

УДК 621.397

DOI 10.18413/2687-0932-2020-47-2-441-451

**АВТОМАТИЧЕСКАЯ СЕГМЕНТАЦИЯ РЕЧИ ПУТЕМ АНАЛИЗА  
ИНФОРМАЦИОННОЙ ОДНОРОДНОСТИ****AUTOMATIC SEGMENTATION OF SPEECH BY ANALYSIS  
OF THE INFORMATIONAL HOMOGENEITY****С.В. Уманец<sup>1</sup>, А.В. Болдышев<sup>1</sup>, П.Г. Лихолоб<sup>2</sup>  
S.V. Umanets<sup>1</sup>, A.V. Boldyshev<sup>1</sup>, P.G. Likholob<sup>2</sup>**<sup>1</sup> Белгородский филиал ПАО «Ростелеком»,  
Россия, 308000, Белгород, п. Б.-Хмельницкого, 81<sup>2</sup> Белгородский государственный национальный исследовательский университет,  
Россия, 308015, Белгород, ул. Победы, 85<sup>1</sup> Belgorod filial PJSC "Rostelekom", 81 B.-Khmelnitskogo Av, Belgorod, 308000, Russia<sup>2</sup> Belgorod State National Research University, 85 Pobeda St, Belgorod, 308015, Russia

e-mail: 286084@bsu.edu.ru; boldyshev@bsu.edu.ru, Likholob@bsu.edu.ru

**Аннотация**

В работе рассмотрен алгоритм анализа цифровых сигналов. Назначение алгоритма – автоматическая сегментация речевого сигнала, что означает разбиение исходного сигнала на фрагменты с устойчивыми информационными признаками. Принцип работы алгоритма основан на анализе информационной однородности. Разработана информационная модель однородности устной речи. Проведен анализ информационной однородности на примере звукозаписи русской фразы. Разработан алгоритм поиска границ смены фонем. В работе также представлены результаты вычислительных экспериментов для различных комбинаций параметров алгоритма. Сделан вывод о работоспособности алгоритма на основе проверки по критериям точность (до 0.84) и полнота (до 0.85). Приведены рекомендации для настройки алгоритма на основе вычислительных экспериментов с речевой базой TIMIT.

**Abstract**

In this paper, an algorithm for digital signal processing is considered. The purpose of the algorithm is automatic segmentation of the speech signal, that is, splitting the original signal into segments generated by action in different conditions. The principle of operation of the algorithm is based on the analysis of information homogeneity. At the beginning of the algorithm, an information model is compiled, then an analysis of information homogeneity is performed and a search for the boundaries of sound change is performed. The application of non-linearity and moving average for confident decision-making about the presence of a boundary between different sounds is considered. The paper also presents the results of computational experiments for various combinations of algorithm parameters. Numerical evaluation of the algorithm was carried out on the material from the database of marked speech fragments of the American Agency for advanced defense research projects DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. The conclusion about the algorithm's performance is made.

**Ключевые слова:** доли энергии, энтропия, взаимная информация, однородность.**Keywords:** fraction of energy, entropy, mutual information, homogeneity.

Для построения систем голосового управления появляется потребность распознавать речь – анализируя речевой сигнал определить элемент устной речи, под воздействием которого сформировался анализируемый отрезок речевого сигнала. Речевой сигнал – это результат регистрации электромагнитных колебаний на выходе микрофона от воздействия

акустических колебаний. Распознавание проводится с помощью анализа набора признаков. [Загоруйко, 1972]. Признаки можно получить, объединяя в группы однородные элементы. Поэтому один из этапов автоматического анализа речевого сигнала – это сегментация – разбиение исходного сигнала на отрезки, порождённые воздействием в разных условиях. Речевой сигнал является нестационарным и его характеристики значительно меняются от отрезка к отрезку. При этом в границах одного элемента речевого сигнала информационная однородность будет больше, чем при переходе между элементами.

Ориентируясь на модели строения слухового аппарата человека [Вологдин, 2004] необходимо анализировать энергию звука в зависимости от частотного интервала. В данной работе анализу будут подвергаться доли энергии отрезков сигнала [Болдышев и др., 2011] путём группировки их в последовательные фрагменты.

### Составление информационной модели

В работе использовалась следующая модель.

Для обработки использовалась звукозапись устной речи

$$\vec{s} = (s_1, s_2, \dots, s_U). \quad (1)$$

Звукозапись разбивалась на отрезки анализа

$$\vec{x}_n = (s_{n+1}, s_{n+2}, \dots, s_{n+N}), \quad (2)$$

представляющие собой вектора размерностью  $N$ , составленные из последовательных отсчётов оцифрованной звукозаписи, взятых начиная с момента времени  $n$ . Момент формирования отрезка анализа  $t$  это

$$t = n / F_s, \quad (3)$$

где  $F_s$  – это частота дискретизации при звукозаписи. Размерность отрезка анализа задаётся параметром модели – временем сбора отрезка анализа  $T_w$ :

$$N = T_w \cdot F_s, \quad (4)$$

это более адекватный параметр, чем просто числовое значение для размерности, потому что звукозапись может быть оцифрована при разных частотах дискретизации.

