

УДК 534.4:621.391

О. Н. Карпов, К. Н. Глушак, О. И. Лучинкина

Днепропетровский национальный университет имени Олеся Гончара

ГРАФО-АНАЛИТИЧЕСКИЙ МЕТОД ОПРЕДЕЛЕНИЯ ГРАНИЦ СЕГМЕНТАЦИИ РЕЧЕВЫХ СИГНАЛОВ

Наведено алгоритм багатоступової сегментації мовленевих сигналів, заснованої на визначенні початку–кінця слова, меж пауз, переходів: шум–тон, тон–шум, тон–тон; графічне удосконалення меж за ізолініями інтенсивності сигналу.

Ключевые слова: *сегментация речи, низкочастотная фильтрация, спектр.*

Приведен алгоритм многоэтапной сегментации речевых сигналов, основанной на определении начала–конца слова, границ пауз, переходов: шумный–тональный, тональный–шумный, тональный–тональный; графическое уточнение границ по изолиниям интенсивности сигнала.

Ключевые слова: *сегментация речи, низкочастотная фильтрация, спектр.*

This article is about multistep segmentation of speech signals. This method is based on determination of the beginning and the end of the word, boundaries of the signal, transitions from tone to noise, noise–tone and tone–tone.

Key words: *speech segmentation, low-pass filtering, spectr, spectrum.*

Постановка проблеми. Известны методы сегментации графических изображений [1]: а) на основе анализа точек или пикселей; б) на основе анализа контуров; в) на основе анализа областей; г) на основе моделирования. Известны методы сегментации речевых сигналов, например, многоэтапное определение границ: а) определение пауз, б) определение чередования тональных–шумных, шумных–тональных звуков [2], в) метод верификации [3], заключающийся в построении функции сегментации.

Изложение основного материала. Определение границ пауз и латентных периодов для исходного речевого сигнала $s_0(t)$ (рис.1) слова «атлас» решается следующим образом. Находится спектрально-временное представление $S_0(\omega_k, T_l)$ (рис.2). Для каждого спектра $S_0(\omega_k)$ интервала T_l находится максимальное значение спектра $S_{0\max}(T_l)$ и затем находится глобальный максимум для всего

спектрально-временного представлення $S_{g \max}$. Затем нормируется масив $S_{0 \max}(T_l)$ максимальных значений спектров относительно $S_{g \max}$, в итоге образуется нормированный массив $S_{0 \max}^n(T_l) \leq 1$.

В зависимости от зашумлённости исходного сигнала $s_0(t)$ и соответственно спектрально-временного представлення $S_0(\omega_k, T_l)$ выбирается порог определения границ пауз и латентных периодов σ . Для дальнейшей обработки формируется сигнал $s(t)$ в границах, образованных при

$$S_{0 \max}^n(T_l) \geq \sigma \tag{1}$$

в начале, конце и внутри слова или фразы.

Для сигнала, произнесённого в обычном, слегка зашумлённом помещении, порог $\sigma = 0,02$. Латентные периоды и паузы надёжно выделяются при зашумлённости до 10%. Однако, может быть удалена тональная пауза перед «б, д, г- твёрдое» в начале слова, например, «два».

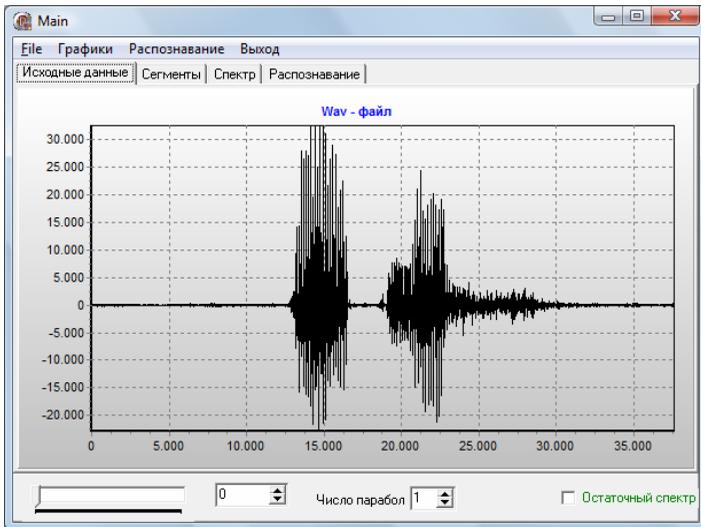


Рис.1. Исходный сигнал слова «атлас»

