



УДК 621.397

**РАСПОЗНАВАНИЕ КЛЮЧЕВОГО СЛОВА НА ОСНОВЕ СУБПОЛОСНОГО  
ПРЕОБРАЗОВАНИЯ С ПРИМЕНЕНИЕМ АЛГОРИТМА ДИНАМИЧЕСКОГО  
ИСКАЖЕНИЯ****KEYWORD RECOGNITION BASED ON SUBBAND TRANSFORMATION  
WITH APPLICATION DYNAMIC TIME WARPING ALGORITHM****С.В. Уманец  
S.V. Umanets**Белгородский государственный национальный исследовательский университет,  
Россия, 308015, Белгород, ул. Победы, 85

Belgorod State National Research University, 85 Pobeda St, Belgorod, 308015, Russia

e-mail: umalub@bk.ru

**Аннотация**

Рассмотрен алгоритм обработки цифрового речевого сигнала, основанный на поиске заранее заданной последовательности признаков. Пространство признаков формируется путём применения субполосного кодирования. Предложены модификации алгоритма динамического искажения.

**Abstract**

An algorithm for processing a digital speech signal based on a search for a predetermined sequence of features is considered. The feature space is formed by applying subband coding. Modifications of the dynamic distortion algorithm are proposed.

**Ключевые слова:** распознавание речи, субполосные матрицы, сравнение последовательностей, алгоритм динамического искажения.

**Keywords:** speech recognition, subband matrixes, comparing sequences, dynamic time warping.

Взаимодействие человека и бытовой техники намного расширится, если управлять машиной обычным голосом в реальном времени. Такую цель можно достичь, решая следующую задачу: необходимо распознать определённое ключевое слово в естественном потоке речи одиночного диктора. При наличии ключевого слова сформировать условно-оповестительный сигнал, который в дальнейшем применении может стать причиной выработки какого-нибудь действия. Ключевое слово может быть записано заранее в словарь. Неплохих успехов в решении подобных задач добилась корпорация Google, предлагающая речевой ввод при осуществлении поиска в сети Интернет, но подробности технологии неизвестны, поэтому есть необходимость продолжить разработку технологии распознавания речи. По классификации теории распознавания [Галунов В.И., 2007] сформулированная задача относится к типу дикторозависимого распознавания с ограниченным словарём.

Для распознавания речи акустический (речевой) сигнал при помощи воспринимающих (микрофона) и оцифровывающих (дискретизирующих) устройств и машинной обработки фиксируется и преобразуется в цифровую форму. В результате дискретизации непрерывный (аналоговый) сигнал переводится в последовательность чисел. Для работы была принята частота дискретизации  $F_d=8000$  Гц и глубина кодирования по уровню 16 бит двоичного кода. Это соответствует формату WAV, наиболее распространённому формату без сжатия, применяемого в телекоммуникационных системах.

Первый этап распознавания речи – выделение различных признаков [Галунов В.И., 2007; Мясников Л.Л., 1970; Хайкин С., 2006]. Затем с помощью некоторой стратегии



обучения формируются шаблоны, с которыми в дальнейшем будут сравниваться неизвестные участки речевого сигнала. Традиционные модели строения слухового аппарата человека [Алдошина И., 2010; Бекеши Г., 1963; Белановский А.С., 1989; Вологдин Э.И. 2004] предполагают наличие способности анализировать энергию звука в зависимости от частотного интервала.

Математически такую гипотезу удобно описывать аппаратом субполосных матриц [Жиляков Е.Г., 2007]. Определение энергии в заданном частотном интервале осуществляется следующим методом. Частотная ось, нормированная по отношению к частоте дискретизации, разбивается на  $R$  равновеликих интервала:

$$v_0 = 0, \quad v_R = \pi, \quad v_r = \frac{r\pi}{R}, \quad r = 1 \dots (R-1), \tag{1}$$

где  $v_r$  – верхняя граница  $r$ -го частотного интервала;

$R$  – количество равновеликих интервалов, на которые разбивается ось частот.

Энергия в каждом частотном интервале оценивается на основе выражения:

$$P_r = \vec{x}^T \cdot A_r \cdot \vec{x}, \tag{2}$$

где  $\vec{x}$  – анализируемый отрезок сигнала, длительностью  $N$  отсчетов;

$T$  – знак транспонирования;

$A_r$  – субполосная матрица с элементами вида:

$$A_r(i, k) = \begin{cases} \frac{\sin(v_r(i-k)) - \sin(v_{r-1}(i-k))}{\pi(i-k)}, & \text{при } i \neq k \\ \frac{(v_r - v_{r-1})}{\pi}, & \text{при } i = k \end{cases}, \tag{3}$$

где  $v_r$  – верхняя граница  $r$ -го частотного интервала;

$$i = 1 \dots N;$$

$$k = 1 \dots N.$$

Субполосные матрицы имеют всего несколько собственных чисел отличных от нуля [Жиляков Е.Г., 2007]. Такое обстоятельство ускоряет вычисление квадратичной формы (2) путем использования формулы:

$$P_r = \sum_{k=1}^J \lambda_k^r (\vec{x} \vec{Q}_k^r)^2, \tag{4}$$

где  $\lambda_k^r$  – собственные числа матрицы  $A_r$ , их количество  $J$ , расположенные по возрастанию собственные числа с индексом больше  $J$  незначительны  $\lambda_{J+k}^r \approx 0$ , для практических вычислений  $J = 8$ ;

$(\vec{x} \vec{Q}_k^r)$  – скалярное произведение вектора анализируемого сигнала на собственный вектор  $\vec{Q}_k^r$  матрицы  $A_r$ , соответствующий собственному числу  $\lambda_k^r$ .

Также надо заметить, что формула (4) соответствует нейронной сети с двумя слоями [Хайкин С., 2006]. На рисунке 1 оценка энергии в  $r$ -ом частотном интервале получается на выходе второго слоя, весовые коэффициенты  $\lambda_k^r, k = 1 \dots K$  которого представляют собой собственные числа от первого до  $K$ -го субполосной матрицы  $A_r$ . Между первым слоем и вторым действует функция активации, которая соответствует действию возведения в квадрат. Первый слой состоит из нейронов, у которых весовые коэффициенты  $q_{ki}^r, k = 1 \dots K, i = 1 \dots N$  – это элементы собственных векторов  $\vec{Q}_k^r$  матрицы  $A_r$ .

Полученные таким образом оценки энергии в частотных интервалах представляют собой полную систему признаков в смысле аддитивности

$$\|\vec{x}\|^2 = \sum_{r=1}^R \vec{x}^T \cdot A_r \cdot \vec{x}. \tag{5}$$



Надо выбрать число отсчётов  $N$  для анализа исходного сигнала. Длина окна анализа не должна быть слишком большой, иначе будут проявляться эффекты наложения фонем друг на друга. Большая длина окна анализа также увеличивает объём основных операций (сложений и умножений), необходимых для дальнейших преобразований исходного сигнала. Маленькая длина окна анализа ухудшает частотное разрешение исследуемого сигнала. Из этих условий длину окна анализа  $N$  надо выбирать из диапазона от 60 до 160 отсчётов.

