вычисленные по сигналу, представляющему собой нормальный и равномерный шум.

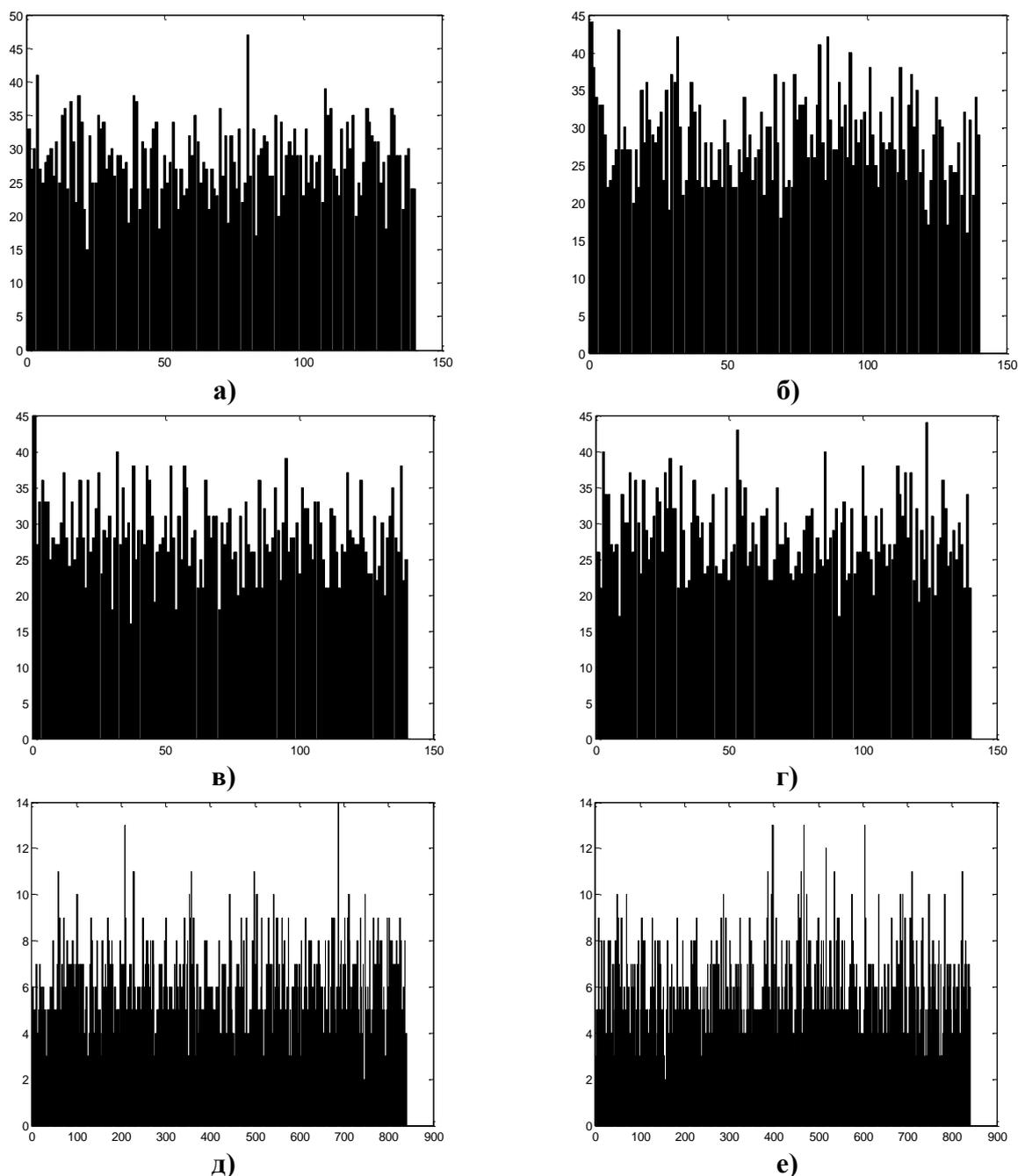


Рис. 2. Гистограммы частоты появления:
а, б – полных групп на операции умножения;
в, г – полных групп на операции сложения; *д, е* – замкнутых групп;
а, в, д – равномерный шум; *б, г, е* – нормальный шум

