

УДК 004.896, 004.934



МЕТОД АВТОМАТИЧЕСКОГО ОПРЕДЕЛЕНИЯ ГРАНИЦ ШУМОВОГО ТУРБУЛЕНТНОГО ИСТОЧНИКА В ЗВУКОВОМ СИГНАЛЕ

Хашан Т.С.

Институт прикладной математики и механики НАН Украины, г. Донецк, khashan@mail.ru

В статье изложен метод автоматического поиска границ шумового турбулентного источника в звуковом сигнале. Предложенный метод основан на многомерном анализе энергетической завершенности исследуемых отрезков сигнала, сравнение последовательностей его характеристик и их сопоставлении. На основе рассматриваемого метода разработан алгоритм и проведены численные исследования.

ШУМОВОЙ ТУРБУЛЕНТНЫЙ ИСТОЧНИК, СЕГМЕНТАЦИЯ, РАСПОЗНАВАНИЕ РЕЧИ, АМПЛИТУДНО-ВОЛНОВОЕ ПРЕДСТАВЛЕНИЕ ВОЛНОВЫХ ПРОЦЕССОВ

Введение

Одним из важных направлений исследований в области искусственного интеллекта является разработка интеллектуальных систем автоматического распознавания звуковых сигналов (ЗС). Качество работы этих систем зависит от первоначальной обработки сигнала и подготовки его к дальнейшим исследованиям. Данная статья является продолжением исследований, связанных с разработкой методов и алгоритмов для интеллектуальной системы технического слуха малых мобильных роботов [1-3], одной из подсистем которой является система автоматической сегментации ЗС.

Следует отметить, что под сегментацией ЗС подразумевают определение границ фрагментов сигнала, на котором могут присутствовать голосовой, шумовой турбулентный, шумовой нетурбулентный источники, а также голосовой и шумовой источники, функционирующие одновременно [4].

Ранее в [2-6] было отмечено, что на сегодняшний день все разрабатываемые системы автоматического сегментации ЗС являются дикторозависимыми, так как идеология их построения основывается на анализе спектральных характеристик, определенных на окнах фиксированной длины (фреймах). При этом размер окна, а также его границы не зависят от природы сигнала. Такой подход приводит к назначению порогов, на основании которых принимается решение о наличии границы между источниками различной природы [7-10]. Очевидно, что заданные значения порогов одного и того же участка для одного и того же диктора изменяются в широких пределах.

Поэтому актуально создание дикторонезависимых методов и алгоритмов автоматического определения границ фрагментов ЗС, соответствующих одному из вышеприведенных источников.

В работах [4-6] предложена идея, позволяющая отказаться от окон фиксированной длины, основанная исключительно на анализе амплитудно-временного представления волнового процесса (АВП ВП) и энер-

гетической завершенности исследуемых фрагментов (ЭЗФ), выдвижении гипотез и проверке их истинности с учетом факторов закономерности и случайности поведения характеристик сигнала на исследуемых ЭЗФ.

Эта идея развита в [2,3], где решена задача автоматического выбора границ фрагментов сигнала, на котором функционирует голосовой источник. В [5] предложен алгоритм, позволяющий выделять сегменты сигнала, соответствующие функционированию шумового турбулентного источника. Однако анализ данного метода показал, что он эффективен лишь в том случае, когда в исследуемом сигнале присутствует только один шумовой источник, либо характеристики других источников близки к первому найденному. Кроме того, границы искомым фрагментов определяются с ошибками. На основании сказанного сформулируем цель и задачи данной работы.

Целью данной работы является усовершенствование дикторонезависимого метода и алгоритма автоматического определения границ фрагментов речевого сигнала, соответствующих функционированию шумовых турбулентных источников.

Для достижения поставленной цели необходимо решить следующие задачи:

- проанализировать метод, представленный в [5];
- на основании анализа усовершенствовать метод автоматического выделения фрагментов АВП ВП, соответствующих функционированию шумовых турбулентных источников (ШТИ);
- программно реализовать алгоритм и провести численные исследования.