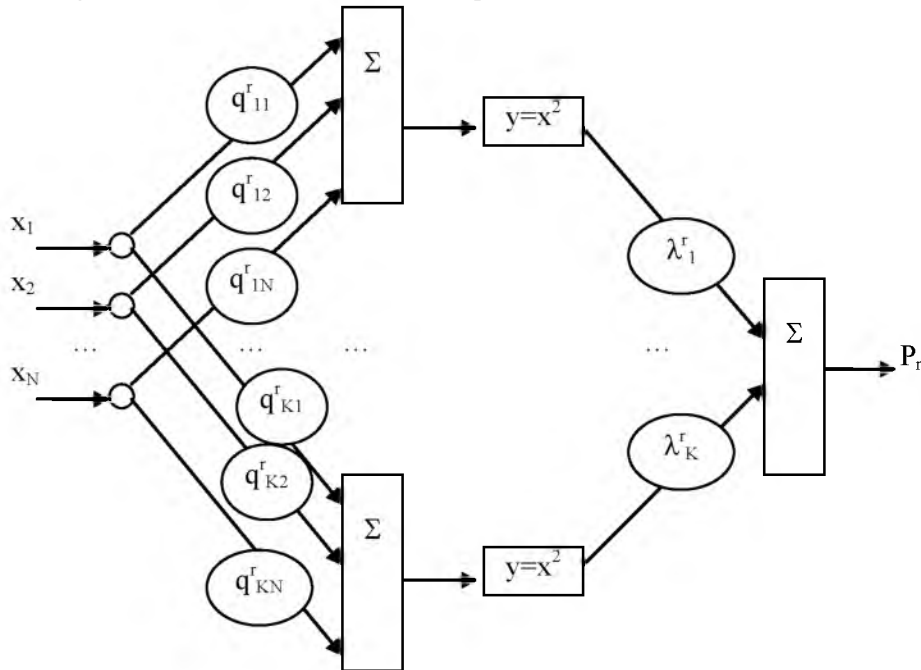


Рис. 1. Нейронная сеть для оценки энергии в частотном интервале  
 Fig. 1. Neural network for estimating energy in the frequency range

Для проведения эксперимента взяты субполосные матрицы со значениями  $N = 60$  и  $R = 15$ . Количество отличных от нуля собственных чисел взято  $J = 8$ . Из звукового файла, содержащего речевой сигнал, поступали отсчёты  $x_n$ , формирующие входной сигнал  $\vec{x}$

$$\vec{x}_n = [x_{n-N+1} \quad x_{n-N+2} \quad \dots \quad x_{n-1} \quad x_n]^T, n > N, \tag{6}$$

где  $n$  – это номер отсчёта, дискретное время.

Представляется рациональным проводить исследование сигнала, сдвигая окно анализа на один отсчёт. Тогда значение оценок энергий можно записывать с таким же обозначением момента времени, как и у входного сигнала

$$P_r(n) = \sum_{k=1}^J \lambda_k^r (\vec{x}_n \vec{Q}_k^r)^2, \quad r = 1 \dots R. \tag{7}$$

Оценки энергий целесообразно усреднить на некотором отрезке времени  $[n - Ts, n], n > Ts$ .

$$\hat{P}_r(n) = \frac{1}{Ts} \sum_{k=n-Ts}^n P_r(k), \quad r = 1 \dots R. \tag{8}$$

Величина интервала  $Ts$  должна быть соизмеримой с длиной участка речевого сигнала, на котором он стационарен. Исходя из практических наблюдений, количество фонем в одну секунду времени не превышает 25. При частоте дискретизации 8000Гц на одну фонему приходится в среднем 320 отсчётов. С учётом среднеквадратичных отклонений от среднего одной фонемы,  $Ts$  рационально выбрать из диапазона от 100 до 200 отсчётов, т. к. точность усреднения падает с уменьшением числа членов. Усреднение по способу «скользящее среднее» представляет собой [Солонина А.И., 2006] СИС-фильтр, очень экономный по вычислительной сложности

$$\hat{P}_r(n) = \frac{1}{Ts} (P_r(n) + \hat{P}_r(n-1) - P_r(n-Ts)), \quad r = 1 \dots R. \tag{9}$$



В эксперименте было взято  $T_s = 200$ .

После усреднения становится возможным анализировать оценки энергии не для всех отсчётов, а только через каждые  $T_s/2$  отсчётов.

$$\hat{E}_r(nl) = \hat{E}_r(n), \quad n = \frac{T_s}{2}nl, \quad r = 1 \dots R, \quad nl = 1, 2, 3, \dots \quad (10)$$

Для проведения простейшего этапа обучения в программе Simulink была составлена компьютерная модель. На вход модели был подан звуковой файл, в котором содержалось несколько раз произнесённое слово «пять». Результат обработки по формулам (6) – (10) представлен на рисунке 2.

Оценка энергии  $\hat{E}_r$

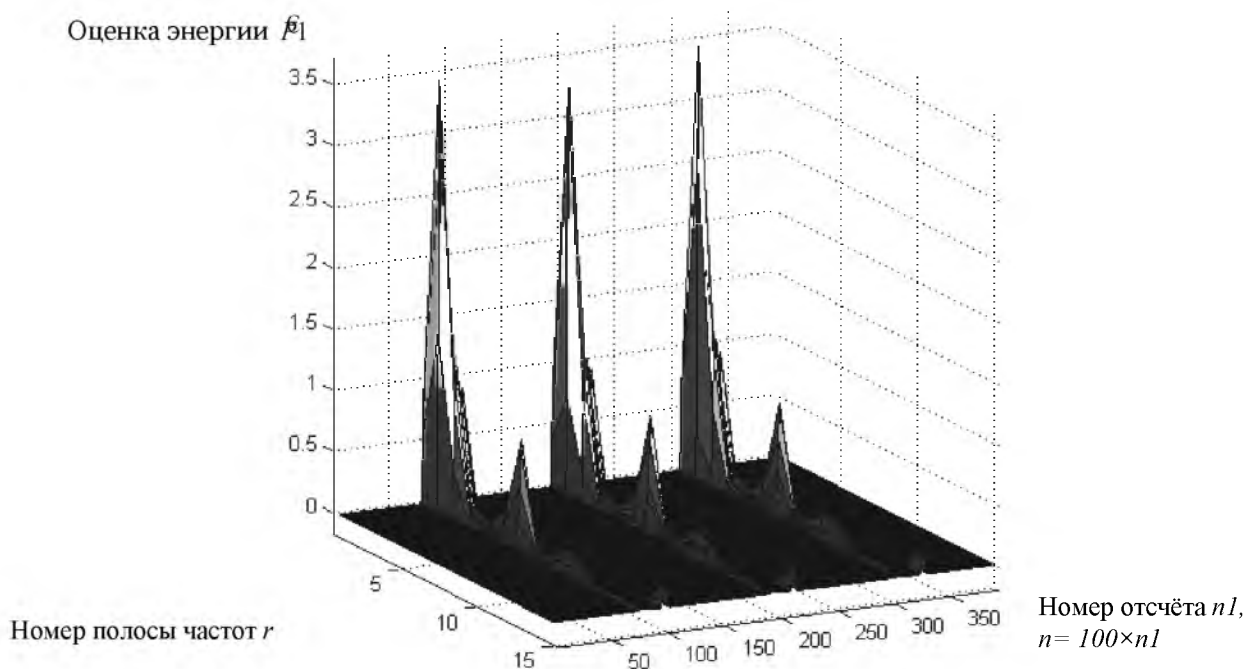


Рис. 2. Оценка энергий для слова «пять» после проведения эксперимента  
 Fig. 2. Estimation of energies for the word "five" after the experiment

Вручную выделено слово целиком, рисунок 3, и сохранено для последующего использования в качестве шаблона. Шаблон представляет собой матрицу  $P_t$  с количеством строк равному количеству анализируемых частотных интервалов. Количество столбцов соответствует количеству отобранных отсчётов  $N_{P_t}$ .

Шаблон  $P_t$

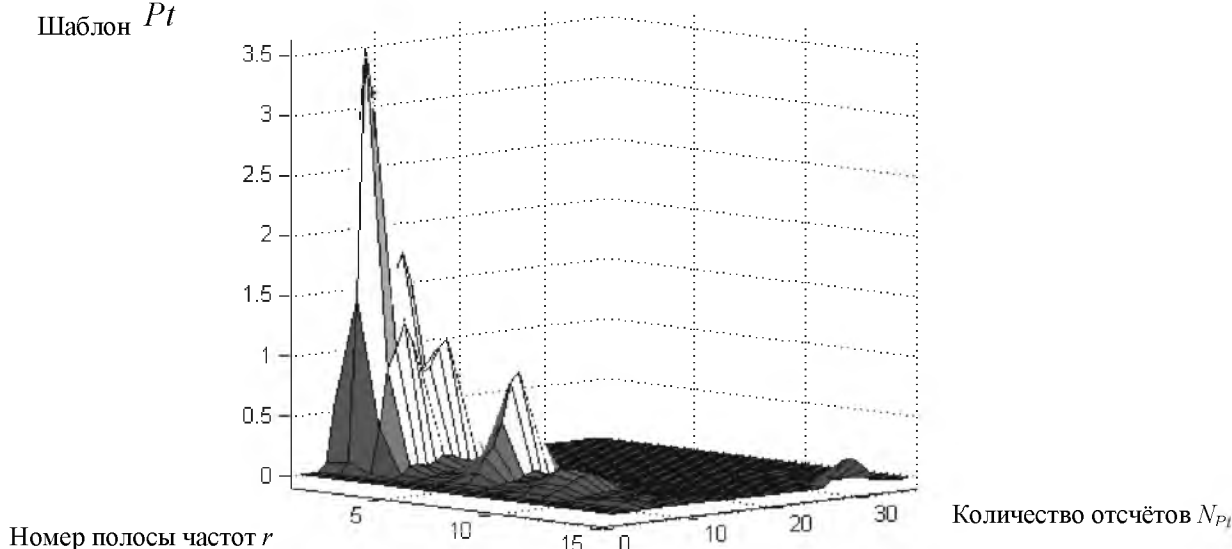


Рис. 3. Шаблон для слова «пять»  
 Fig. 3. Template for the word "five"



