Литература

1. Харалик Р.М. Статистический и структурный подходы к описанию текстур [Текст] / Р.М. Харалик // ТИИЭР. – 1979. – №5.– С. 98-120.

2. T. Mikolov. Efficient Estimation of Word Representations in Vector Space [Текст]/ T. Mikolov, K. Chen, G. Corrado et al. // Google Inc., Mountain View, CA. – 2013.

А.К. Алимурадов, П.П. Чураков

МЕТОД ПОВЫШЕНИЯ ТОЧНОСТИ СЕГМЕНТАЦИИ СИГНАЛ/ПАУЗА НА ОСНОВЕ КОМПЛЕМЕНТАРНОЙ МНОЖЕСТВЕННОЙ ДЕКОМПОЗИЦИИ НА ЭМПИРИЧЕСКИЕ МОДЫ

(Пензенский государственный университет)

Сегментация на информативные участки и паузы является одной из важных задач при обработке речевых сигналов. Точное обнаружение границ сигнала не только повышает качество обработки, но и уменьшает количество вычислительных и расчетных операций. Поэтому исследование и разработка методов, повышающих точность сегментации сигнал/пауза, являются весьма актуальными.

На сегодняшний день существует много различных подходов к сегментации сигнал/пауза, которые успешно решают проблему эффективного обнаружения границ речевого сигнала. Среди наиболее известных методов сегментации можно выделить следующие:

– методы, основанные на использовании значений кратковременной энергии (*Short-time Energy, STE*) и количества переходов сигнала через нулевое значение в короткие промежутки времени (*Short-time Zero-crossing Rate, ZCR*) [1];

– методы, основанные на использовании значений информационной энтропии (*Information Entropy, IE*) [2];

– методы, основанные на использовании мел-частотных кепстральных коэффициентов (МЧКК, *Mel-frequency cepstrum coefficients, MFCC*) [3];

Проведенные авторские исследования [4] данных методов выявили низкую эффективность в условиях зашумленной обстановки. При отношении сигнал/шум (*Signal-to-Noise Ratio SNR*) 10 дБ коэффициент действительного обнаружения (*Detection rate, DR*) у метода на основе *STE + ZCR* равен всего лишь 72,1%, а у метода на основе на использовании МЧКК - 76,2%.

Основная причина больших погрешностей в сегментации связана с использованием неэффективных и неадаптивных методов обработки сложных нестационарных и зашумленных речевых сигналов. В данной статье авторами предлагается метод сегментации сигнал/пауза, с использованием: адаптивного анализа на основе комплементарной множественной декомпозиции на эмпирические моды (КМДЭМ) [5]; правила разграничения на основе физиологическо-

го аспекта формирования речи и функционала слухового аппарата человека [6]. Статья является продолжением ранее опубликованных работ авторов [7 - 9].

Суть метода заключается в сегментации речевого сигнала на кратковременные фрагменты для адаптивного анализа с помощью КМДЭМ, с последующим формированием адаптивного порога на основе физиологического аспекта формирования речи, функционала слухового аппарата человека и оценки энергии эмпирических мод (ЭМ). На рисунке 1 представлена блок-диаграмма алгоритма сегментации сигнал/пауза на основе предложенного метода.