Для отрезка анализа вычислялось распределение энергии по частотам (здесь  $T$  – знак транспонирования)

$$W_m = \vec{x}_n \cdot \mathbf{A}_r \cdot \vec{x}_n^T. \quad (5)$$

Квадратичная форма содержит субполосную матрицу с элементами в виде разницы синусов [Жилияков, 2007]:

$$a_r(i, k) = \begin{cases} \frac{\sin(v_r(i-k)) - \sin(v_{r-1}(i-k))}{\pi(i-k)}, & \text{при } i \neq k \\ (v_r - v_{r-1}) / \pi, & \text{при } i = k \end{cases}, \quad (6)$$

где  $v_r$  – верхняя граница  $r$ -го частотного интервала; индексы  $i, k$  меняются от 1 до  $N$ ,  $r$  – это номер частотного интервала,  $r = 1, 2, \dots, R$  всего интервалов  $R$ , количество частотных интервалов является параметром модели.

Разбиение на интервалы удовлетворяло условию:

$$\Omega_r = [-v_r, -v_{r-1}) \cup [v_{r-1}, v_r) \quad v_0 = 0, \quad \Omega_R = \bigcup_{r=1}^R \Omega_r = [-\pi, \pi). \quad (7)$$

Квадратичная форма (5) соответствует вычислению части энергии отрезка сигнала, содержащегося в  $r$ -м частотном интервале

$$S_r(x) = \frac{1}{2\pi} \int_{\omega \in \Omega_r} |X(\omega)|^2 d\omega \quad (8)$$

где

$$X(\omega) = \sum_{n=1}^N x_n e^{-j(n-1)\omega}, j = \sqrt{-1} \tag{9}$$

это трансформанта Фурье [Ильин и др., 1985] с областью определения

$$-\pi \leq \omega \leq \pi. \tag{10}$$

Имеет место и обратное преобразование

$$x_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) e^{j(n-1)\omega} d\omega \tag{11}$$

Отсюда можно получить равенство Парсеваля для энергии сигнала в отрезке анализа

$$\|\vec{x}\|^2 = \sum_{n=1}^N x_n^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\omega)|^2 d\omega = \sum_{r=1}^R S_r(x). \tag{12}$$

В данной работе отрезки для анализа выбирались из звукозаписи (1) через *Nstep* отсчётов

$$Nstep = (1 - Cw) \cdot N, \tag{13}$$

где *Cw* – это доля перекрытия отрезков анализа, является параметром модели. На рисунке 1 представлен пример отбора отрезков анализа из исходной звукозаписи при значении доли перекрытия  $Cw = \frac{3}{5}$ .

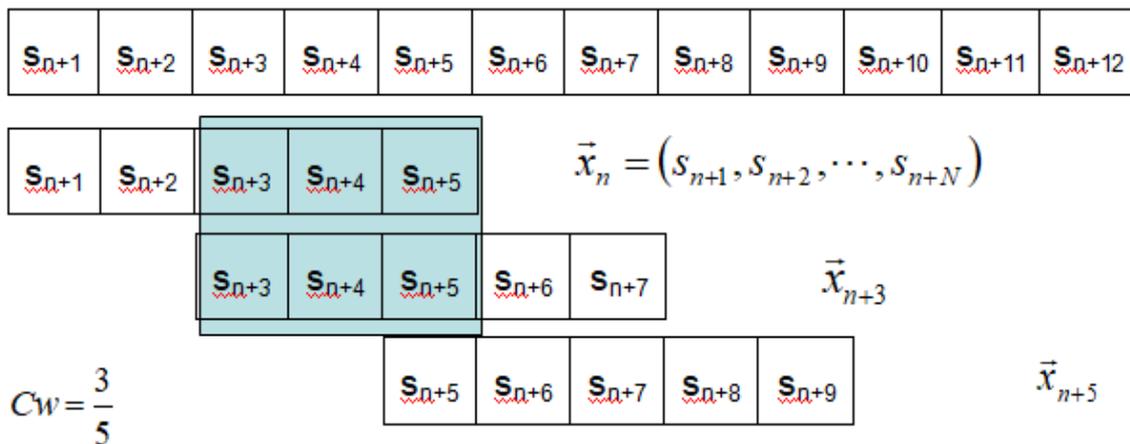


Рис. 1. Выбор отрезков анализа с перекрытием  
Fig. 1. Select analyze intervals with intercross