1. Метод автоматического определения границ фрагментов, соответствующие функционированию ШТИ. Анализ его эффективности

В основе метода решения поставленной задачи лежит анализ функции $U = \{u_n\}_{n=1}^N$ – АВП ВП (рис.1) и ее характеристик: $D = \{d_i\}_{i=1}^M$ – длин полных колебаний (ПК), $L = \{[l_1^i, l_2^i, l_3^i, l_4^i, l_5^i, l_6^i]\}_{i=1}^M$ – границ ПК,

$R = \{r_i\}_{i=1}^M$ – размахов амплитуд на ПК (рис.2), где M – количество ПК функции U , а также энергетически завершенных фрагментов R_j^1 (ЭЗФ) с границами $\{\{lr_1^j, lr_2^j, lr_3^j, lr_4^j, lr_5^j\}\}$, где $j = \overline{1, J}$ – количество ЭЗФ функции U (рис.3). Отметим, что характеристики функции U получены согласно правилам, представленным в [2].



Рис.1. АВП ВП слова «СУШУ»

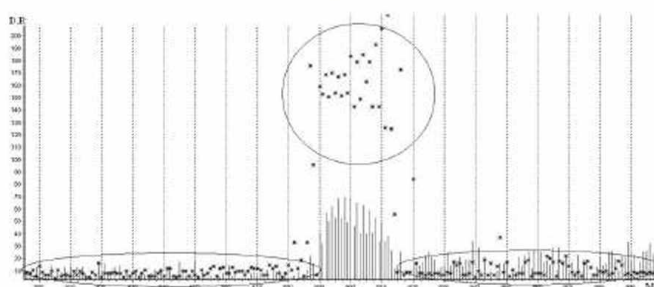


Рис. 2. Фрагмент «СУШ» слова «сушу» в виде $\{d_i, r_i\}_{i=1}^N$: d_i представлено знаком *, r_i – знаком |

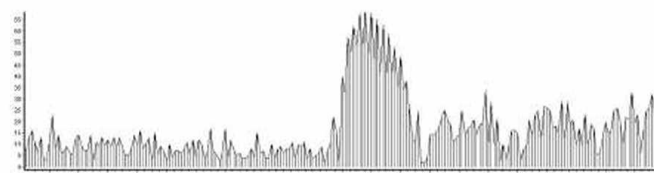


Рис. 3. Фрагмент «СУШ» слова «сушу», представленного R_j^1

Анализируя характеристики функции U на фрагментах, соответствующих функционированию ШТИ и его окружения, в [5] были сделаны следующие выводы: длины ПК, принадлежащих фрагментам ШТИ, меньше длин ПК, не принадлежащих объекту (см. рис. 2); длины ПК, принадлежащих фрагментам ШТИ, близки.

Однако, как было показано авторами [5], невозможно определить близость ПК на исследуемых фрагментах, так как они ведут себя случайным образом. Анализ ЭЗФ также подтверждает случайность поведения характеристик. Поэтому для описания поведения длин ПК вводятся:

– функция F , которая отражает количество ПК в единицу времени:

$$F = \{F(R_j^1) : F(R_j^1) = \frac{N(R_j^1) \cdot f}{D(R_j^1)}, j = \overline{1, J}\}, \quad (1)$$

где $N(R_j^1)$ – количество ПК в ЭЗФ R_j^1 ; $D(R_j^1)$ – дли-

на ЭЗФ R_j^1 ; f – частота дискретизации сигнала по времени; J – количество ЭЗФ.

– множество

$$\begin{aligned} G^1 &= \{R_k^1 : F(R_k^1) > \frac{(\overline{M} + \overline{m})}{2}, \\ k &= \overline{k', k''}, F(R_{k-1}^1) \leq \frac{(\overline{M} + \overline{m})}{2}, \\ F(R_{k+1}^1) &\leq \frac{(\overline{M} + \overline{m})}{2}, \end{aligned} \quad (2)$$

где $\overline{M} = \max_{1 \leq j \leq J} (F(R_j^1))$ и $\overline{m} = \min_{1 \leq j \leq J} (F(R_j^1))$.