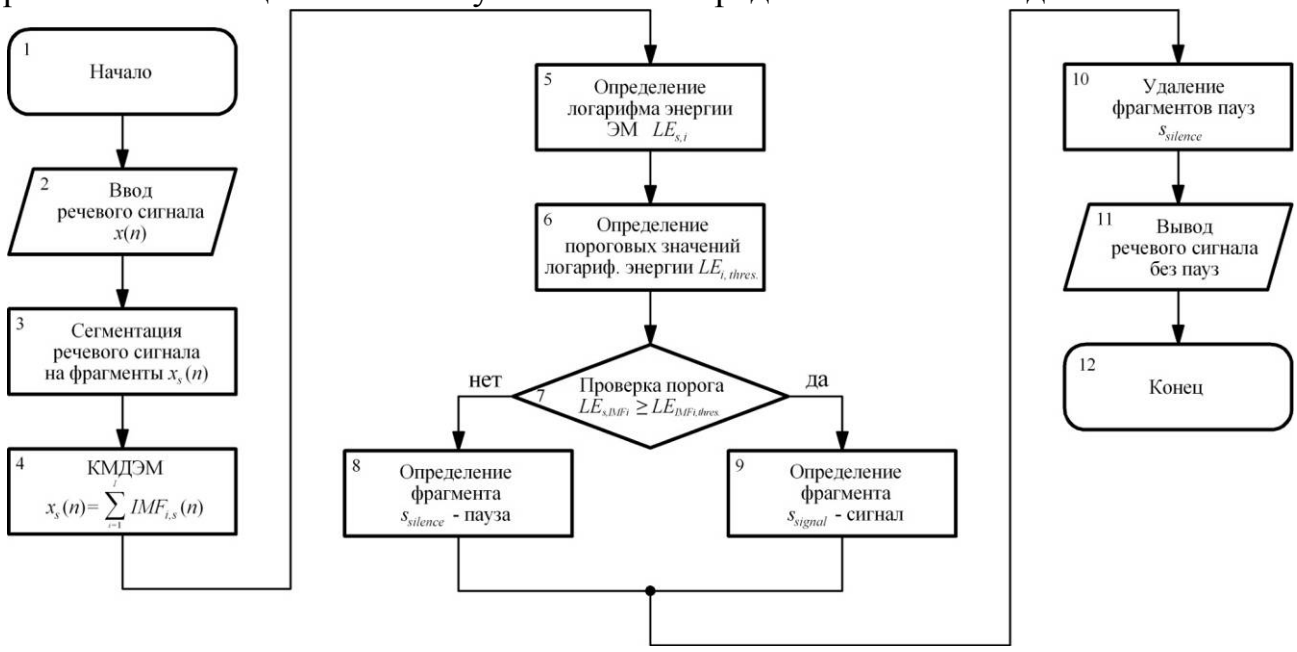

Рисунок 1 – Блок-диаграмма алгоритма сегментации сигнал/пауза на основе предложенного метода

Рассмотрим подробнее основные этапы работы алгоритма. Важным условием помехоустойчивой сегментации речевых сигналов, является возможность формирования адаптивного базиса, функционально зависимого от содержания исходного сигнала. Такой подход реализуется в математическом аппарате, называемом КМДЭМ, который представляет собой адаптивную технологию разложения сигнала на внутренние функции, называемые ЭМ. В результате КМДЭМ, из каждого фрагмента речевого сигнала $x_s(n)$ извлекается конечное число ЭМ (блок 4):

$$x_s(n) = \sum_{i=1}^{I-1} IMF_{s,i}(n) \qquad (1),$$

где $x_s(n)$ - фрагмент речевого сигнала, n - дискретный отсчет времени, s=(0, 1, 2,…, S-1) - номер фрагмента, $IMF_{s,i}(n)$ - полученные ЭМ, i=1,2,…I - номер ЭМ.

В отличии от других методов декомпозиции, особенностью КМДЭМ является многократное добавление к исходному речевому сигналу белого шума с прямыми и инверсными значениями амплитуды и вычисление среднего значения ЭМ, как конечного истинного результата [5]:

$$\begin{bmatrix} y_{s,j}(n) \\ y_{s,j}(n)^* \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} * \begin{bmatrix} x_s(n) \\ w_j(n) \end{bmatrix} \qquad (2),$$

где $w_j(n)$ - добавленный белый шум; $y_j(n)$ - сумма зашумленного фрагмента речевого сигнала $x_s(n)$ с белым шумом; $y_j(n)^*$ - сумма зашумленного фрагмента речевого сигнала $x_s(n)$ с инверсным значением амплитуды белого шума.

$$IMF_{s,i}(n) = \frac{\sum_{j=1}^{J} IMF_{s,ji}(n)}{J} \quad (3),$$

где $IMF_{s,ji}(n)$ - ЭМ, полученные при различных декомпозициях сигналов $y_{s,j}(n)$ и $y_{sj}(n)^*$, j=1,2,…J - количество циклов декомпозиций (добавлений к сигналу белого шума).