Результаты вычислений распределения энергий по частотам группировались в таблицу *G*, определяемую как фрагмент анализа

$$G = \begin{pmatrix} W_{11} & W_{12} & \dots & W_{1r} & \dots & W_{1R} \\ W_{21} & W_{22} & \dots & W_{2r} & \dots & W_{2R} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ W_{m1} & W_{m2} & \dots & W_{mr} & \dots & W_{mR} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ W_{M1} & W_{M2} & \dots & W_{Mr} & \dots & W_{MR} \end{pmatrix}. \tag{14}$$

Индекс *r* – это номер частотного интервала, а индекс *m* – соответствует времени выборки отрезка анализа, относительно первого отрезка при формировании фрагмента:

$$t/F_s = n + (m - 1) \cdot Nstep, \tag{15}$$

то есть *m* – это индекс для окна анализа, взятого через время *t* относительно выбора первого отрезка во фрагменте.

Фрагмент анализа собирался за время  $Tg$  – это ещё один параметр модели. Вертикальный размер  $M$  для матрицы фрагмента вычислялся как

$$M = \left\lfloor \frac{Tg \cdot Fs}{Nstep} \right\rfloor + 1, \quad (16)$$

где полуквадратные скобки обозначают действие взятия целой части. Время сбора фрагмента обязательно было больше времени сбора отрезка анализа

$$Tg > Tw. \quad (17)$$

Расстояние между фрагментами анализа составляло величину

$$Nstep\_g = M \cdot (1 - Cg), \quad (18)$$

т. е. в следующем фрагменте анализа первый отрезок выбирается через  $Nstep\_g$  отрезков относительно первого отрезка текущего фрагмента, где шаг сдвига (18) вычислялся с помощью доли перекрытия фрагментов  $Cg$ . Доля перекрытия фрагментов анализа является параметром модели. На рисунке 2 представлен пример формирования фрагментов анализа из отрезков анализа при значении доли перекрытия  $Cg = \frac{2}{4}$ .

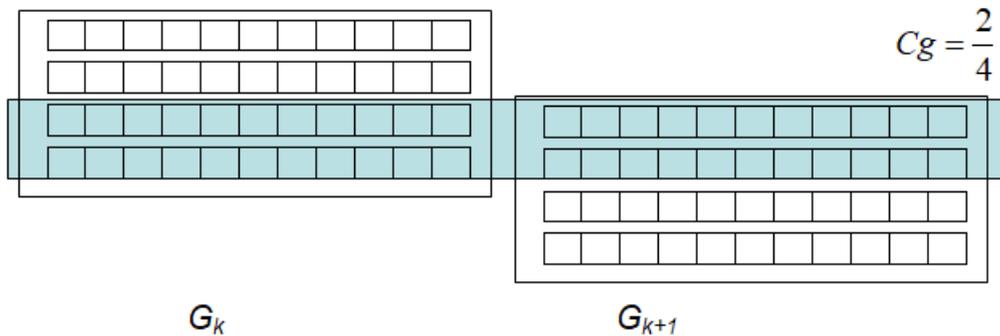


Рис. 2. Формирование отрезков анализа с перекрытием  
Fig. 2. Forming analyze fragments with intercross

### Анализ информационной однородности речевого сигнала

Следующим этапом в работе был анализ информационной однородности речевого сигнала.

Для фрагмента анализа  $G$  (14) определялись:

доля энергии в  $r$ -ом частотном интервале (5) в момент времени  $m$  (15):

$$P_{mr} = \frac{W_{mr}}{W_{\Sigma}}, \quad (19)$$

где величина

$$W_{\Sigma} = \sum_{m=1}^M \sum_{r=1}^R W_{mr} \quad (20)$$

является суммарной энергией во фрагменте анализа;  
и краевые доли:

доля энергии в  $r$ -ом частотном интервала по отношению ко всей энергии фрагмента:

$$P_r = \frac{\sum_{m=1}^M W_{mr}}{W_{\Sigma}} \quad (21)$$

и доля энергии в  $m$ -ый момент времени по отношению ко всей энергии во фрагменте:

$$P_m = \frac{\sum_{r=1}^R W_{mr}}{W_{\Sigma}}. \quad (22)$$

Если рассматривать весь фрагмент как пространство элементарных событий, а за событие принять наличие энергии, то доли энергии можно рассматривать как вероятности.