Значения функции $F(R_k^1)$, принадлежащие фрагментам ШТИ, будут больше, чем значения функции $F(R_{k+1}^1)$ и $F(R_{k-1}^1)$, принадлежащие их окружению, $R_{k+1}^1 \notin G^1$ $R_{k-1}^1 \notin G^1$ время существования ЭЗФ

$$R_k^1 < R_{k+1}^1 \text{ и } R_k^1 < R_{k-1}^1, \\ R_k^1 \in G^1 \quad R_{k+1}^1 \notin G^1 \quad R_k^1 \in G^1 \quad R_{k-1}^1 \notin G^1$$

Из определения множества G^1 следует, что $F(R_s^1) \in [\min_{R_k^1 \in G^1} F(R_k^1); \overline{M}]$, $F(R_s^1) \in G^1$.

Алгоритм, разработанный на основании данного метода позволяет «близко» определить границы объектов, соответствующие ШТИ. На рис. 4 показан пример функционирования алгоритма на примере слова «силаща». Типизация выделенных фрагментов (их объединения с учетом случайности поведения функции на соседних фрагментах) (рис. 5а, б) устраняет «провалы и всплески», однако не уточняет границы искомого фрагмента.

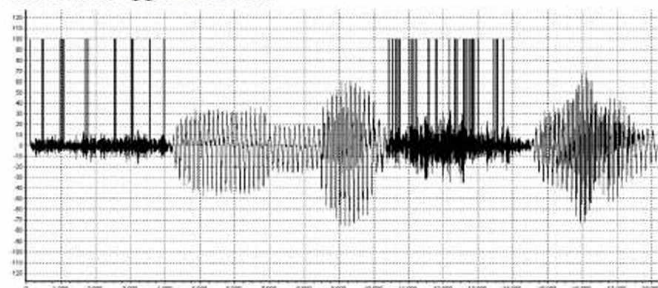


Рис. 4. Фрагменты ШТИ, выделенные в результате работы алгоритма (слово «силаща»)

2. Усовершенствование метода автоматического определения границ ШТИ в ЗС

Для уточнения границ фрагментов, соответствующих функционированию ШТИ, проведем анализ последовательностей ЭЗФ первого и второго уровней, а также ПК для каждой из последовательностей.

Отметим, что последовательности ЭЗФ второго уровня строятся из последовательностей ЭЗФ первого уровня следующим образом: выбираются максимальные значения размахов амплитуд на R_j^1 и анало-

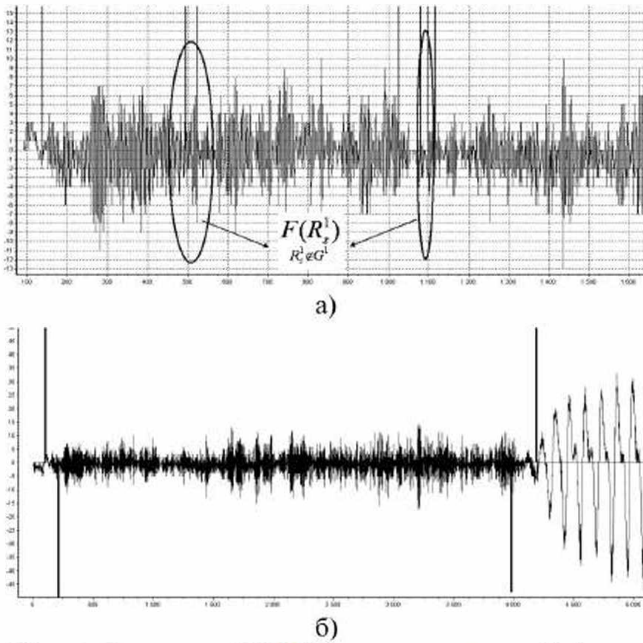


Рис. 5. Фрагмент АВП ВП, соответствующий функционированию ШТИ: а) до типизации объектов (звук «с»); б) после типизации объектов (границы после типизации отмечены знаком |, нижняя часть графика, «си»)