Человеческий слуховой аппарат фиксирует речь нелинейно, различия между энергиями участков полезного сигнала и паузы должно быть значительным, чтобы человек фиксировал изменение амплитуды. Для увеличения амплитуды в 2 раза необходимо чтобы энергия увеличилась в 8 раз. Для сжатия амплитуды сигнала при большом динамическом диапазоне применяют логарифмирование энергии фрагментов, максимально приближая работу алгоритма к функционалу слухового аппарата человека (блок 5).

$$LE_{s,i} = \log_2 \sum_{n=1}^{N} \left[IMF_{s,i}(n)\right]^2 \quad (4),$$

где $LE_{s,i}$ - логарифм энергии ЭМ фрагмента речевого сигнала.

В соответствии с физиологическим аспектом формирования речи человек перед произношением делает кратковременную паузу - обычно 200 мс или более. Этот участок не содержит речи и соответствует тишине с фоновым шумом, т.е. значения логарифмов энергии ЭМ фрагментов первых 200 мс будут соответствовать значениям паузы. Используя усредненные значения логарифмов энергии ЭМ можно сформировать пороговые значения логарифмов энергии для обнаружения границы полезного сигнала и паузы. Далее выполняется сравнение значений логарифмов энергии ЭМ остальных фрагментов с пороговыми значениями $LE_{s,IMFi} \geq LE_{IMFi,thres.}$. В случае если условие выполняется, то фрагмент считается полезным сигналом s=s_{signal}, а если условие не выполняется, то фрагмент считается паузой s=$s_{silence}$.

В качестве критерия эффективности предложенного метода сегментации сигнал/пауза использовался коэффициент действительных обнаружений:

$$DR_{speech} = \frac{S_{cor.speech}}{S_{cor.speech} + S_{n.cor.speech}} \times 100\% \quad (5),$$

где $S_{cor.speech}$ - действительный фрагмент сигнала, $S_{n.cor.speech}$ - мнимый фрагмент сигнала.

Для исследования предложенного метода сформирована тестовая выборка из 50 зашумленных речевых сигналов со значениями *SNR* от 10 до 35 дБ с шагом 5 дБ. Результаты исследования оценивались в сравнении с известными методами сегментации сигнал/пауза (см. таблицу 1):

Таблица 1 – Сравнительный анализ результатов сегментации сигнал/пауза.

Параметр входного сигнала SNR_{IN}, дБ	DR_{speech}, %				
	STE	*STE+ZCR*	*IE*	*MFCC*	Предложенный метод
10	66,3	67,1	73,2	76,2	71,3
15	71,5	74,9	80,5	81,4	84,3
20	77,9	79,3	82,1	84,5	87,6
25	80,2	82,6	87,4	88,1	92,5
30	85,4	88,3	90,3	90,7	94,3
35	88,3	90,2	92,4	92,1	95,2

Как видно из результатов предложенный метод обеспечивает наилучший результат сегментации (особенно с малыми значениями *SNR*): в среднем на 9 % лучше, чем метод *STE* и *STE+ZCR*; на 5 % лучше, чем метод *IE*; на 4 % лучше, чем метод *MFCC*. Таким образом, использование предложенного метода в обработке речевых сигналов позволит значительно повысить точность сегментации сигнал/пауза.

Литература

1. R.G. Bachu, S. Kopparthi, B. Adapa, and B.D. Barkana, "Separation of voiced and unvoiced using zero crossing rate and energy of the speech signal," Proc. Amer. Soc. Eng. Educ. (ASEE) Zone Conference, 2008, pp. 1–7.