Рассмотрим алгоритм, который выполняет сравнение моделей двух сигналов:

1) формируются гистограммы частот появления полных и замкнутых групп, входящих в модель диктора (для каждой модели сигнала формируются три гистограммы): AH_{psi} , AH_{pnia} , AH_{pnim} , BH_{psi} , BH_{pnia} , BH_{pnim} ;

2) вычисляется расстояние между моделями:

$$DST[M_A, M_B] = \sum_{i=1}^{140} (AH_{pnia} - BH_{pnia}) + \sum_{i=1}^{140} (AH_{pnim} - BH_{pnim}) + \sum_{i=1}^{840} (AH_{psi} - BH_{psi}).$$

Меньшее значение расстояния DST соответствует большему подобию сигналов.

Похожий алгоритм используется в теории цифровой обработки изображений для сравнения двух изображений по гистограммам.

3. Исследование предложенного алгоритма

Описание эксперимента

Исследования предложенного алгоритма проводились на основе базы данных, содержащей записи голосов 20 дикторов. Рассмотрим этапы, из которых состоит эксперимент:

- 1) в i -й записи ($i \in [1; N]$, $N = 20$) выделяются два непересекающихся участка (A_i и B_i);
- 2) для A_i и B_i ($i \in [1; N]$, $N = 20$) формируются модели диктора: M_{A_i} и M_{B_i} ;
- 3) выполняется сравнение моделей дикторов на основе алгоритма, описанного в пункте 2 (формируется матрица): $D[i, j] = DST[M_{A_i}, M_{B_j}]$, $i \in [1; N]$, $j \in [1; N]$;
- 4) $\forall i \in [1; N]$ проверяется условие: если $D[i, i] = \min_{j \in [1; N]} (D[i, j])$, тогда считается, что идентификация диктора выполнена.

Результаты экспериментов

В табл. 2 приведены результаты исследования алгоритма идентификации в случае, если сравниваемые сигналы искажены равномерным шумом. Использование других параметров алгоритма при построении моделей дикторов показало невозможность устойчивой идентификации. Также результаты исследования алгоритма продемонстрировали, что при сравнении двух записей голоса одного диктора, одна из которых искажена, а другая – нет, выполнить идентификацию невозможно.

Таблица 1

Исследование алгоритма к искажениям сравниваемых сигналов

ОСШ (в Дб)	Длина сигнала	Длина сегмента	Ошибка идентификации (EER)
9	512000	512	1
15	512000	512	0
18	512000	512	0

В табл. 3 приведены временные характеристики производительности описанного алгоритма.

Таблица 2

Производительность алгоритма

Длина сегментов, на которые разбивается сигнал	Время формирования описания диктора (в секундах)	Время сравнения описаний двух дикторов (в секундах)
512	246	0.01
256	283	0.01
128	351	0.02
64	473	0.06
32	729	0.14

Анализируя табл. 3, можно отметить, что время сравнения описаний дикторов и время формирования описаний зависит от длины сегмента сигнала, по которому формируется спектр, и увеличивается с увеличением длины сегмента.

Таблица 3

Результаты идентификации дикторов

Длина сигнала	Длина сегмента	Ошибка идентификации (EER)
256000	32	15%
256000	50	10%
512000	50	5%
512000	128	0%
512000	256	0%
512000	512	0%
512000	384	0%

Результаты, приведённые в табл. 4, указывают на то, что с увеличением длины сегментов, на которые разбивается сигнал, увеличивается точность идентификации дикторов.

В [7] приводятся результаты идентификации дикторов на основе различных методов (табл. 5).