Для решения поставленной задачи распознавания далее надо сравнивать отрезок речевого сигнала  $\hat{f}_1(n), n \in [nlb, nle]$ , где  $nlb$  – начало отрезка,  $nle$  – конец отрезка, с записанным в словарь шаблоном  $Pt$ . Основная сложность сравнения – изменчивость слова, как по длительности, так и по звучанию. Таким образом, надо сравнить две матрицы между собой, количество строк в матрицах одинаковое и соответствует количеству рассматриваемых частотных интервалов  $R$ , а количество столбцов – разное. У матрицы шаблона количество столбцов фиксированное  $N_{pt}$ , у матрицы распознаваемого речевого сигнала количество столбцов  $N1 = nle - nlb$  берётся по длительности звучания предполагаемого слова. При таком подходе эти матрицы представляют собой последовательности с разным количеством элементов, которыми являются столбцы.

Для измерения схожести двух последовательностей, которые имеют разное количество элементов, часто применяется алгоритм динамического искажения (dynamic time warping). Сравнение производится по правилу «наилучшего соответствия» [Pavel S., 2008; Sakoe H., 1978]. По этому алгоритму сначала надо составить матрицу  $D$  с расстояниями между элементами последовательностей. Предлагается использовать относительные меры различия [Семкин Б.И., 2008; Скворцов В.А., 2002] при определении расстояния:

$$D(i, k) = \frac{\sum_{r=1}^R |\hat{f}_1(r, i) - Pt(r, k)|}{\sum_{r=1}^R \hat{f}_1(r, i) + Pt(r, k)}, \quad i = 1 \dots N1, \quad k = 1 \dots N_{pt}. \quad (11)$$

В знаменателе (11) нет знака взятия модуля, потому что оценки энергии всегда положительные.

Затем составляется матрица  $C$  с накопительной дистанцией между одиночными элементами двух последовательностей:

$$C(i, k) = D(i, k) + \min(C(i, k-1), C(i-1, k), C(i-1, k-1)), \quad (12)$$

где минимум берётся среди трёх соседних элементов;

$$\begin{aligned} C(1, 1) &= 0, \quad C(1, k) = \infty, \quad C(i, 1) = \infty; \\ k &= 2 \dots N_{pt}; \\ i &= 2 \dots N1. \end{aligned}$$

В решаемой задаче отсчётом является столбец со значениями оценок энергии.

Затем в матрице  $C$  восстанавливается путь из точки  $(N1, N_{pt})$  в точку  $(1, 1)$  по правилу движения в сторону с наименьшим значением:

$$(i_s, k_s) = \arg \min(C(i, k-1), C(i-1, k), C(i-1, k-1)), \quad (13)$$

где  $s$  – это номер шага по пути от  $(N1, N_{pt})$  до  $(1, 1)$ . Итоговое количество шагов  $S$  не постоянно, шаги совершаются пока оба индекса  $i$  и  $k$  не изменятся от своих максимальных значений  $N1$  и  $N_{pt}$  до 1.

Затем вычисляется общее различие  $SH$  между двумя последовательностями. В классическом варианте алгоритма:

$$SH = \sum_{s=1}^S D(i_s, k_s), \quad (14)$$

в данной работе предлагается

$$SH = \max_x (D(i_s, k_s)), s = 1 \dots S. \quad (15)$$

При сравнении неизвестной последовательности с шаблоном решение о соответствии шаблону принимается после сравнения  $SH$  с заданным порогом.

Для эксперимента один диктор надиктовал пять звуковых файлов с одинаковым содержанием: произнесены по порядку числительные: «1», «2», «3», «4», «5», «6», «7», «8», «9» и «0». Речевой сигнал из звуковых файлов сравнивался с заранее сохранённым шаблоном для слова «пять». Оценки энергий  $\hat{f}_1(n)$  на шаге  $n$  формировались в матрицу с количеством столбцов  $N1 = 50$ , т. к. у шаблона количество столбцов оказалось  $N_{pt} = 36$ .



Матрица неизвестного сигнала берётся на большее количество отсчётов, чем у шаблона (50>36), в расчёте на возможно медленное произнесение искомого слова.

Результат моделирования на рисунке 4 для речевого сигнала из файла №2 и на рисунке 5 – из файла №4.

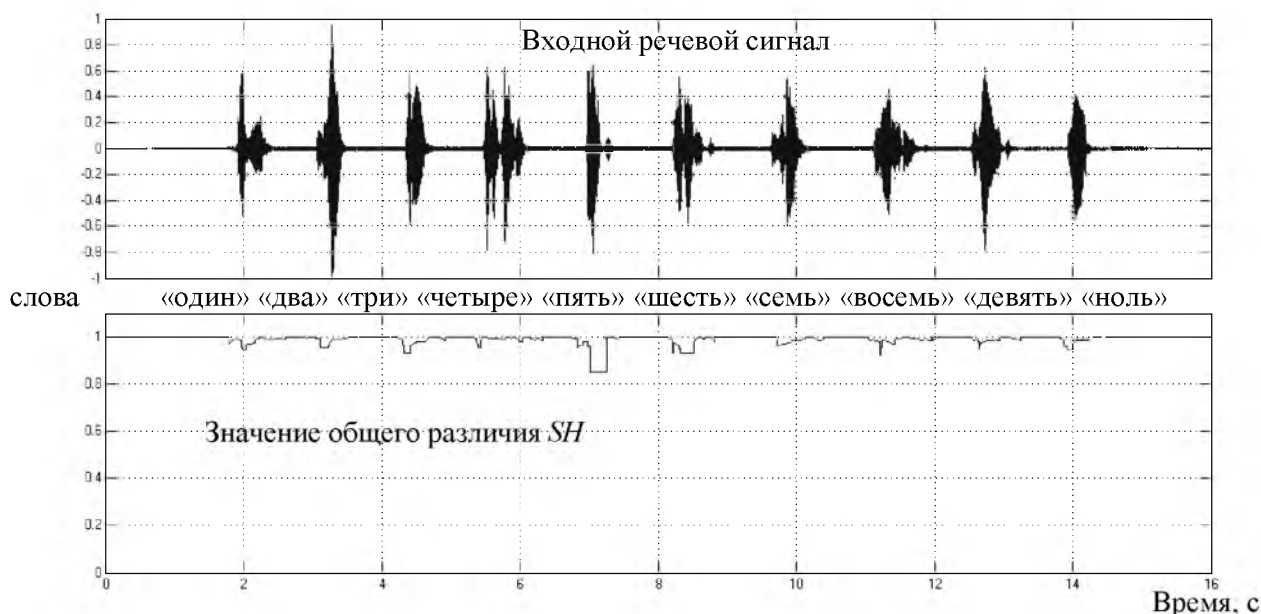


Рис. 4. Графики входного сигнала и величины отлчия от шаблона после использования относительной меры. Входной сигнал из файла №2

Fig. 4. Plots of the input signal and the value of the difference from the template after using relative measure. Input signal from file №2

