# Теоретические основы непрерывных саундлетов для распознавания вокальных звуков речи

***Фёдоров Е.Е.,*** *доктор технических наук, доцент*

***Слесорайтите Э.,*** *старший преподаватель*

В статье изложены теоретические основы непрерывных саундлетов, которые применяются в метрическом подходе к распознаванию вокальных звуков. Предложены материнский и дочерний непрерывные саундлеты и исследованы свойства саундлетных отображений, которые позволяют учитывать структуру квазипериодического сигнала и сопоставлять образцы вокальных звуков речи разной длины. На основе саундлетов и саундлетных отображений разработаны метод создания образцов, метод формирования эталонных образцов и модель распознавания образцов, которые используются в режимах обучения и распознавания интеллектуальной системы.

*• непрерывный саундлет • материнский саундлет • дочерний саундлет • саундлетное отображение • эталонные образцы вокальных звуков • метрический подход*

---

In article theory continuous soundlets which are applied in the metric approach to recognition of vocal sounds are stated. Are offered parent and child continuous soundlets and properties soundlets mapping which allow to consider structure quasi-periodic a signal are investigated and to compare patterns of vocal sounds of speech of different length. On a basis soundlets and soundlets mapping methods of creation of patterns, a method of formation of reference patterns and model of recognition of patterns which are used in modes of training and recognition of intellectual system are developed.

*• continuous soundlet • parent soundlet • child soundlet •soundlet mapping • reference patterns of vocal sounds • the metric approach.*

## ОБЩАЯ ПОСТАНОВКА ПРОБЛЕМЫ

На сегодняшний день актуальной является разработка специализированных процессоров компьютерных систем предназначенных для распознавания речи человека, синтеза речи и др., и используемых в интеллектуальных компьютерных системах.

В основе данной задачи лежит проблема построения эффективных методов, обеспечивающих высокую скорость обучения модели распознавания, а также высокую вероятность, адекватность и скорость распознавания речевых сигналов.

## АНАЛИЗ ИССЛЕДОВАНИЙ

Современные системы распознавания речевых образов используют следующие подходы: логический, метрический, байесовский, нейросетевой, структурный. Существующие методы и модели распознавания речевых образов обычно основаны на скрытых марковских моделях (СММ) [1-3], алгоритме динамического программирования DTW [1-2], и искусственных нейронных сетях [4-6] и обладают следующими недостатками [7]:

- время обучения несколько месяцев;
- хранение большого количества эталонов звуков или слов, а также весовых коэффициентов;
- большое время распознавания;
- вероятность распознавания меньше 95 %;
- наличие сотен тысяч обучающих образцов/

## ПОСТАНОВКА ЗАДАЧ ИССЛЕДОВАНИЯ

Целью работы является разработка теоретических основ непрерывных саундлетов и формируемых на их основе саундлетных отображений для метрического подхода к распознаванию вокальных звуков речи.

## РЕШЕНИЕ ЗАДАЧ И РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ

Для достижения поставленной цели необходимо:

1. Разработать метод создания образцов вокальных звуков.
2. Создать теоретические основы формирования семейств материнских и дочерних непрерывных саундлетов, характеризующих образцы вокальных звуков.
3. Формализовать саундлетные отображения, действующие между семействами образцов и саундлетов, результатом которых является образец, находящийся в заданном амплитудно-временном окне.
4. Разработать метод формирования эталонных образцов на основе семейства непрерывных саундлетов и саундлетных отображений.
5. Разработать модель метрического распознавания вокальных звуков на основе семейства непрерывных саундлетов и саундлетных отображений и эталонных образцов.
6. Создать критерии оценки эффективности модели.
7. Формализовать условия распознавания вокального звука по эталонным образцам на основе семейства непрерывных саундлетов и саундлетных отображений для оценивания результатов распознавания
8. Разработать логико-формальные правила для оценивания результатов метрического распознавания по модели.

## 1. МЕТОД СОЗДАНИЯ ОБРАЗЦОВ ВОКАЛЬНЫХ ЗВУКОВ

Образцом вокального звука речи назовем участок вокального звука в речевом сигнале, расположенный между соседними пиковыми значениями амплитуды, длина которого соответствует квазипериоду сигнала.

При формировании образца в режиме обучения экспертом вводится левая и правая границы $T^l, T^r$ вокального звука в сигнале $g$, а в режиме распознавания автоматически определяется (на основе энергий последовательно идущих участков сигнала равной длины) левая и правая границы $T^l, T^r$ вокальной части сигнала $g$.