гично, только уже для второго уровня, из этих значений строится ЭЗФ уровня 2 – $R_{j_2}^2$.

$$R^2 = (r_{i_2}^2 : r_{i_2}^2 = \max_{r_i^1 \in R_j^1} (r_i^1)), \quad (3)$$

$$\mu \in [lr_{1}^{j_2,1}; lr_{5}^{j_2,1}], i_2 = \overline{1; M_2}$$

$$R_{j_2}^2 : r_{i-1}^2 > r_i^2 = r_{i+p_1}^2, \quad (4)$$

$$r_{i+p_1}^2 < r_{i_2}^2 \leq r_{i_2+p_2}^2 < r_{i_3}^2 = r_{i_3+p_3}^2,$$

$$r_{i_3+p_3}^2 > r_{i_4}^2 \geq r_{i_4+p_4-1}^2 > r_{i_4+p_4}^2,$$

$$r_{i_4+p_4}^2 = r_{i_5}^2 = r_{i_5+p_5}^2 < r_{i_5+p_5+1}^2,$$

где $\exists s1 \in N : \forall p1 = \overline{0, s1}, s2 \in N : \forall p2 = \overline{0, s2},$

$$s3 \in N : \forall p3 = \overline{0, s3}, s4 \in N : \forall p4 = \overline{0, s4},$$

$$s5 \in N : \forall p5 = \overline{0, s5}, j_2 = \overline{1, J_2},$$

J_2 – количество ЭЗФ второго уровня;

M_2 – количество ПК и, соответственно, размахов амплитуд, составляющих уровень 2.

Рассматривая длины ПК на R_j^1 , а также определяя их близость согласно условиям (5) и (6), принимается решение о выборе границ искомым фрагментов. В случае, когда эти условия не выполняются (то есть на таких ЭЗФ фиксируются «всплески» или «провалы») определяется количество «близких» между собой ПК внутри ЭЗФ. Если их большинство, то принимается решение о случайности поведения функции U на этих

участках (рис. 5б – границы отмечены знаком |, верхняя часть рисунка).

$$l_1^{i+k1} \leq l_{\max}^i + d_i + S < l_4^{i+k1}, \quad (5)$$

$$l_3^{i+k1} \leq l_{\min}^i + d_i + S < l_6^{i+k1},$$

$$l_1^i \leq l_{\max}^{i+k1} - d_{i+k1} - S < l_4^i \leq l_{\min}^{i+k1} - d_{i+k1} - S < l_6^i.$$

$$0 < \frac{r_i^+}{2} \leq r_{i+k1}^+ < \frac{r_{i+k1}^+}{2} \leq r_i^+, \quad (6)$$

$$r_{i+k1}^- \leq \frac{r_i^-}{2} < 0, r_i^- \leq \frac{r_{i+k1}^-}{2} < 0.$$

$S = l_1^{i+k1} - l_6^i$ – расстояние между ПК.

Дополнительным условием принадлежности к искомым фрагментам является условие (7) (это условие выполнимо для большинства сигналов, однако оно не является обязательным):

$$G^m = (r_m^1 : r_m^1 < \frac{\sum_{i=1}^M r_i^1}{M}, m = \overline{m_1, m_2}). \quad (7)$$

На отрезках АВП ВП, определенных вышеуказанным способом, строится ЭЗФ 2 уровня – условия (3), (4). Далее на нем уточняется поведение функции $F(R_{j_2}^2)$.

$$F^2 = \{F(R_{j_2}^2) : F(R_{j_2}^2) = \frac{N(R_{j_2}^2) \cdot f}{D(R_{j_2}^2)}, j_2 = \overline{1, J_2}\}, \quad (8)$$

где $N(R_{j_2}^2)$ – количество ПК в ЭЗФ $R_{j_2}^2$; $D(R_{j_2}^2)$ – длина ЭЗФ $R_{j_2}^2$; f – частота дискретизации сигнала по времени; J_2 – количество ЭЗФ второго уровня.