Таблица 4

Результаты идентификации на основе различных методов

Способ описания диктора	Метод сравнения	Тип идентификации	Количество дикторов	Ошибка идентификации
Кепстр	Сопоставление с образцом	Текстозависимая	10	2%@0.5с
Линейное предсказание	Долговременная статистика	Текстонезависимая	17	2%@39с
LAR (log area ratio)	Непараметрическая плотность распределения вероятностей	Текстонезависимая	21	2.5%@2с
Линейное предсказание, Кепстр	Сопоставление с образцом	Текстонезависимая	11	21%@3с

После символа «@» в столбце «Ошибка идентификации» указывается время идентификации.

Сравнивая результаты идентификации дикторов, приведённые в табл. 5, с результатами идентификации, полученными на основе предложенного алгоритма, можно отметить, что предложенный алгоритм, по качеству идентификации, не уступает другим алгоритмам идентификации дикторов.

4. Анализ полных и замкнутых групп

Анализируя результаты идентификации дикторов, можно отметить, что не для каждого сегмента сигнала можно вычислить полную и замкнутую группу.

На рис. 3 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. В табл. 6 представлены возможные сочетания групп, в столбце «Частота появления» приводится вероятность появления группы в сигнале (в процентах).

Таблица 5

Возможные сочетания групп

№ п/п	Полная группа на операции умножения	Полная группа на операции сложения	Замкнутая группа	Частота появления	Рисунок
1	-	-	-	0	-
2	+	-	-	0	-
3	-	+	-	1	рис. 3, а, б
4	+	+	-	0	-
5	-	-	+	1	рис. 3, в, г, д
6	+	-	+	0	-
7	-	+	+	0	-
8	+	+	+	98	-

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.