Если бы события «доля энергии в частотном интервале» и «доля энергии в отсчёте» были полностью зависимые, то перемножение вектора-столбца  $P_m$  на вектор-строку  $P_r$  было бы равно матрице  $P_{mr}$ . Вычислительные эксперименты обнаруживают отсутствие равенства.

$$\begin{pmatrix} P_1 \\ \vdots \\ P_m \\ \vdots \\ P_M \end{pmatrix} \times (P_1 \quad \dots \quad P_r \quad \dots \quad P_R) \neq \begin{pmatrix} P_{11} & \dots & P_{1r} & \dots & P_{1R} \\ \dots & \dots & \dots & \dots & \dots \\ P_{m1} & \dots & P_{mr} & \dots & P_{mR} \\ \dots & \dots & \dots & \dots & \dots \\ P_{M1} & \dots & P_{Mr} & \dots & P_{MR} \end{pmatrix}. \quad (23)$$

Поэтому должна быть взаимная информация между изменением энергии по частоте и изменением энергии по времени.

Взаимная информация [Хайкин, 2006] вычислялась как

$$I = H_m + H_n - H_{mn}, \quad (24)$$

где  $H$  – это энтропия [Шеннон, 1948]:

для моментов времени

$$H_{mr} = - \sum_{m=1}^M \sum_{r=1}^R P_{mr} \log(P_{mr}), \quad (25)$$

для частотных интервалов

$$H_{mr} = - \sum_{m=1}^M \sum_{r=1}^R P_{mr} \log(P_{mr}), \quad (26)$$

совместная

$$H_{mr} = - \sum_{m=1}^M \sum_{r=1}^R P_{mr} \log(P_{mr}). \quad (27)$$

Взаимная информация определяется как разница суммы краевых энтропий и совместной. Взаимная информация должна быть положительной величиной. Вычисляя величину взаимной информации от фрагмента к фрагменту можно построить график и по нему проанализировать информационную однородность речевого сигнала.

На рис. 3 и рис. 4 представлены графики речевого сигнала для слова «ВЛЕВО» и поведения взаимной информации.

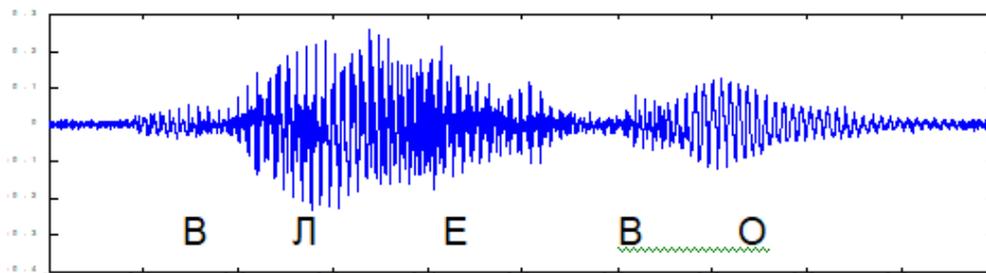


Рис. 3. График речевого сигнала: слово «ВЛЕВО»

Fig. 3. Plot of speech signal: word «ВЛЕВО»

Возникла гипотеза, что между пиками на графике поведения взаимной информации звучит фонема. Такой анализ позволил бы определять границы фонем, составляющих речевой сигнал. Однако наблюдается множество нежелательных пиков локальных максимумов.

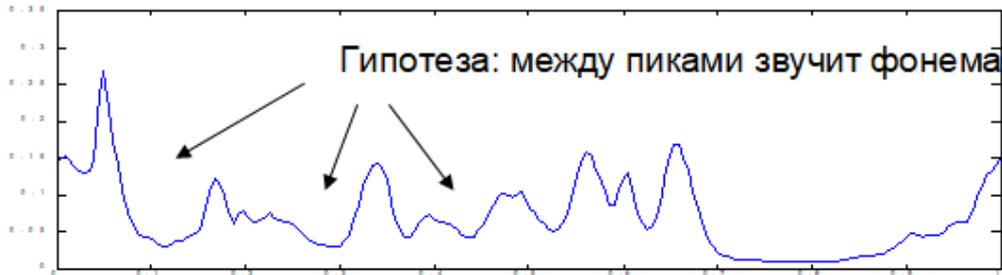


Рис. 4. График поведения взаимной информации  
Fig. 4. Plot of behavior of mutual information

### Поиск границ смены звуков речи

Чтобы обнаруживать локальный максимум в работе использовалось условие:

$$\begin{aligned} I_k &< I_s, \\ k &\in [s-d, s+d], \\ k &\neq s \end{aligned} \quad (28)$$

где  $s$  – индекс фрагмента, дающий локальный максимум на графике взаимной информации,  $d$  – количество соседних фрагментов, участвующих в сравнении.

Значение  $d$  вычислялось по параметрам модели (4), (13), (17), (18)

$$d = 2 \left\lfloor \frac{N_a}{2 \cdot G_d} \right\rfloor + 1, \quad (29)$$

где  $G_d$  – расстояние между соседними фрагментами анализа, измеряемое в отсчётах входного речевого сигнала

$$G_d = Nstep \cdot Nstep\_g, \quad (30)$$

$N_a$  – количество отсчётов входного речевого сигнала в интервале времени для поиска локального максимума

$$N_a = \tau_a \cdot Fs, \quad (31)$$

интервал времени  $\tau_a = 40ms$  выбран по материалам результатов психоакустических исследований [Цвикер 1971; Алдошина, 2010].