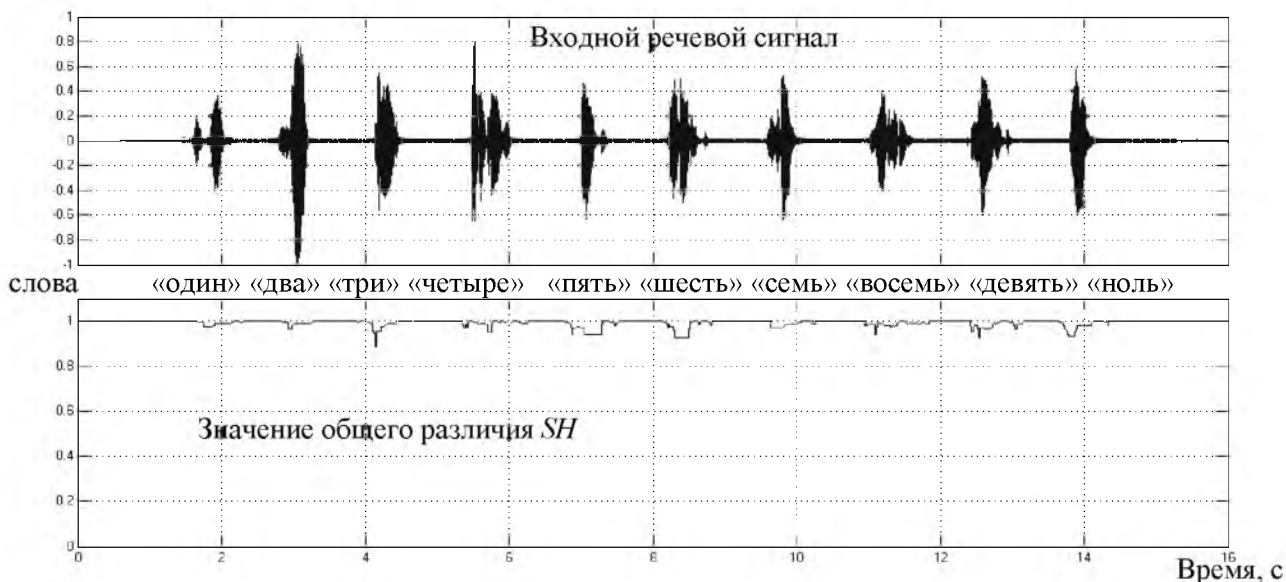


Рис. 5. Графики входного сигнала и величины отлчия от шаблона после использования относительной меры. Входной сигнал из файла №4

Fig. 5. Plots of the input signal and the value of the difference from the template after using relative measure. Input signal from file №4

На рисунке 4 графики показывают чёткую реакцию, а на рисунке 5 более убедительной кажется реакция на слово «шесть», хотя для слова «пять» отличие от шаблона меньше, чем для других слов.

Для повышения чёткости при анализе различия между шаблоном и входным сигналом можно воспользоваться нелинейным преобразованием величины оценок энергии. Согласно

эмпирическому психофизиологическому закону Вебера-Фехнера [Вебера Э., 1834; Фехнера Г., 1860] сила ощущения пропорциональна логарифму интенсивности раздражителя. Другой учёный, Стивенсон [Стивенсон С. С., 1957], предложил использовать степенную функцию для описания зависимости силы ощущений от величины раздражения. Показатель степенной функции для разных ощущений используется разный от 3,5 до 0,67 в экспериментах Стивенсона.

Для решения поставленной задачи распознавания можно применить нелинейность извлечение квадратного корня из величины оценок энергий, что будет похоже на степенной закон Стивенсона, но не внесёт заметных вычислительных сложностей.

Новые оценки с учётом нелинейностей:

$$W(n1) = \sqrt{k1(n1)} \quad (16)$$

$$W_{pt} = \sqrt{Pt}$$

Результат моделирования при наличии нелинейности на рисунке 6 и 7.

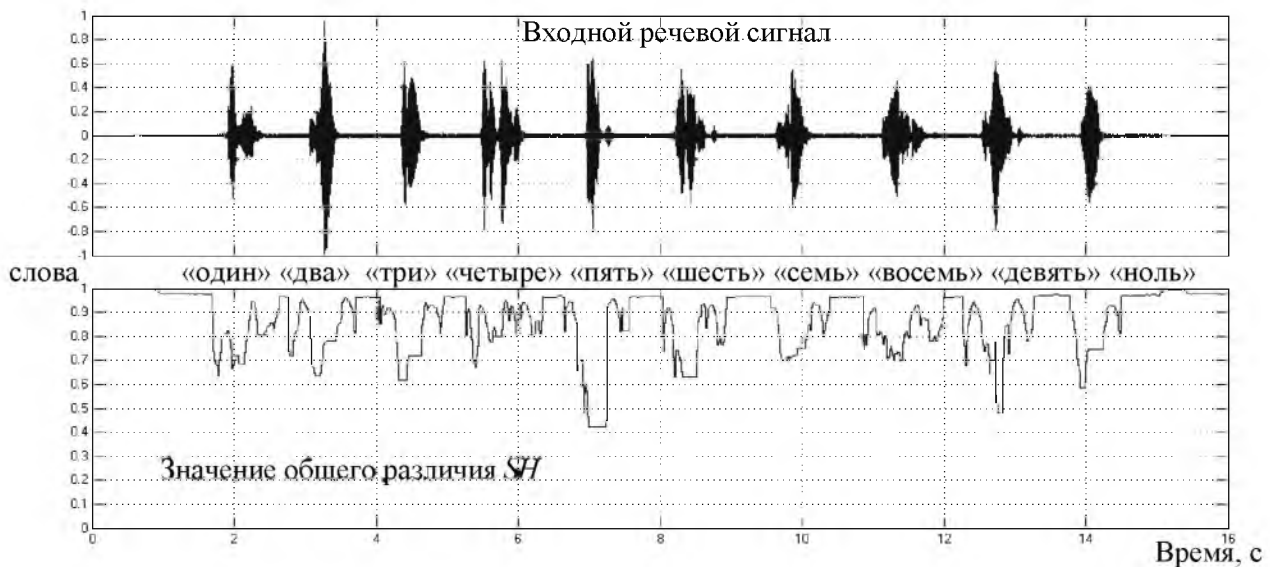


Рис. 6. Графики входного сигнала и величины отлчия от шаблона после использования относительной меры и нелинейности. Входной сигнал из файла №2

Fig. 6. Plots of the input signal and the value of the difference from the template after using relative measure and nonlinear. Input signal from file №2

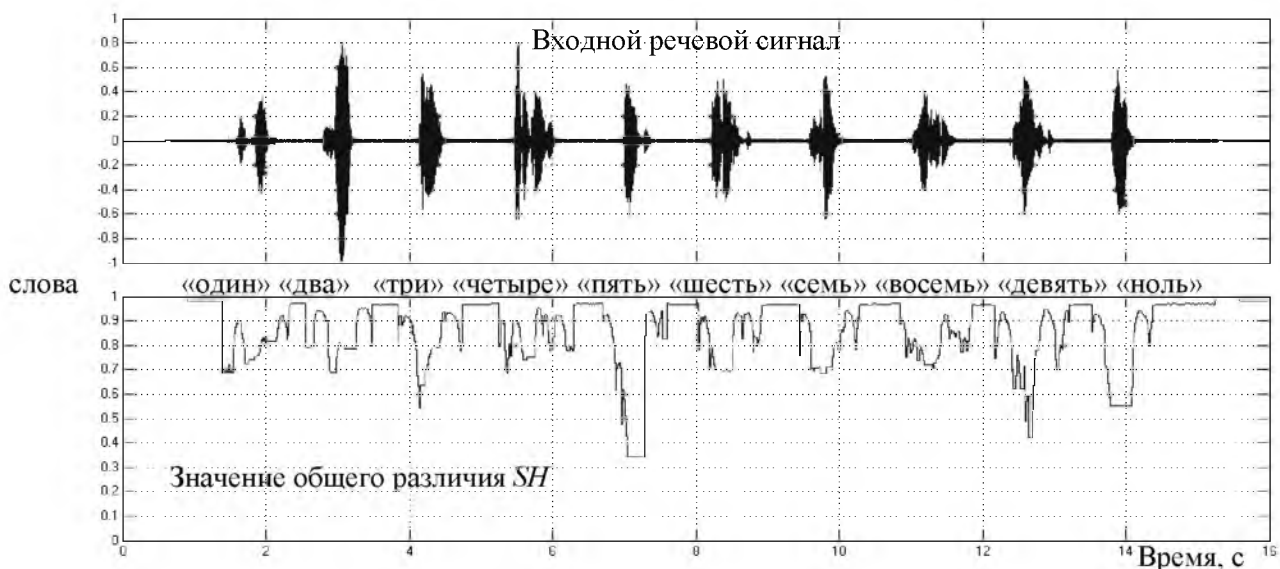


Рис. 7. Графики входного сигнала и величины отлчия от шаблона после использования относительной меры и нелинейности. Входной сигнал из файла №4

Fig. 7. Plots of the input signal and the value of the difference from the template after using relative measure and nonlinear. Input signal from file №4



Применение нелинейности позволило повысить различительную способность алгоритма. Минимум для слова «пять» выделяется сильнее по отношению к другим словам и графикам предыдущих экспериментов. Для наглядности на рисунках 8-10 приведены графики для речевых сигналов из файлов №1, №3, №5. Поведение графиков одинаковое.