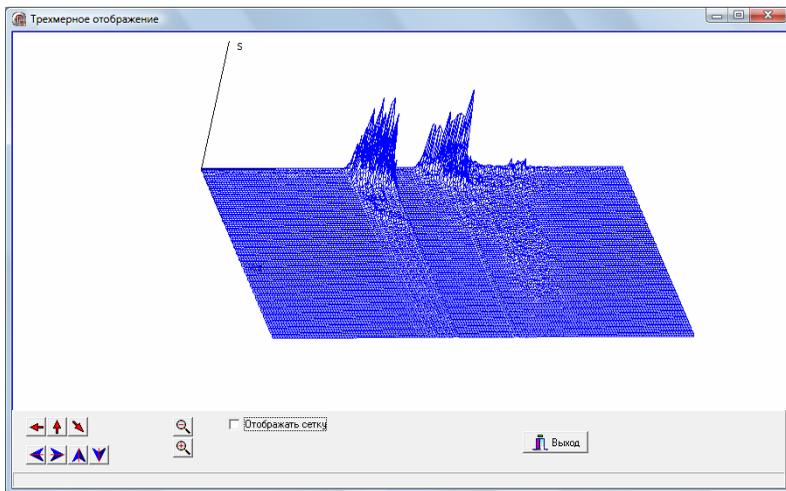


Рис.2. Спектрально-временное представление слова «атлас»