Для учёта порога слышимости была введена нелинейность «зона нечувствительности». На рис. 5 представлены графики нелинейности в виде кусочно-непрерывной функции и в виде гладкой аппроксимации.

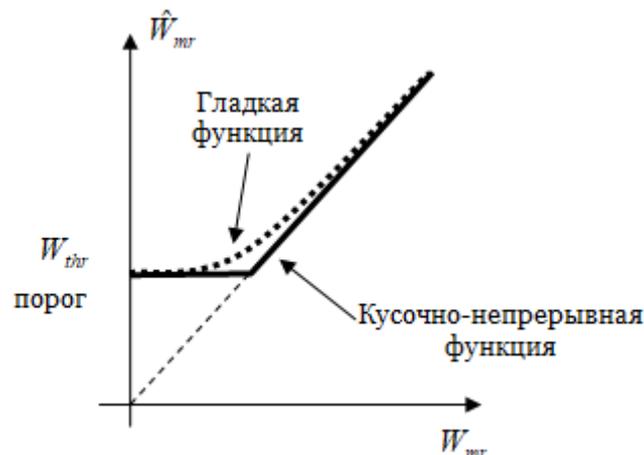


Рис. 5. График нелинейности  
Fig. 5. Plot of nonlinearity

Кусочно-непрерывная функция, задающая нелинейность:

$$\hat{W}_{mr} = \begin{cases} W_{mr}, & W_{mr} > W_{thr} \\ W_{thr}, & W_{mr} \leq W_{thr} \end{cases}, \tag{32}$$

гладкая аппроксимация задавалась функцией:

$$\hat{W}_{mr} = W_{thr} + \frac{W_{mr}^2}{W_{mr} + W_{thr}}. \tag{33}$$

В формулах (32) и (33) пороговое значение  $W_{thr}$  вычислялось для каждого фрагмента анализа индивидуально как среднеарифметическое значение энергий [Жиляков и др., 2011] по трём смежным фрагментам, но не меньше значения эквивалентного -50 дБ как абсолютного порога слышимости.

Дополнительно, для устранения ложных пиков, проводилась процедура сглаживания кривой взаимной информации

$$\hat{I}_s = \frac{1}{2d+1} \sum_{k=(s-d)}^{k=(s+d)} I_k. \tag{34}$$

Значение для сглаженной кривой взаимной информации для индекса  $s$  вычислялось как среднее арифметическое между значениями взаимных информаций в соседствующих фрагментах. Отрезок со значениями индексов для соседствующих фрагментов брался такой же, как и для определения локального максимума.

На рис. 6 изображён график речевого сигнала с отметками алгоритма, на рис. 7 представлена кривая взаимной информации после учёта нелинейности и сглаживания. Кружочками обозначены локальные максимумы.

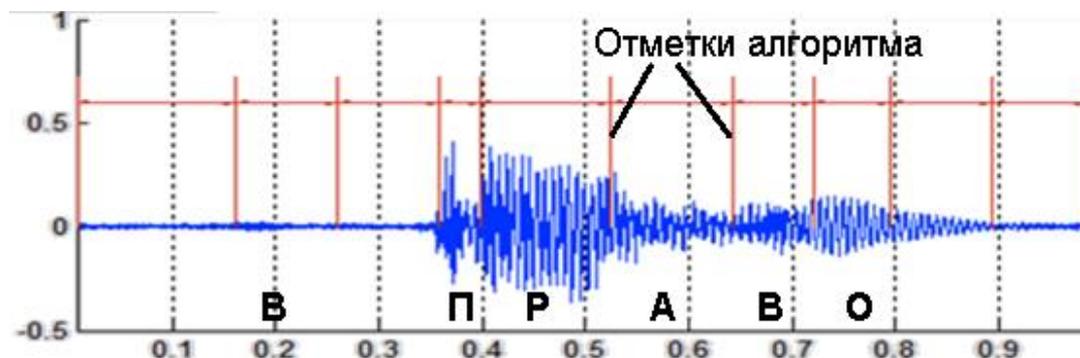


Рис. 6. График речевого сигнала: слово «ВПРАВО»  
Fig. 6. Plot of speech signal: word «ВПРАВО»