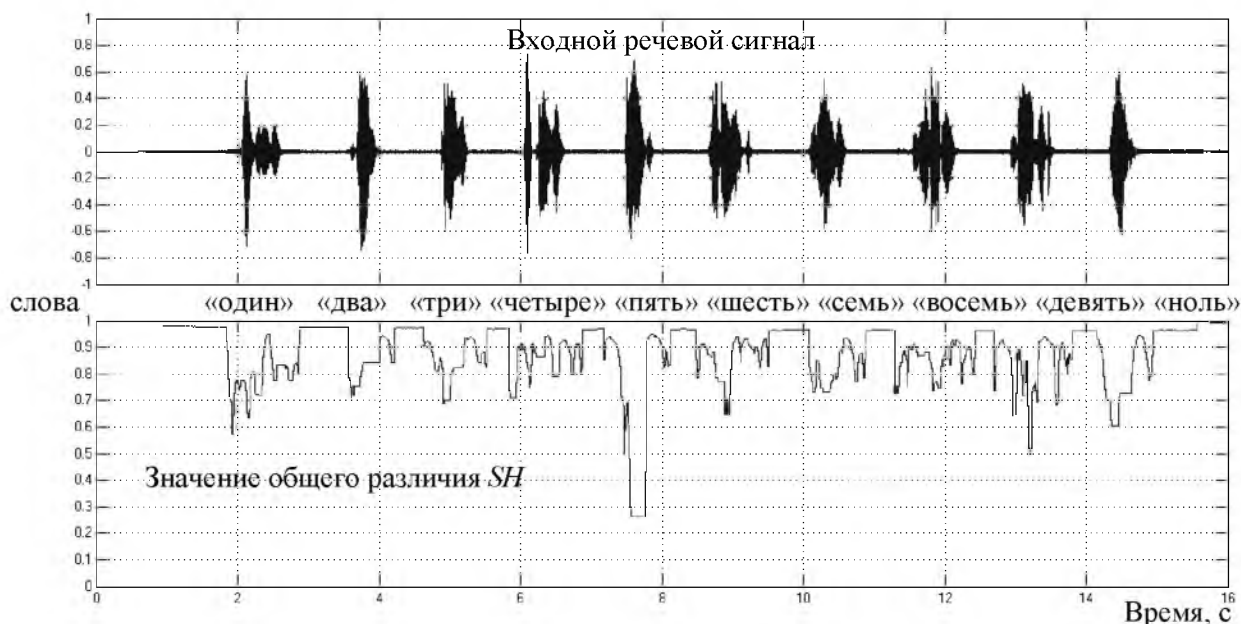


Рис. 8. Графики входного сигнала и величины отклонения от шаблона после использования относительной меры и нелинейности. Входной сигнал из файла №1

Fig. 8. Plots of the input signal and the value of the difference from the template after using relative measure and nonlinear. Input signal from file №1

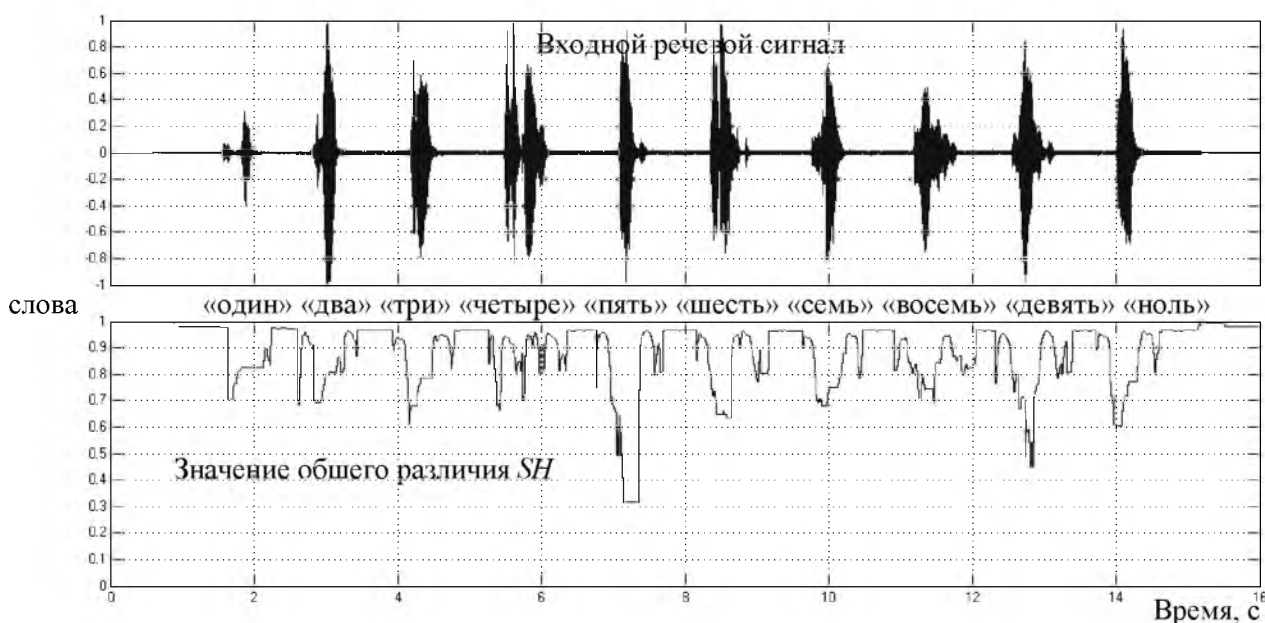


Рис. 9. Графики входного сигнала и величины отклонения от шаблона после использования относительной меры и нелинейности. Входной сигнал из файла №3

Fig. 9. Plots of the input signal and the value of the difference from the template after using relative measure and nonlinear. Input signal from file №3