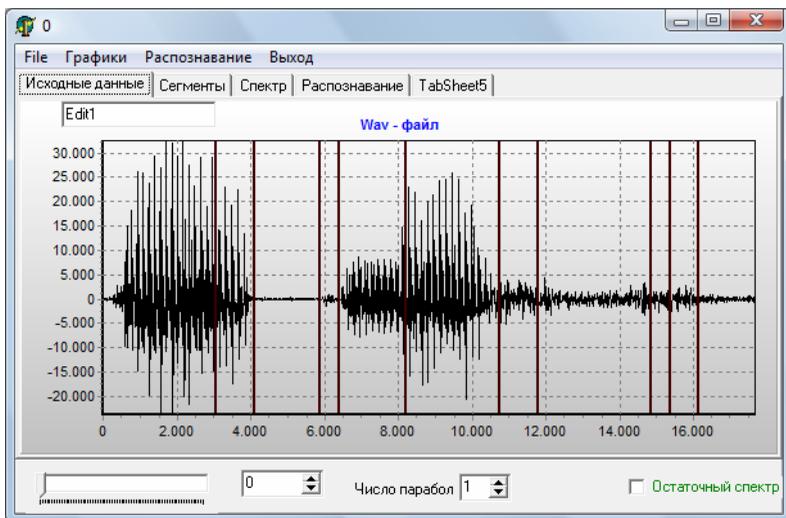


Рис.3. Сегментированный сигнал слова «атлас» без латентных периодов

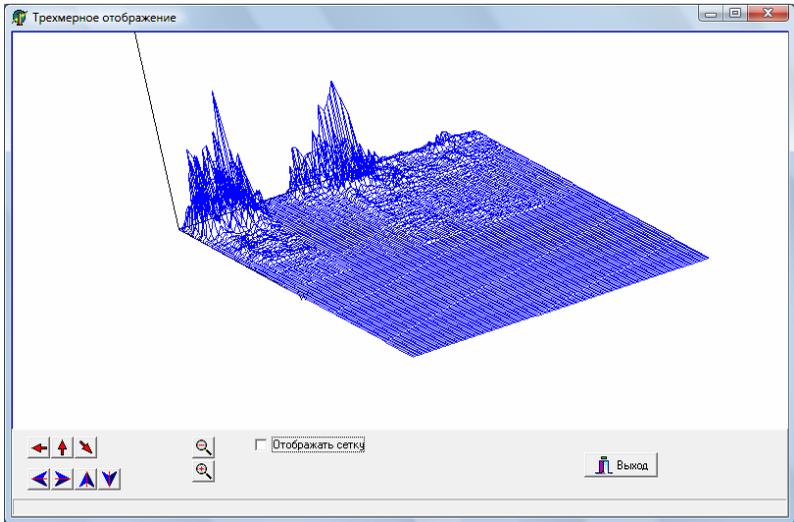


Рис.4. Спектрально-временное представление слова «атлас» без латентных периодов

На рис. 3, 4 показаны сигнал $s(t)$ и спектрально-временное представление $S(\omega_k, T_l)$ без латентных периодов.

Решение задачи выявления чередования тональных и шумных звуков строится по параметрам тональности, шумности и интенсивности на интервалах анализа.

Графики, которые иллюстрируют процесс формирования функции частоты перехода через нуль $\rho(T_l)$, представлены на рис. 1.

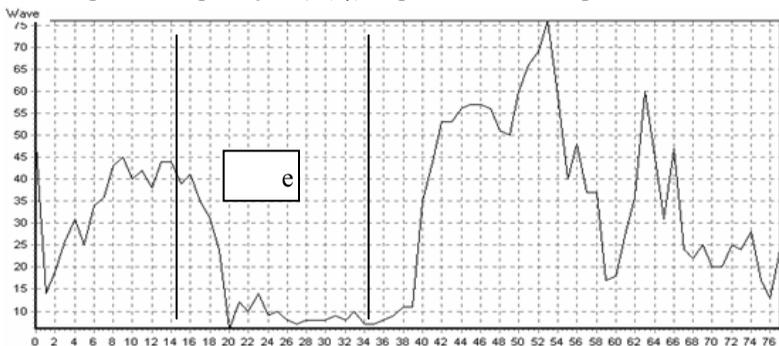


Рис. 5. Функция частоты переходов через нуль $\rho(T_l)$ сигнала слова «шесть»

Как видно из графика рис. 5 уровень частоты переходов через нуль $\rho(T_l)$ для тональных звуков не превышает 15, в то же время как для шумных звуков и глухой паузы в зависимости от фонемы может достигать 60 для интервала анализа $T_{lN} = 220$ отсчётов сигнала $S(t)$.

Допустим, каждая эталонная последовательность Y_k содержит $[y_{k0}(\omega), \dots, y_{ki}(\omega), \dots, y_{km}(\omega)]$ элементов, объединяемых в μ_k групп последовательности, каждая группа в эталонной последовательности содержит γ_j фонем, $j=1 \div \mu_k$, входная последовательность X состоит из $x_0(\omega), \dots, x_i(\omega), \dots, x_r(\omega)$ элементов, объединяемых в θ групп последовательности, каждая группа входной последовательности содержит σ_p фонем, $p=1 \div \theta$. Задача сегментации – определить границ между группами или фонемами [3]. *Сегментация* речи верификацией (рис.6) на принадлежность элементов $[y_{k0}(\omega), \dots, y_{ki}(\omega), \dots, y_{km}(\omega)]$; $[x_0(\omega), \dots, x_i(\omega), \dots, x_r(\omega)]$ к некоторому классу с близкими параметрами осуществляется по правилу [3]

$$d = \underset{u,t}{\text{extrem}} \{x_u\} \# \{x_t\}. \quad (2)$$

Граница определяется как

$$d_1 = \max \{x_u\} \# \{x_{u+\Delta}\}, \quad (3)$$

$$u = v + \delta; v = 0, \Delta, 2\Delta, \dots, r - 2\Delta; \delta = 1 \div \Delta$$

при условии

$$T_{\text{seg}} > 0,2 \text{сек}, \quad (4)$$

где Δ – шаг верификации (в данной работе $\Delta = 3; 2; 2$); $\#$ – операция сопоставления (битового в хемминговом или десятичного в Эвклидовом пространствах).