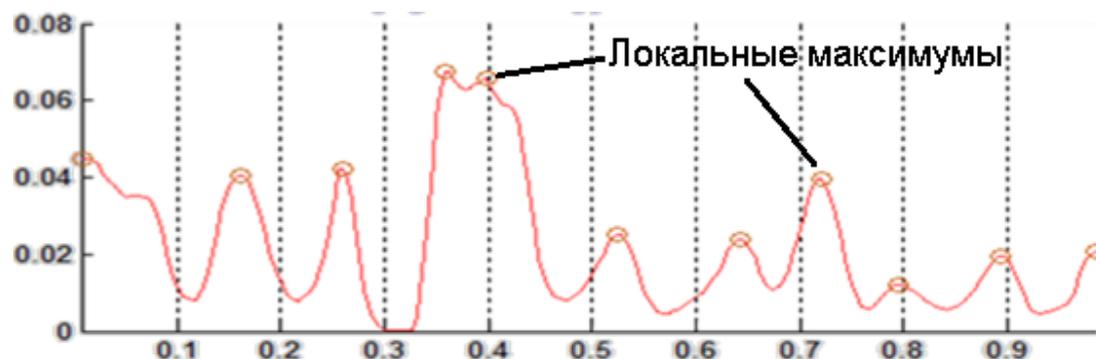


Рис. 7. График поведения взаимной информации после учета нелинейности и сглаживания  
Fig. 7. Plot of behavior of mutual information after use nonlinearity and smoothing

### Вычислительный эксперимент

Для проверки и численной оценки адекватности работы алгоритма надо подсчитать, сколько раз была верной постановка границы. Верной считалась постановка, если расстояние между отметкой ручного разбиения и отметкой автоматического разбиения оказывалось меньше величины времени допустимого расхождения:

$$C_0 = \sum_{i,j} \begin{cases} 1, \min_{i,j} (|\alpha_i - \beta_j|) < \theta \\ 0 \end{cases}, \quad (35)$$

где  $C_0$  – количество верных постановок границ,  $\beta \in \{B\}$  – множество отметок ручной разметки,  $\alpha \in \{A\}$  – множество отметок автоматической разметки,  $\theta = 20\text{мс}$  – величина времени допустимого расхождения. Выбрана по материалам результатов психоакустических исследований [Цвикер и др., 1971; Алдошина, 2010].

Кроме количества верных оценок, ещё необходимо подсчитать сколько раз алгоритм не поставил отметку наличия границы, когда такая отметка присутствует при ручной разметке, это ошибки первого рода «пропуск цели».

$$C_I = [B] - C_0, \quad (36)$$

где  $C_I$  – количество ошибок первого рода,  $[B]$  – количество элементов в множестве  $\{B\}$ .

Также необходимо подсчитать сколько произошло ошибочных постановок границ, это ошибки второго рода «ложная тревога».

$$C_{II} = [A] - C_0, \quad (37)$$

где  $C_{II}$  – количество ошибок второго рода,  $[A]$  – количество элементов в множестве  $\{A\}$ .

Оценка алгоритма проводилась по критериям точности  $\phi$  и полноты  $\chi$

$$\phi = \frac{C_0}{C_0 + C_I}, \quad (38)$$

$$\chi = \frac{C_0}{C_0 + C_{II}}. \quad (39)$$

Численная оценка работы алгоритма проводилась на материале из базы размеченных речевых фрагментов американского агентства передовых оборонных исследовательских проектов DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. В базе содержится много звукозаписей. Для эксперимента были взяты 100 звукозаписей. С целью выяснения влияния параметров модели на точность вычислений, расчёты выполнялись для одной звукозаписи при разных комбинациях значений параметров.

На рис. 8 представлен рабочий момент алгоритма. Синяя линия – это сигнал из звукозаписи. Зелёные линии – отметки, сделанные вручную, горизонтальные концы – интервалы допуска  $\pm 20$  миллисекунд. Розовая линия внизу – график взаимной информации, розовые вертикальные линии – отметки автоматической сегментации. Есть эпизоды верной сегментации и ошибочной.

По результатам численных экспериментов посчитаны точность и полнота (таблица 1).

В табл. 2 приведены результаты оценки параметров алгоритма, при которых достигаются максимальные значения точности или полноты. Максимальная точность была при наборе параметров (табл. 2, строка 1), полнота при этом составила  $\chi = 0,689$ . Максимальная полнота была при наборе параметров (таблица 2, строка 2), точность при этом составила  $\phi = 0,595$ , т. е. была минимальной. Максимальное произведение точности и полноты было при наборе параметров (табл. 2, строка 3), полнота при этом составила  $\chi = 0,751$ , а точность оказалась  $\phi = 0,82$ .

Выводы: возможно проводить сегментацию речевого сигнала, используя анализ информационной однородности распределения энергий сигнала по частотным интервалам. Параметры алгоритма оказывают влияние на результативность.