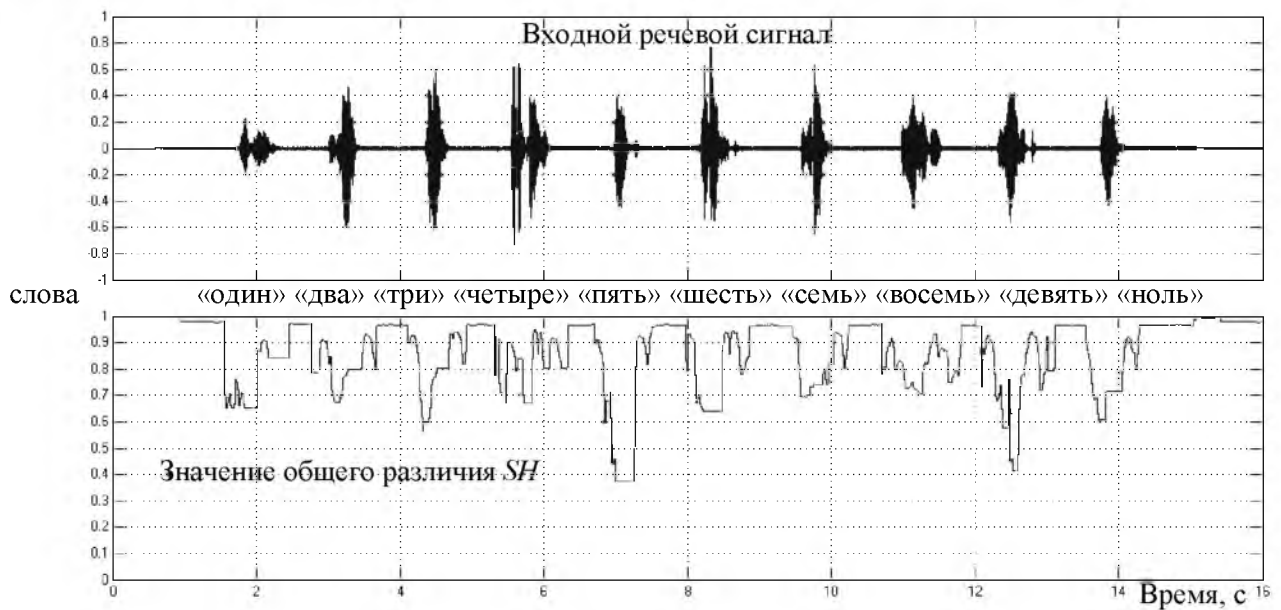


Рис. 10. Графики входного сигнала и величины отлчия от шаблона после использования относительной меры и нелинейности. Входной сигнал из файла №5

Fig. 10. Plots of the input signal and the value of the difference from the template after using relative measure and nonlinear. Input signal from file №5

Пороговое значение необходимо определить с учётом разнообразия произношения ключевого слова. Для этого из располагаемых звуковых файлов вручную были выделены и сохранены в переменные матрицы  $\mathcal{F}_i$ ,  $i=1,2,3,4,5$  с оценками энергии для слова «пять». Затем матрицы сравнивались каждая с каждой алгоритмом динамического искажения с применением относительной меры между столбцами и нелинейным преобразованием оценок энергии путём извлечения квадратного корня.

$$SH_{ij} = DTW\left(\sqrt{\mathcal{F}_i}, \sqrt{\mathcal{F}_j}\right), \quad (17)$$

где  $i, j=1,2,3,4,5$  – индексы файлов, откуда было взято слово «пять»,

$\mathcal{F}_i, \mathcal{F}_j$  – матрицы признаков для слова «пять», элементы матрицы вычисляются по формуле (4) с последующим применением скользящего среднего (9); количество строк в матрице равно количеству частотных интервалов, количество столбцов соответствует длительности звучания слова после прореживания (10), определено вручную;

$DTW()$  – функция алгоритма динамического искажения, описанная формулами (11)-(13), (15).

В отличие от непрерывного речевого сигнала, нет необходимости скользить окном анализа, отмечая величину отлчия в каждый момент времени. На данном этапе имеются фиксированные матрицы, описывающие слово «пять». Поэтому в результате сравнения получается одно число, а не график во времени.

Результаты парных сравнений показывают, насколько разнообразно произносилось слово «пять» одним диктором. Полученные значения приведены в таблице 1.

Таблица 1

Table 1

Величина различия  $SH_{ij}$  между словами «пять» одного диктора  
Value of difference  $SH_{ij}$  between word «five» of single announcer

Откуда взято слово «пять»	Из файла №1	Из файла №2	Из файла №3	Из файла №4	Из файла №5
Из файла №1	0	0,470	0,755	0,414	0,576
Из файла №2	0,470	0	0,450	0,238	0,249
Из файла №3	0,755	0,450	0	0,544	0,405
Из файла №4	0,414	0,238	0,544	0	0,281
Из файла №5	0,576	0,249	0,405	0,281	0



Среднее значение по этой выборке, игнорируя нули на диагонали, составляет 0.438, среднеквадратичное отклонение 0.161. По этим числам можно принять порог из диапазона от  $h_{min}=0.438-0.161=0.277$  до  $h_{max}=0.438+0.161=0.599$ . Пусть  $h=0.5$ , чтобы уменьшить ошибку типа «пропуск цели».

Чтобы уменьшить реакцию на одиночные случайные выбросы, сравнение с порогом предлагается учитывать не на одном шаге вычислений, а на десяти.

$$Flag = (SH(n1) < h) \wedge (SH(n1-1) < h) \wedge \dots \wedge (SH(n1-9) < h), \tag{18}$$

где *Flag* – логический результат сравнения с порогом,

*SH*(*n1*) – значение общего различия, вычисленное на шаге *n1*.

Результат отработки сравнения с порогом на рисунках 11-15.

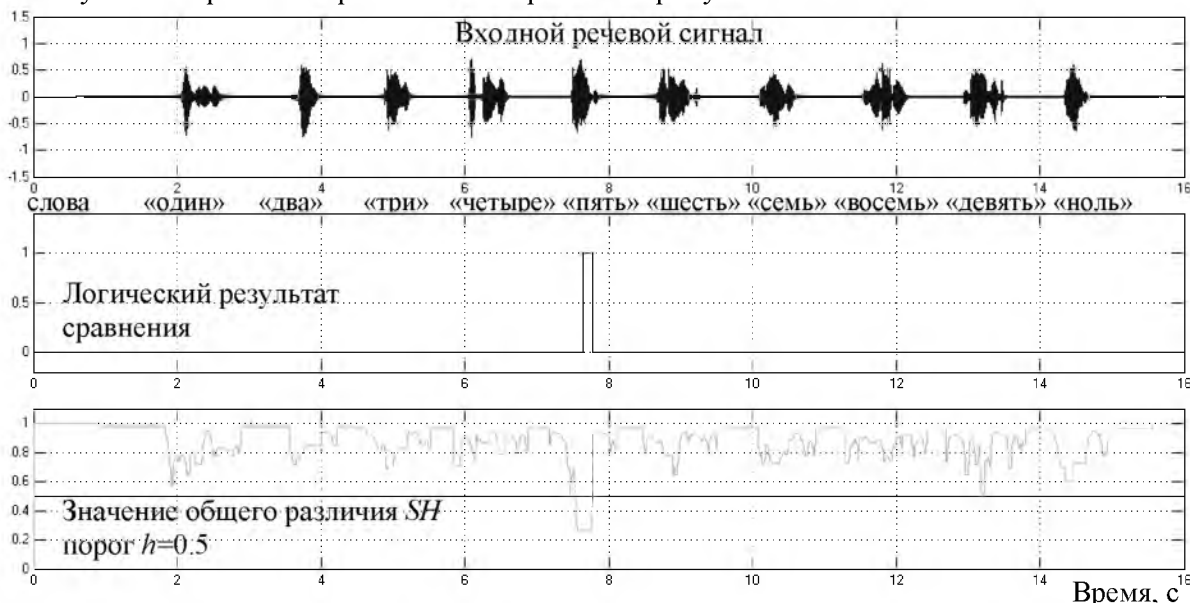


Рис. 11. Графики: входной сигнал (верхний), флаг обнаружения слова (средний) и величина отлчия от шаблона для речевого сигнала из файла №1 вместе с пороговым значением 0.5 (нижний)  
 Fig. 11. Plots: input signal graphs (upper), word detection flag (average) and the difference values from the template for the speech signal from file № 1 together with a threshold value of 0.5 (lower)