Очевидно, что значения функции $F(R_{k_2}^2)$, принадлежащие области $R_{k_2}^2 \in G^2$

лежащие фрагментам ШТИ, будут больше, чем значения функций $F(R_{k_2+1}^2)$ и $F(R_{k_2-1}^2)$, принадлежащих его окружению. Область G^2 строится следующим образом:

$$G^2 = \{R_{k_2}^2 : F(R_{k_2}^2) > \frac{(\overline{M_2} + \overline{m_2})}{2}, \quad (9)$$

$$k_2 = \overline{k'_2, k''_2}, F(R_{k_2-1}^2) \leq \frac{(\overline{M_2} + \overline{m_2})}{2},$$

$$F(R_{k_2+1}^2) \leq \frac{(\overline{M_2} + \overline{m_2})}{2}\},$$

где $\overline{M_2} = \max_{1 \leq j_2 \leq J_2} (F(R_{j_2}^2))$ и $\overline{m_2} = \min_{1 \leq j_2 \leq J_2} (F(R_{j_2}^2))$.

На основании предложенного метода был разработан алгоритм и проведены следующие численные исследования.

3. Численные исследования

Основные параметры: частота дискретизации $f = 22050$ Гц, 8-битный сигнал, запись сигнала производилась при помощи программы SoundForge.

Аналізу подвергались около 300 слов, произнесенных 10 дикторами (3 женщины и 7 мужчин, возраст от 22 до 60 лет). Во всех случаях границы фрагментов ШТИ были выявлены точно и совпадают с результатами сегментации ЗС оператором вручную на основании прослушивания соответствующих фрагментов. Это свидетельствует о высоком показателе по сравнению с базовым методом. На рис.6 показан результат функционирования модифицированного алгоритма на примере слова «лошак» (диктор – мужчина, 23 года). Как видно из рисунка, алгоритм точно определил все фрагменты, соответствующие ШТИ.

Следует отметить, что в данном методе не устранен пока недостаток неразделимости двух подряд идущих фрагментов, соответствующих ШТИ, например слово «шить» (рис.7). Этот недостаток поясняется тем, что характер поведения функции F на данных фрагментах, а также все характеристики U идентичны.

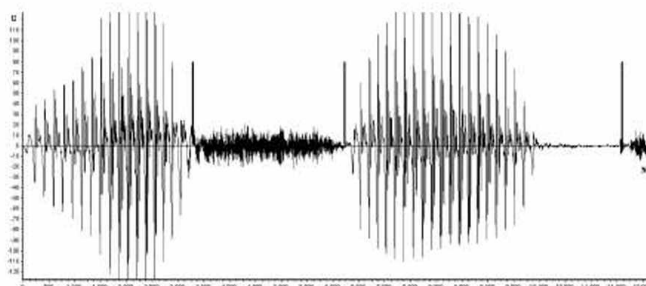


Рис. 6. Выявленные на АВП ВП фрагменты ШТИ слова «лошак»: границы фонем «ш» и «к»

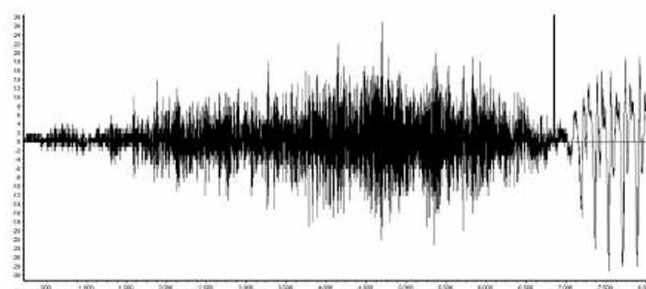


Рис. 7. Фрагмент АВП ВП «СШИ» слова «шить»

Выводы

В статье предлагается модифицированный метод выделения границ фрагментов звуковых сигналов, соответствующих функционированию шумового турбулентного источника. Предлагаемый метод не зависит от диктора, а также от фонетического и фонологического состава. В отличие от базового метода [5], подход, описанный в данной статье, позволяет выделить все искомые объекты, а также точно определить их границы.