После задания или вычисления границ $T^l, T^r$ на интервале $[T^l, T^r]$ сигнала $g$ вычисляется функции автокорреляции, с помощью которой определяется длина периода основного тона $T^{FT}$ вокального звука.

Для формирования образца как структурообразующего элемента вокального звука интервал $[T^l, T^r]$ сигнала $g$ разбивается на участки на основе вычисленной длины периода основного тона $T^{FT}$ согласно следующему правилу:

$$T_0^{\max} = \arg\max_t g(t),\ t \in [T^l - 0.5 \cdot T^{FT}, T^l + 0.5 \cdot T^{FT}]$$

$$T_{i-1}^{\max} \le T^r \Rightarrow \left(T_i^{\min} = T_{i-1}^{\max}\right) \wedge \left(T_i^{\max} = \arg\max_t g(t)\right)$$
$$t \in [T_i^{\min} + 0.5 \cdot T^{FT}, T_i^{\min} + 1.5 \cdot T^{FT}]$$

На основе этого разбиения формируется конечная совокупность образцов, описываемых множеством вещественнозначных ограниченных финитных непрерывных функций $\{\varphi_i \mid i \in \{1,..., I\}\}$ в виде

$$\varphi_i(t) = \begin{cases} 0, & t \le T_i^{\min} - \Delta t \\ g(T^{\min})\left(\dfrac{t - (T^{\min} - \Delta t)}{\Delta t}\right), & t \in [T_i^{\min} - \Delta t, T_i^{\min}] \\ g(t), & t \in [T_i^{\min}, T_i^{\max}] \\ g(T^{\max})\left(\dfrac{(T^{\max} + \Delta t) - t}{\Delta t}\right), & t \in [T_i^{\max}, T_i^{\max} + \Delta t] \\ 0, & t \ge T_i^{\max} + \Delta t \end{cases}, \quad i \in \{1,..., I\},$$

$$A_i^{\min} = \min_t g(t),\ t \in [T_i^{\min}, T_i^{\max}],\ i \in \{1,..., I\},$$

$$A_i^{\max} = \max_t g(t),\ t \in [T_i^{\min}, T_i^{\max}],\ i \in \{1,..., I\},$$

где параметр $\Delta t \in (0,1)$ задан оператором.

На рис.1. представлен пример разбиения звука «о» слова «со» на образцы, при этом $T^{FT} = 5.8,\ T^l = 214, T^r = 396,\ \Delta t = 1/22050.$

Для дальнейшего сопоставления образцов между собой при формировании эталонных образцов и распознавании по ним тестовых образцов необходимо привести их к единообразию (т.е. к единому прямоугольному амплитудно-временному окну, в которое точно вписана только та часть образца, которая находится на компактном носителе). Для этого в статье разрабатываются теоретические основы материнского и дочернего саундлетов.

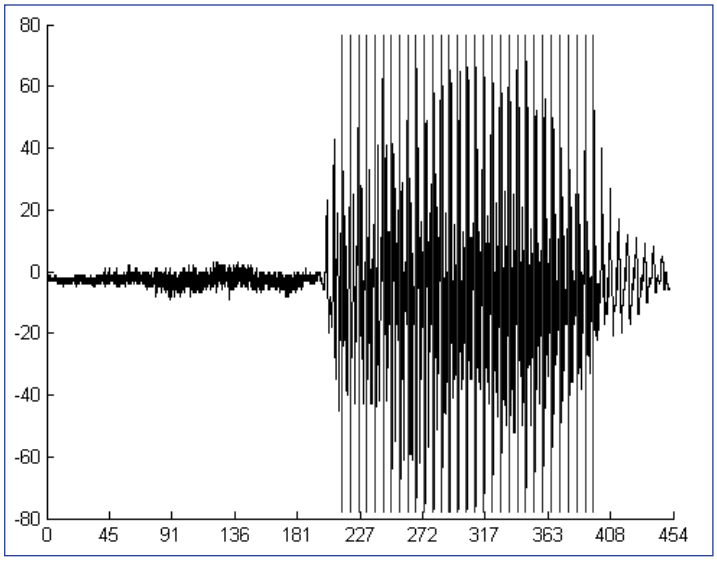


*Рис.1. Разбиение звука «о» слова «со» на образцы*

## 2. СОЗДАНИЕ СЕМЕЙСТВА МАТЕРИНСКИХ НЕПРЕРЫВНЫХ САУНДЛЕТОВ

Материнским непрерывным саундлетом образца вокального звука речи назовем образец, сдвинутый по времени и амплитуде в левый нижний угол положительной плоскости.