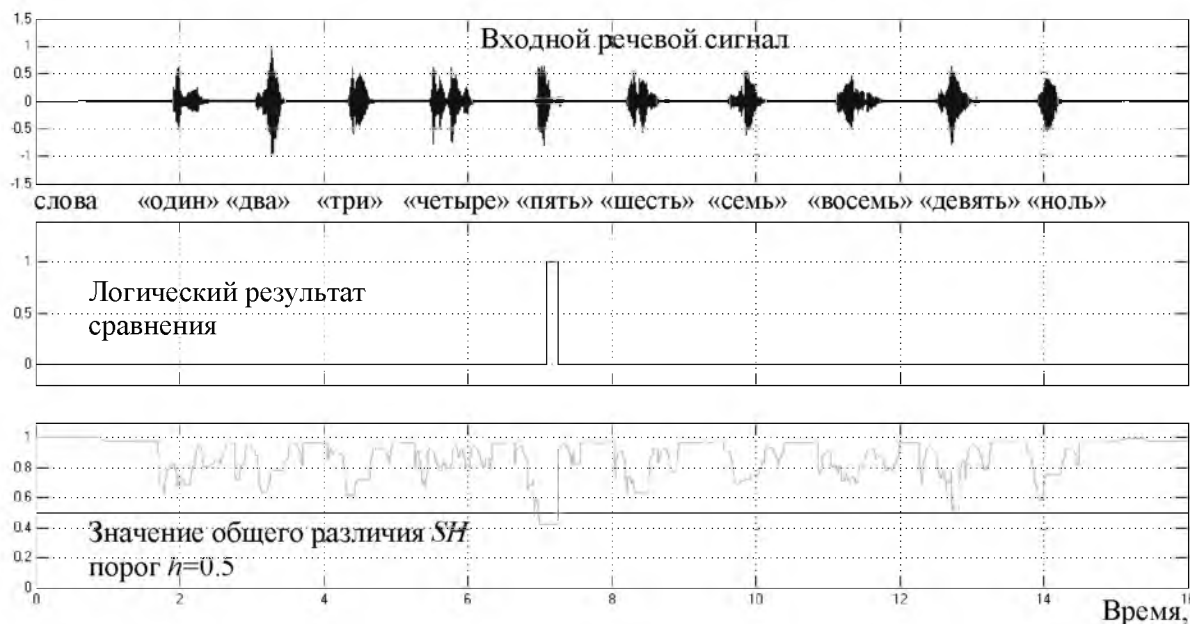


Рис. 12. Графики: входной сигнал (верхний), флаг обнаружения слова (средний) и величина отлчия от шаблона для речевого сигнала из файла №2 вместе с пороговым значением 0.5 (нижний)  
 Fig. 12. Plots: input signal graphs (upper), word detection flag (average) and the difference values from the template for the speech signal from file № 2 together with a threshold value of 0.5 (lower)

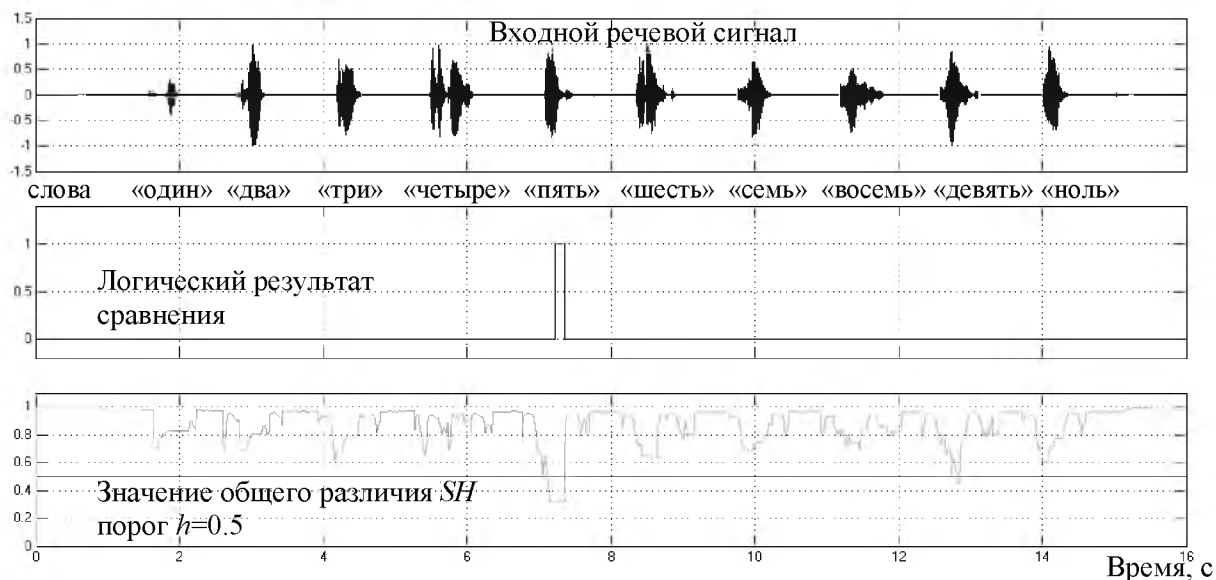


Рис. 13. Графики: входной сигнал (верхний), флаг обнаружения слова (средний) и величина отличия от шаблона для речевого сигнала из файла №3 вместе с пороговым значением 0.5 (нижний)  
Fig. 13. Plots: input signal graphs (upper), word detection flag (average) and the difference values from the template for the speech signal from file № 3 together with a threshold value of 0.5 (lower)

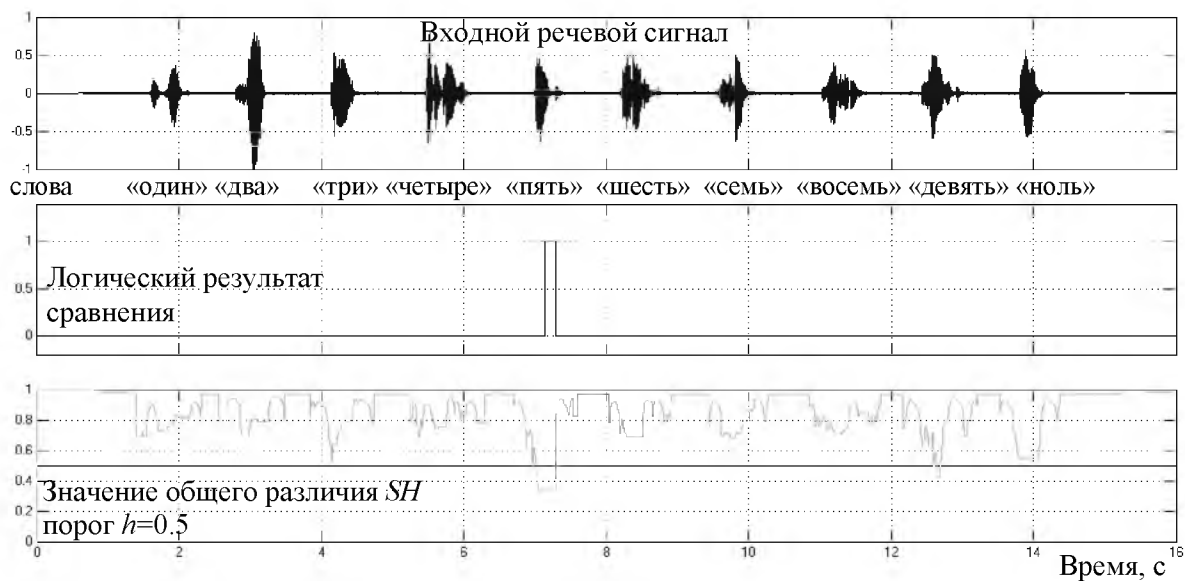


Рис. 14. Графики: входной сигнал (верхний), флаг обнаружения слова (средний) и величина отличия от шаблона для речевого сигнала из файла №4 вместе с пороговым значением 0.5 (нижний)  
Fig. 14. Plots: input signal graphs (upper), word detection flag (average) and the difference values from the template for the speech signal from file № 4 together with a threshold value of 0.5 (lower)