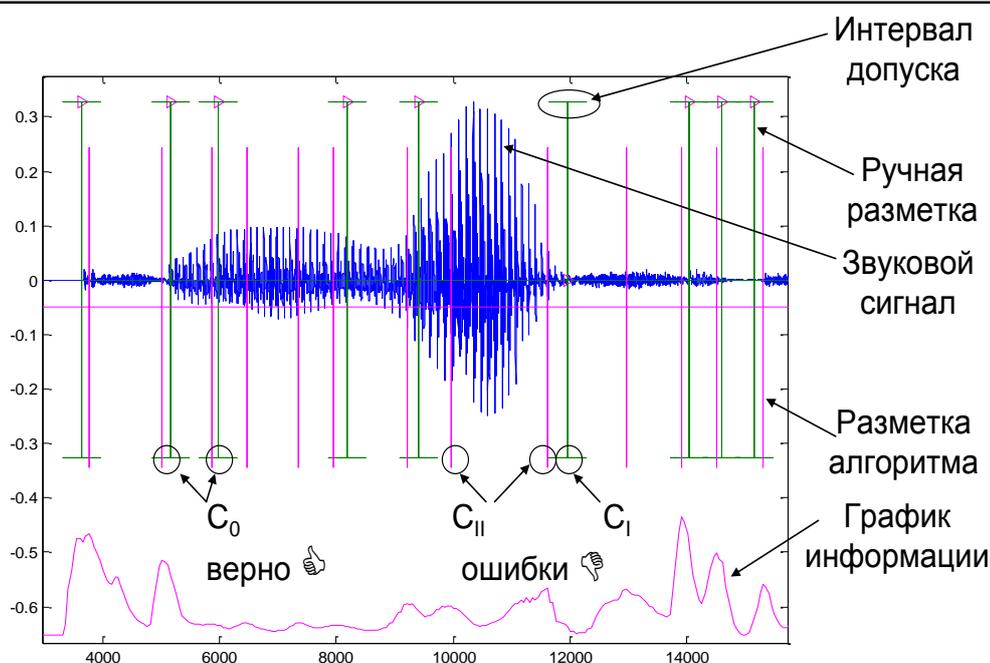


Рис. 8. Пример работы алгоритма  
Fig. 8. Example of working algorithm

Таблица 1  
Table 1

Значения точности и полноты работы алгоритма  
Value of precision and fullness working algorithm

	Минимальное значение	Максимальное значение
$\phi$	0,607	0,841
$\chi$	0,595	0,85

Таблица 2  
Table 2

Значения параметров при максимальной точности  
Value of parameters for maximal precision

$\phi$	$\chi$	$T_w$ , мс	$C_w$	$T_g$ , мс	$C_g$
0,841	0,689	8	0,75	60	0,95
0,595	0,85	10	0,65	45	0,7
0,82	0,751	13	0,9	50	0,95

### Список литературы

1. Алдошина И. 2010. Основы психоакустики. Подборка статей. URL: <http://www.625-net.ru> (дата обращения: 11 февраля 2010).
2. Белов С.П., Белов А.С. 2008. О различиях частотных свойств информационных и неинформационных звуковых сигналов речевого диапазона. Научные ведомости БелГУ Сер. Информатика, 7 (38): 214–221.
3. Вологдин Э.И. 2004. Слух и восприятие звука: Учеб. пособие. СПб. СТ «Факультет ДВО», 52.
4. Жилияков Е. Г., Прохоренко Е. И., Болдышев А. В. и др. 2011. Сегментация речевых сигналов на основе анализа распределения энергии по частотным интервалам. Научные Ведомости БелГУ. Сер. Экономика. Информатика, 7 (102): 187–196.

5. Жилияков Е. Г., Трубицына Д. И., Прохоренко Е. И., Болдышев А. В. 2019. Об использовании субполосного анализа и синтеза сигналов в области определения косинус-преобразования при решении задач сжатия речевых сигналов. Научные Ведомости БелГУ. Сер. Экономика. Информатика, 4 (46): 700–709.
6. Жилияков Е.Г. 2007. Вариационные методы анализа и построения функций по эмпирическим данным: моногр. Белгород: Изд-во БелГУ. 160.
7. 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.
8. Загоруйко Н.Г. 1972. Методы распознавания и их применение. М. Сов. Радио: 135–147.
9. Ильин В.А., Садовничий В.А., Сендов Бл.Х. 1985. Математический анализ. Продолжение курса. М. Изд-во МГУ, 358.
10. Фирсова А.А. 2013. Разработка и исследование субполосных методов и алгоритмов сегментации речевых сигналов. Автореф. дис. ... канд. техн. наук. Белгород, 22.
11. Хайкин С. 2006. Нейронные сети: полный курс, 2-е издание. Пер. с англ. М. Издательский дом «Вильямс», 1104.
12. Цвикер Э., Фельдкеллер Р. 1971. Ухо как приемник информации. Пер. с нем. М. Связь, 64.
13. Шелухин О.И., Лукьянцев Н.Ф. 2000. Цифровая обработка и передача речи. Москва. Радио и связь, 456.
14. DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. URL: <https://www.kaggle.com/mfekadu/darpa-timit-acousticphonetic-continuous-speech> (дата обращения: 11 February 2020).
15. Shannon C.E. 1948. A mathematical theory of communication. Bell System Technical Journal, 27: 379–423, 623–656.