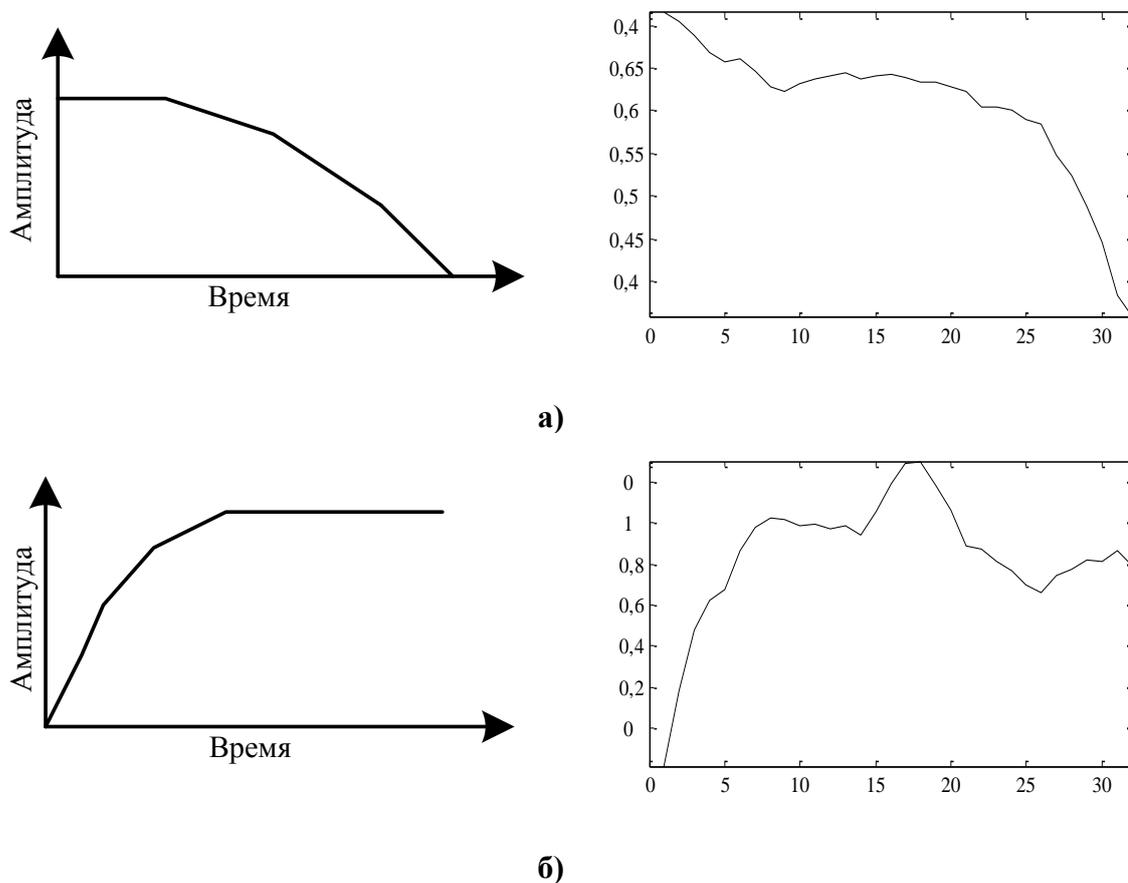
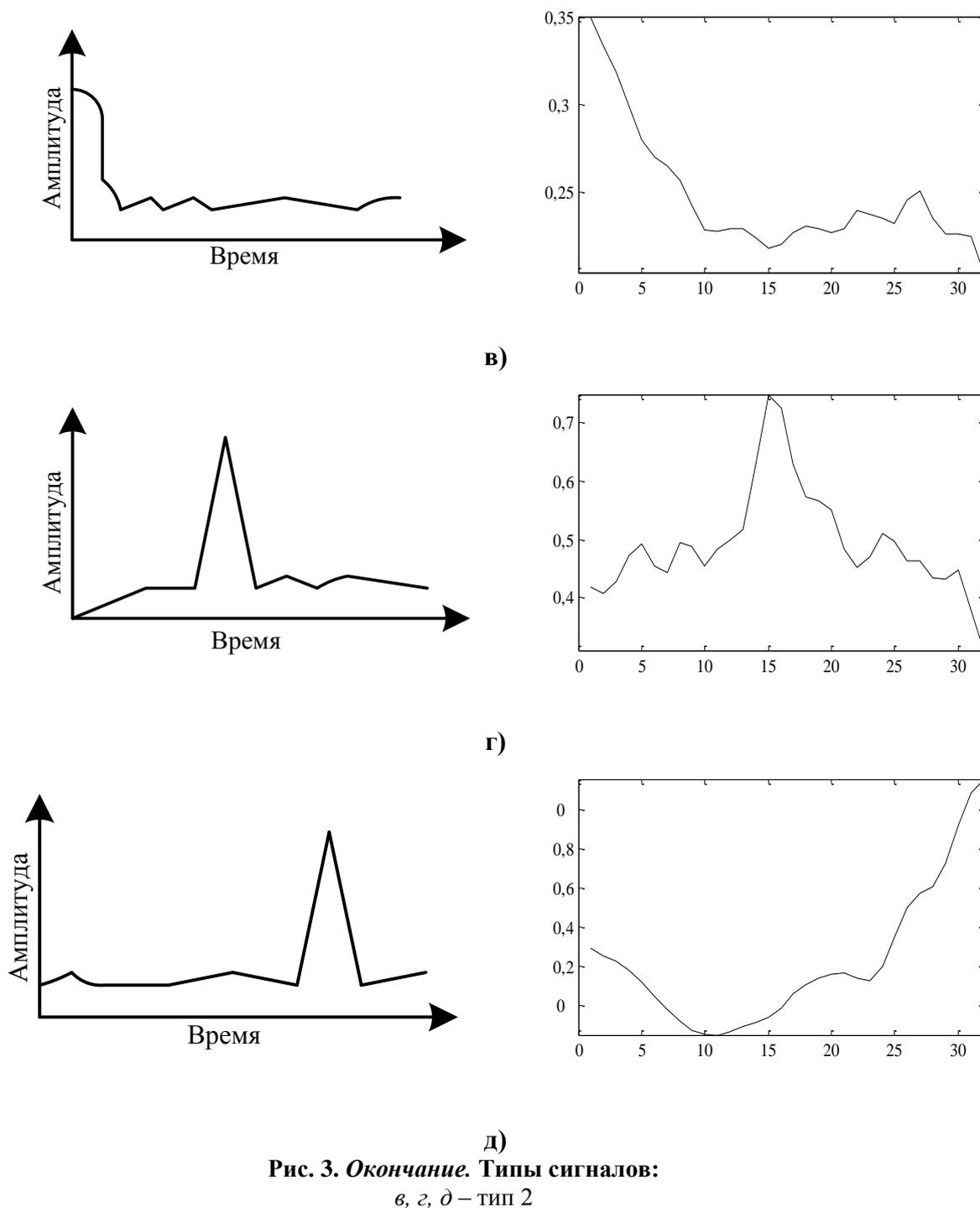