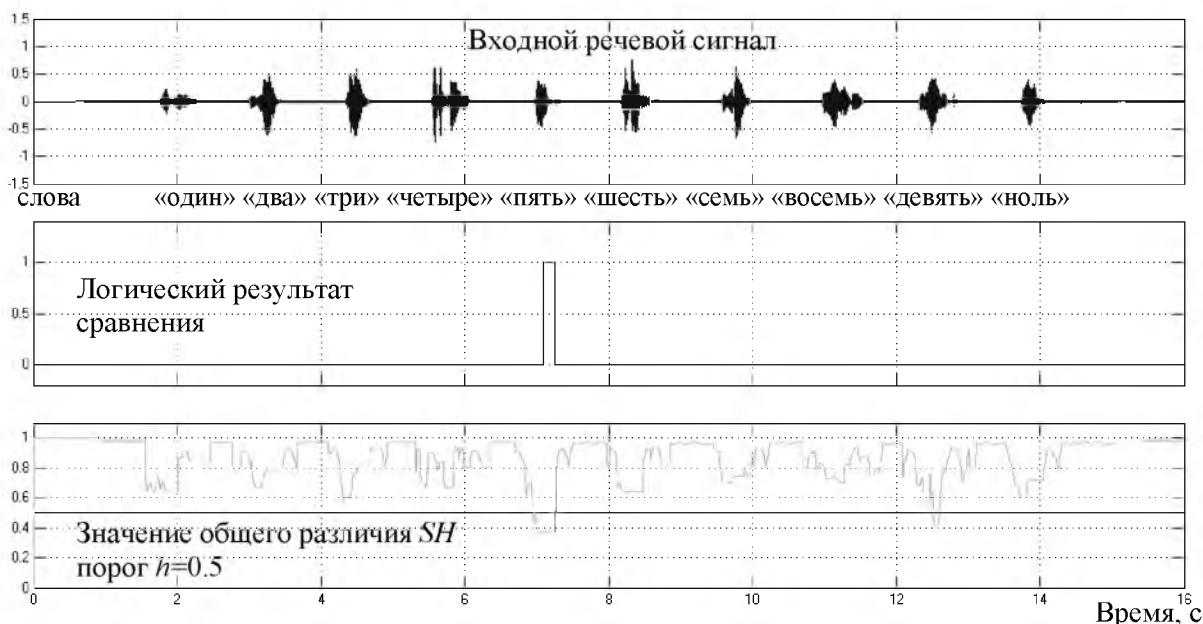


Рис. 15. Графики: входной сигнал (верхний), флаг обнаружения слова (средний) и величина отличия от шаблона для речевого сигнала из файла №5 вместе с пороговым значением 0.5 (нижний)  
 Fig. 15. Plots: input signal graphs (upper), word detection flag (average) and the difference values from the template for the speech signal from file № 5 together with a threshold value of 0.5 (lower)

Для всех файлов слово «пять» обнаружено правильно. Ошибки «пропуск цели» не было, ошибки «ложная тревога» не было. Для некоторых практических применений это приемлемо.

### Список литературы References

1. Материал из Википедии - свободной энциклопедии Распознавание речи. Available at: <http://ru.wikipedia.org/wiki/>. (in Russian)
2. Галунов В.И. 2007. Верификация и идентификация говорящего. С-Петербургский государственный университет.  
Galunov V.I. 2007. Verifikacija i identifikacija govorjashhego. S-Peterburgskij gosudarstvennyj universitet. (in Russian)
3. Мясников Л.Л., Мясникова.Н.Е. 1970. Автоматическое распознавание звуковых образов Л., «Энергия», 183.  
Mjasnikov L.L., Mjasnikova N.E. 1970. Avtomaticheskoe raspoznavanie zvukovyh obrazov L., «Jenergija», 183. (in Russian)
4. Хайкин Саймон, 2006. Нейронные сети: полный курс, 2-е издание. Пер. с англ. М. Издательский дом «Вильямс».1104.  
Hajkin Sajmon, 2006. Nejronnye seti: polnyj kurs, 2-e izdanie. Per. s angl. M. Izdatel'skij dom «Vil'jams».1104. (in Russian)
5. Белановский А.С. 1989. Основы биофизики в ветеринарии: Учеб. пособие для студентов высш. учеб. заведений. М. Агропромиздат. 271.  
Belanovskij A.S. 1989. Osnovy biofiziki v veterinarii: Ucheb. posobie dlja studentov vyssh. учеб. zavedenij. M. Agropromizdat. 271. (in Russian)
6. Бекеша Г., Розенблат В.А. 1963. Механические свойства уха. Психоакустическая лаборатория Гарвардского университета. Экспериментальная психология (том 2). редактор-составитель С.С. Стивенс: Пер. с англ. Под ред. П.К. Анохина, В.А. Артёмова. Издательство иностранной литературы, Москва.  
Bekeshi G., Rozenblat V.A. 1963. Mehanicheskie svojstva uha. Psihoakusticheskaja laboratorija Garvardskogo universiteta. Jeksperimental'naja psihologija (tom 2). redaktor-sostavitel' S.S. Stivens: Per. s angl. Pod red. P.K. Anohina, V.A. Artjomova. Izdatel'stvo inostranoj literatury, Moskva. (in Russian)
7. Алдошина И. Основы психоакустики. Подборка статей. Available at: <http://www.625-net.ru> (11 февраля 2010).



Aldoshina I. Osnovy psihoakustiki. Podboroka statej. Available at: <http://www.625-net.ru> (11 February 2010). (in Russian)

8. Вологдин Э.И. 2004. Слух и восприятие звука: Учеб. пособие. СТ«Факультет ДВО», СПб.

Vologdin Je.I. 2004. Sluh i vosprijatie zvuka: Ucheb. posobie. ST«Fakul'tet DVO», SPb. (in Russian)

9. Жилияков Е.Г., Белов С.П., Прохоренко Е.И. 2007. Методы обработки речевых данных в информационно-телекоммуникационных системах на основе частотных представлений. Белгород: Изд-во БелГУ. 136.

Zhiljakov E.G., Belov S.P., Prohorenko E.I. 2007. Metody obrabotki rechevyh dannyh v informacionno-telekommunikacionnyh sistemah na osnove chastotnyh predstavlenij. Belgorod: Izd-vo BelGU. 136. (in Russian)

10. Жилияков Е.Г. 2007. Вариационные методы анализа и построения функций по эмпирическим данным: моногр. Белгород: Изд-во БелГУ. 160.

Zhiljakov E.G. 2007. Variacionnye metody analiza i postroeniya funkcij po jempiricheskim dannym: monogr. Belgorod: Izd-vo BelGU. 160. (in Russian)

11. Черноморец А.А., Прохоренко Е.И., Голощапова В.А., 2009. О свойствах собственных векторов субполосных матриц. Научные ведомости БелГУ. Сер. История. Политология. Экономика. Информатика. 7(62): 122-128.

Chernomorec A.A. Prohorenko E.I., Goloshhapova A.A., 2009. About properties of subband matrices eigenvectors. Nauchnye vedomosti BelGU. Istoriya. Politologiya. Ekonomika. Informatika. [Belgorod State University Scientific Bulletin. History Political science Economics Information technologies]. 7(62): 122-128. (in Russian)

12. Солонина А.И., Улахович Д.А., Арбузов С.М., Соловьева Е.Б. 2006. Основы ЦОС: Учеб. пособие. Государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича, 747.

Solonina A.I., Ulahovich D.A., Arbuzov S.M., Solov'eva E.B. 2006. Osnovy COS: Ucheb. posobie. Gosudarstvennyj universitet telekommunikacij im. prof. M. A. Bonch-Bruevicha, 747. (in Russian)

13. Жилияков Е.Г., Черноморец А.А., Болгова Е.В., 2015. Оценивание производных дискретных функций. Научные ведомости БелГУ. Сер. Экономика. Информатика. 19(216): 96–100.

Zhilyakov E.G., Chernomorets A.A., Bolgova E.V., 2015. Discrete functions derivatives estimation. Nauchnye vedomosti BelGU. Ekonomika. Informatika. [Belgorod State University Scientific Bulletin. Economics Information technologies]. 19(216): 96–100. (in Russian)

14. Pavel Senin, 2008. Dynamic time warping algorithm review. Information and computer science department University of Hawaii at Manoa Honolulu. 23 .

15. H. Sakoe and S. Chiba, 1978. Dynamic programming algorithm optimization for spoken word recognition, Acoustics, Speech and Signal Processing, IEEE Transactions on, 1(26): 43-49. Available at: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=1163055](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1163055)

16. Скворцов В.А. 2002. Примеры метрических пространств. М. МЦНМО. 24.

Skvorcov V.A. 2002. Primery metricheskih prostranstv. M. MCNMO. 24. (in Russian)

17. Семкин Б.И., Горшков М.В. 2008. Аксиоматическое введение мер сходства, различия, совместимости и зависимости для компонентов биоразнообразия. Вестник ТГУЭ. 4.

Semkin B.I., Gorshkov M.V. 2008. Aksiomaticeskoe vvedenie mer shodstva, razlichija, sovmestimosti i zavisimosti dlja komponentov bioraznoobrazija. Vestnik TGUJe. 4. (in Russian)