## References

1. Aldoshina I. Osnovy psihoakustiki. Podborka statej [Fundamentals of psychoacoustics. A selection of articles]. Available at: <http://www.625-net.ru> (accessed: 11 February 2020).
2. Belov S.P., Belov A.S. 2008. O razlichiyah chastotnyh svojstv informacionnyh i neinformacionnyh zvukovyh signalov rechevogo diapazona [On the differences in the frequency properties of information and non-information audio signals of the speech range]. Nauchnyje Vedomosti BelGU. Ser. Ekonomika. Informatika. 7 (38): 214–221.
3. Vologdin Je.I. 2004. Sluh i vospriyatie zvuka [Hearing and sound perception]: Ucheb. posobie. ST«Fakul'tet DVO». Saint Petersburg, 52.
4. Zhiljakov E.G., Prokhorenko E.I., Boldyshev A.V. et al. 2011. Segmentatsija rechevyh signalov na osnove analiza raspredelenija energii po chastotnym intervalam [Segmentation of speech signals based on analysis of energy distribution over frequency intervals]. Nauchnyje Vedomosti BelGU. Ser. Ekonomika. Informatika. 7 (102): 187–196.
5. Zhiljakov E.G., Trubitsyna D.I., Prokhorenko E.I., Boldyshev A.V. 2019. Ob ispolzovanii subpolosnogo analiza i sinteza signalov v oblasti opredeleniya cosinus-preobrazovaniya pri reshenii zadach szhatiya rechevikh signalov [On the use of subband analysis and synthesis of signals in the field of determining the cosine transform in solving problems of compression of speech signals]. Nauchnyje Vedomosti BelGU. Ser. Ekonomika. Informatika. 4 (46): 700–709.
6. Zhiljakov E.G. 2007. Variacionnye metody analiza i postroenija funkcion po jempiricheskim dannym [Variational methods of analysis and construction of functions based on empirical data]: monogr. Belgorod. Izd-vo BelGU, 160.
7. Zhiljakov E.G., Belov S.P., Prohorenko E.I. 2007. Metody obrabotki rechevyh dannyh v informacionno-telekommunikacionnyh sistemah na osnove chastotnyh predstavlenij [Methods of processing voice data in information and telecommunication systems based on frequency representations]. Belgorod. Izd-vo BelGU, 136.
8. Zagorujko N.G. 1972. Metody raspoznavaniya i ih primenenije [Recognition Methods and Their Application]. М. Sov. radio: 135–147.
9. Il'in V.A., Sadovnichij V.A. Sendov Bl.Kh. 1985. Matematicheskij analiz. Prodolgenije kursa [Mathematical analysis. Continuation of the course]. Moskov. Izd-vo MGU, 358.
10. Firsova A.A. 2013. Razrabotka i issledovanie subpolosnyh metodov i algoritmov segmentacii rechevyh signalov [Development and research of subband methods and algorithms of segmentation of speech signals]. Abstract. dis. ... cand. of technical sciences. Belgorod, 22.

11. Hajkin Sajmon, 2006. Neironnye seti: polnyj kurs [Neural networks: full course], 2-e izdanie. Per. s angl. Moskov. Izdatel'skij dom «Vil'jams», 1104.
12. Cviker E., Feldkeller R. 1971. Uho kak prijemnik informacii [Ear as a receiver of information]. Per. s nemeckogo. Moskov, Svyaz, 64.
13. Shelukhin O.I., Lukyantsev N.F. 2000. Cifrovaya obrabotka i peredacha rechi [Digital speech processing and transmission]. Moscow. Publ. Radio i svyaz, 456.
14. DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. Available at: <https://www.kaggle.com/mfekadu/darpa-timit-acousticphonetic-continuous-speech> (accessed: 11 February 2020).
15. Shannon C.E. 1948. A mathematical theory of communication. Bell System Technical Journal, 27: 379–423, 623–656.

**Ссылка для цитирования статьи  
For citation**

Уманец С.В., Болдышев А.В., Лихолоб П.Г. 2020. Автоматическая сегментация речи путем анализа информационной однородности. Экономика. Информатика. 47 (2): 441–451. DOI: 10.18413/2687-0932-2020-47-2-441-451.

Umanets S.V., Boldyshev A.V., Likholob P.G. 2020. Automatic segmentation of speech by analysis of the informational homogeneity. Economics. Information technologies. 47 (2): 441–451 (in Russian). DOI: 10.18413/2687-0932-2020-47-2-441-451.