В эвклидовом пространстве

$$d_{iu} = \sum_{v=0}^{r-2\Delta} \sum_{u=v}^{v+\Delta} \sum_{j=1}^n [x_u(\omega_j) - x_{u+\Delta}(\omega_j)]^2,$$

$$d_{2u} = \sum_{v=0}^{r-2\Delta} \sum_{u=v}^{v+\Delta} \sum_{\delta=0}^{\Delta-1} \sum_{j=1}^n [x_u(\omega_j) - x_{u+\Delta+\delta}(\omega_j)]^2.$$

Если $y_{ki}(\omega)$ и $x_l(\omega)$ – битовые последовательности по ω , то хеммингово расстояние

$$d_{1u}^h = \sum_{v=0}^{r-2\Delta} \sum_{u=v}^{v+\Delta} \sum_{j=1}^n \text{unc}[x_u(\omega_j) \wedge x_{u+\Delta}(\omega_j)],$$

$$d_{2u}^h = \sum_{v=0}^{r-2\Delta} \sum_{u=v}^{v+\Delta} \sum_{\delta=0}^{\Delta-1} \sum_{j=1}^n \text{unc}[x_u(\omega_j) \wedge x_{u+\Delta+\delta}(\omega_j)],$$

где $\text{unc}()$ – функция несовпадений бит операции «исключающее ИЛИ – \wedge ».

Верификацию и, соответственно, сегментацию можно осуществлять на всей длине $T_{\text{фр}}$ слова или фразы. Если же вводится несколько этапов сегментации, отличающихся по принципу определения границ, то верификация осуществляется для несегментированных длительных интервалов. При сегментации по групповым признакам могут образоваться длинные сегменты типа Ш-Ш, С-С, Г-Г, С-Г-С, Г-С-Г (Ш – шумный, С – согласный тональный, Г – гласный). Обычно эти сегменты неразделимы по групповым или отдельным признакам, но всегда делятся по совокупности признаков на этапе верификации на сегменты, соответствующие фонемам.

В качестве признака необходимости проверки сегмента на наличие внутренней границы применяется условие (4). Если внутренней границы нет, то на данном интервале не будет максимума d . Вместе с тем на интервале может быть несколько максимумов. В этом случае выбирается глобальный максимум на интервале $T_{\text{сег}}$ и сегмент делится на два новых. Каждый вновь образовавшийся сегмент проверяется по условию (4) и, если необходимо, делится на две части, реализуя бинарное деление рекурсивной процедурой по правилу (3) и условию (4). Для полной сегментации достаточно три рекурсии с шагом $\Delta=3;2;2$ отсчетов.

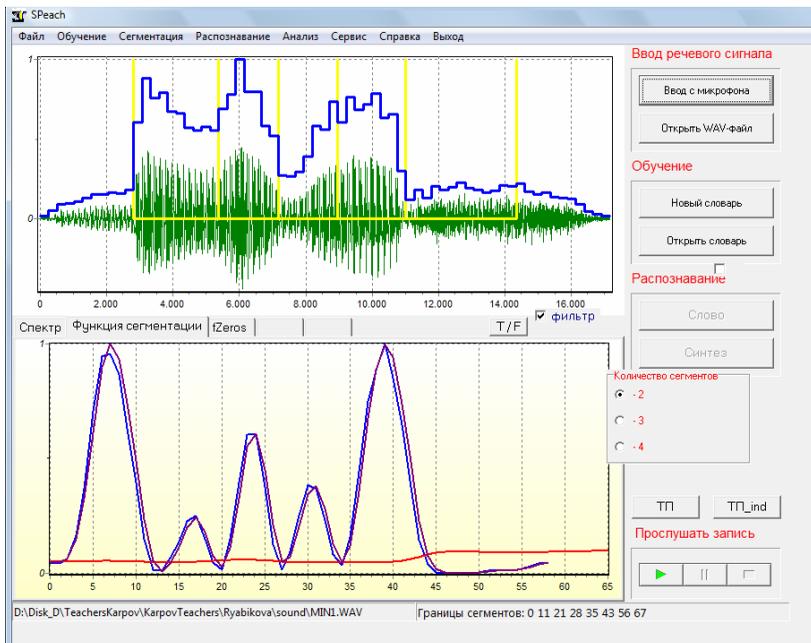


Рис. 6. Функция сегментации слова «минимум»