2. N. Mattias, "Entropy and speech," Sound and Image Processing Laboratory School of Electrical Engineering KTH (Royal Institute of Technology), Stockholm, 2006, p. 54.

3. S. Ahmadi and A.S. Spanias, "Cepstrum-based pitch detection using a new statistical V/UV classification algorithm," IEEE Transactions on Speech and Audio Processing, vol. 7, No. 3, 2002, pp. 333–338.

4. Алимурадов А.К. Помехоустойчивый адаптивный алгоритм сегментации «сигнал/пауза» для систем распознавания речи / А.К. Алимурадов, П.П. Чураков // Известия высших учебных заведений. Поволжский регион. Технические науки. - 2015. - № 2 (34). - С. 82 - 94.

5. J.-R. Yeh, J.-S. Shieh, and N.E. Huang, "Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method," Adv. Adapt. Data Anal., vol. 2, No. 2, 2010, pp. 135–156, doi: 10.1142/S1793536910000422.

6. V. Sarma and D. Venugopal, "Studies on pattern recognition approach to voiced-unvoiced-silence classification," Proc. IEEE Int. Conf. ICASSP'78 Acoust., Speech, and Signal Process., vol. 3, Apr. 1978, pp.1–4.

7. Алимурадов А.К. Применение преобразования Гильберта-Хуанга в задаче выделения информативных признаков речевых сигналов / А.К. Алимурадов, А.Ю. Тычков // Международный научно-исследовательский журнал. - 2013. - № 5-1 (12). - С. 57 - 58.

8. Алимурадов А.К., Муртазов Ф.Ш. Методы повышения эффективности распознавания речевых сигналов в системах голосового управления. Измерительная техника. - 2015. - № 10. - С. 20 - 24.

9. Алимурадов А.К. Адаптивная компенсация помех речевых сигналов с использованием комплементарной множественной декомпозиция на эмпирические моды / А.К. Алимурадов // «Молодежь и XXI век - 2015»: материалы V Международной молодежной научной конференции (26-27 февраля 2015 года), в 3-х томах, Том 2, Юго-Зап. гос. ун-т., ЗАО «Университетская книга», Курск, 2015, С. 96 - 99.

М.Е. Бурлаков

ИССЛЕДОВАНИЕ ДИНАМИКИ АКТИВНОСТИ ПУБЛИКАЦИИ УГРОЗ В ОТКРЫТЫХ И ЗАКРЫТЫХ ИСТОЧНИКАХ ДАННЫХ

(Самарский национальный исследовательский университет им. С.П. Королёва)

Введение. В настоящее время остро стоит вопрос, связанный с обнаружением угроз и уязвимостей в области информационных технологий и программных комплексов. Более 87% пользователей компьютерных систем используют в качестве рабочей операционной системы - ОС *Windows* (в вариациях версий *Windows XP*, *Windows 7*, *Windows 8* и *Windows 10*) [1]. Исходя из этой статистики, злоумышленники не первый год уделяют повышенное внимание как данной операционной системе, так и программным решениям, функционирующих в рамках ОС Windows. Так, например, по данным компании *Hewlett Packard* [2] на сегодняшний день можно выделить более 500 классов различных уязвимостей в ПО.

К наиболее частым уязвимостям, по версии команды *TeamSHATTER*, в рамках информационных систем можно отнести [3]:

1. наличие специализированных логинов и паролей (пустые, по умолчанию, легко поддающиеся подбору);
2. SQL инъекции в различных реализациях взаимодействий с базами данных;
3. некорректно выданные права пользователя-исполнителя, ошибки в настройке привилегий групп;
4. слабое администрирование баз данных в части настройки необходимого функционала;
5. некорректная настройка разного рода конфигураций;
6. переполнение буфера (стека);
7. повышение привилегий;
8. атаки отказа в обслуживании (*DDoS*);
9. отсутствие своевременного обновление компонент безопасности баз данных;
10. хранение данных в отрытом (незащищенном виде).