На основании базового и предложенного методов, представляется возможным перейти к решению задач по автоматическому определению границ фрагментов

АВП ВП, соответствующих функционированию шумового нетурбулентного источника, а также голосового и шумового источников, функционирующих одновременно. Решение указанных задач позволит в дальнейшем перейти к решению задачи распознавания звуковых сигналов.

Список литературы: 1. *V.E.Pavlovsky, A.S.Myagkov, T.S.Khashan, V.V.Pavlovsky.* Concept, Simulation and Elaboration of Audition Sensors for Robots // Proc. of The IARP Workshop «Adaptive and Intelligent Robots: Present and Future». – Moscow, Nov 24-26 2005. – IPMech RAS, RAS. – vol.1. – pp. 90–100. 2. *Xashan T.C.* Сегментация речевого сигнала // Искусственный интеллект. – 2007. – №.4. – С. 386–397. 3. *Xashan T.C., Писаренко Ю.В.* Алгоритм пеленгаши речевых и специальных сигналов системой технического слуха в интеллектуальной робототехнике // КЗИМ. – 2008. – №.7. – с. 66–75. 4. *Мышко С.В., Блохин С.В., Доценко А.В.* Определение на амплитудно-временном представлении звуковых волн элементарных и однородных квазипериодических структур // Биофизика сложных систем. – 1993. – С. 92–100. 5. *Доценко А.В., Мышко С.В., Шевченко А.И., Блохин С.В.* Выделение фрагментов амплитудно-временного представлении звуковых волн, являющихся результатом функционирования шумового турбулентного источника // Биофизика сложных систем. – 1993. – С. 85–92. 6. *Мышко Н.В., Доценко А.В., Блохин С.В., Мышко С.В.* Определение энергетически завершенных фрагментов амплитудно-временного представления звуковых волн // Биофизика сложных систем. – 1993. – С. 103–117. 7. *Старушко Д.Г., Шелепов В.Ю.* Новая система признаков для распознавания речевых единиц // Искусственный интеллект. – 2002. – № 4. – С. 286–288. 8. *Yermolenko T.V., Fedorov E.E.* The procedures of the noise clipping in the signal, based on Fourier- and wavelet-transform and on classification of sounds of speech // Труды межд. конф. SPECOM'2006. – 2006. – P.245–250. 9. *Карпов О.Н.* Технология построения устройств распознавания речи: Моногр. – Днепропетровск: Из-во Днепропетровского ун-та, 2001. – 184 с. 10. *Шелепов В.Ю., Нищенко А.В.* Амплитудная сегментация речевого сигнала, использующая фильтрацию и известный фонетический состав // Искусственный интеллект. – 2003. – № 3. – С. 421–426.

Поступила в редколлегию 19.02.2009

УДК 004.896, 004.934

Метод автоматичного визначення границ шумового турбулентного джерела у звуковому сигналі / Т. С. Хашан // Біоніка інтелекту: наук.-техн. журнал. – 2009. – №1(70). – С. 68–71.

Запропоновано вдосконалений метод виділення границь фрагментів звукового сигналу, що відповідають функціонуванню шумового турбулентного джерела. Запропонований метод не залежить від диктора, а також від фонетичного й фонологічного складу.

Л.: 7. Бібліогр.: 10 найм.

UDC 004.896, 004.934

Method for automatically determining the boundaries of the turbulent noise source in audio signal / Khashan Tatiana S. // Bionics of Intelligence: Sci. Mag. – 2009. – №1(70). – P. 68–71.

Proposed an improved method for the speech signal fragment's boundaries allocation, which corresponds to the turbulent noise source activity. The proposed method is a speaker-independent and does not depends from phonetic and phonological structure of the speech.

Fig.: 7. Ref.: 10 items.