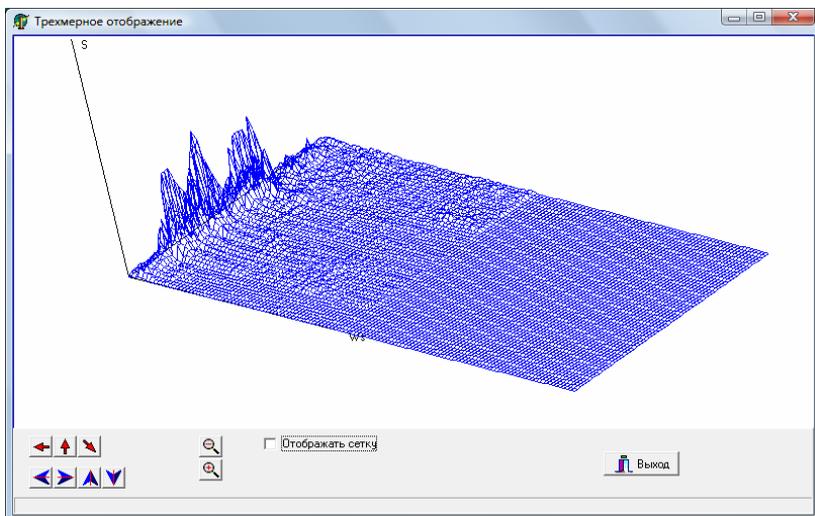


Рис.7. Спектрально-временное представление слова «минимум»

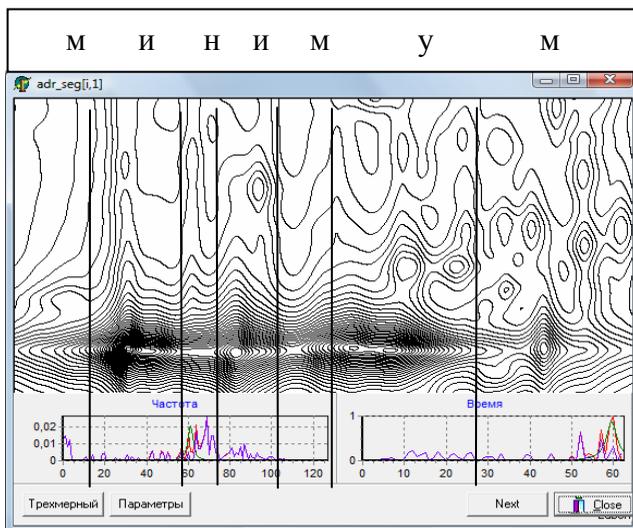


Рис.8. Изолинии интенсивности спектрально-временного представления слова «минимум»

Выводы. В процессе многоэтапной сегментации может возникнуть ситуация ложной сегментации или неполной сегментации. В этом случае можно привлечь очередной этап сегментации, основанный на анализе графического спектрально - временного представления в виде изолиний урвней интенсивности сигнала (рис.7,8). В этом случае подходящими являются два метода: – сегментация на основе анализа контуров; – сегментация на основе анализа областей. В предполагаемых сомнительных точках осуществляется проверка возможности появления границы исходя из оценки существенности и значимости свойств изменения изолиний в окрестностях предполагаемых границ.

Библиографические ссылки

1. **Яне Б.** Цифровая обработка изображений. / Б. Яне – М., 2007-584с.
2. **Карпов О. М.** Методи та алгоритми оцінки ситуативних відхилень параметрів мови людини: навч. пос. / О.М. Карпов, О.А. Чугай, Г.В. Зірнєєва, В.А. Асадулін – Д., 2007. – 64 с.
3. **Карпов О.Н.** Технология построения устройств распознавания речи: монограф. /О.Н. Карпов – Д.: Изд-во Днепропетр. ун-та,2001. –184 с.

Надійшла до редколегії 25.09.11