Рис. 3. Типы сигналов:
а, б – тип 1



Проверена возможность использования спектров указанных типов при решении задачи идентификации. Результаты исследования показали, что их использование влияет на результат идентификации незначительно. Также, при использовании указанных спектров увеличивается время идентификации, так как выполняется обработка большего объема данных.

Заключение

В данной работе описан метод решения задачи идентификации диктора на основе теории активного восприятия. Для описания речевого сигнала используется U -представление сигнала. Результаты экспериментов говорят об эффективности предложенного метода. К основным достоинствам предложенного метода можно отнести:

- 1) высокую точность идентификации по сравнению с другими методами идентификации дикторов;
- 2) низкую ошибку ложных срабатываний;
- 3) небольшое количество настраиваемых параметров.

Результаты исследований могут найти применение при построении систем обработки речевых сообщений, в системах текстозависимой и текстонезависимой идентификации, а также в системах верификации дикторов.

Библиографический список

1. **Зубов, Г.** Состояние рынка биометрических технологий / Г. Зубов, М. Хитров // Chip News 2007. №10 (123). С. 51-61.
2. **Первушин, Е. А.** Обзор основных методов распознавания дикторов // Математические структуры и моделирование, 2011. Вып. 24. С. 41-54.
3. **Цыпкин, Я. З.** Адаптация и обучение в автоматических системах / Я. З. Цыпкин. – М.: Наука, 1968. – 400 с.
4. **Цыпкин, Я. З.** Информационная теория идентификации / Я. З. Цыпкин. – М.: Наука. Физматлит, 1995. – 336 с.
5. **Сорокин, В. Н.** Верификация диктора по спектрально-временным параметрам речевого сигнала / В. Н. Сорокин, А. И. Цыплихин // Информационные процессы. 2009. Т. 10. №2. С. 87-104.
6. **Утробин, В. А.** Физические интерпретации элементов алгебры изображения // Успехи физических наук. 2004. Т. 174. № 10. С. 1089–1104.
7. **Anil, K. Jain, Ruud Bolle, Sharath Pankanti** Biometrics: personal identification in networked society // Springer. 1999. P. 411.

*Дата поступления
в редакцию 08.02.2013*

V.E. Gay

METHOD FOR PROCESSING BIOMETRIC INFORMATION IN SYSTEMS FOR ACCESS CONTR

Nizhny Novgorod state technical university n.a. Alexeev

Purpose: The paper considers the description of method for speaker identification.

Design/methodology/approach: As a theoretical basis for the suggested method of identification used the theory of active perception. This theory is based on the use of the U-conversion. U-transform create multilevel (roughly-precision) representation of a signal using filter Walsh Hartmut system.

Findings: Developed algorithm of speaker identification may be used in biometric control systems. Study on algorithm of point to the possibility of its use in solving the problem. Advantages of the proposed algorithm is a low computational complexity and simplicity of implementation.

Research results: The results will be used in research on the application of the theory of active perception for processing sound signals. One possible practical applications of this algorithm is its use in locating records in the databases of the audio data, while monitoring the radio.

Key words: speaker identification, digital signal processing, analysis of speech signal.