Материнский непрерывный саундлет образца вокального звука речи представлен в виде вещественнозначной ограниченной финитной непрерывной функции

$$\psi^m(t) = (G0\varphi)(t) = \begin{cases} 0, & t \le -\Delta t \\ (\varphi(b0) - d0)\left(\dfrac{t+\Delta t}{\Delta t}\right), & t \in [-\Delta t, 0] \\ \varphi(t+b0) - d0, & t \in [0, T] \\ (\varphi(T+b0) - d0)\left(\dfrac{(T+\Delta t)-t}{\Delta t}\right), & t \in [T, T+\Delta t] \\ 0, & t \ge T+\Delta t \end{cases}$$

$$b0 = T^{\min}, \quad d0 = A^{\min}, \quad T = T^{\max} - T^{\min}, \quad A = A^{\max} - A^{\min},$$

где параметр $\Delta t \in (0,1)$ задан оператором,

$G0$ – преобразование, переводящее образец в материнский саундлет,

$b0,\ d0$ – параметры сдвига функции $\varphi$ по времени и амплитуде,

$A^{\min}, A^{\max}$ – минимальное и максимальное значение функции $\varphi$ на компакте $[T^{\min}, T^{\max}]$.

Таким образом, часть материнского саундлета, которая находится на компактном носителе $[-\Delta t, T + \Delta t]$, точно вписана в амплитудно-временное окно высотой $A$ и шириной $T + 2\Delta t$.

Определим конечное семейство материнских непрерывных саундлетов образцов вокального звука речи как $\Psi^m = \{\psi^m\}$, причем все функции $\psi^m$ ограничены снизу и сверху числами 0 и $A$ соответственно.

На рис.2 представлен образец вокального звука «о» при $T^{\min} = 331.5$, $T^{\max} = 339.3$, $A^{\min} = -59$, $A^{\max} = 54$, $\Delta t = 1/22050$ на рис. 3 представлен материнский саундлет вокального звука «о» при $T = 7.8$, $A = 113$.

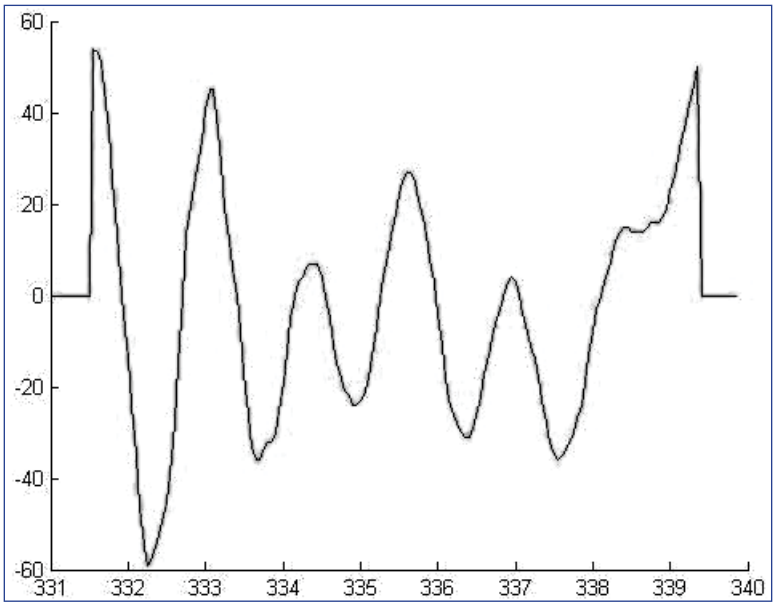


*Рис.2. Образец вокального звука «о»*

От материнского саундлета породим дочерний саундлет, описывающий образец вокального звука речи, который находится в заданном амплитудно-временном окне.

## 3. СОЗДАНИЕ СЕМЕЙСТВА ДОЧЕРНИХ НЕПРЕРЫВНЫХ САУНДЛЕТОВ

Дочерним саундлетом назовем сдвинутый и масштабированный по времени и амплитуде материнский саундлет.

Дочерний саундлет представлен в виде вещественнозначной ограниченной финитной непрерывной функции

$$\psi^c(t) = (G1\psi^m)(t) = \begin{cases} 0, & t \le \breve{T}^{\min} - \Delta t \\ \left(d + c\psi^m\left(\dfrac{\breve{T}^{\min} - b}{a}\right)\right)\left(\dfrac{t - (\breve{T}^{\min} - \Delta t)}{\Delta t}\right), & t \in [\breve{T}^{\min} - \Delta t, \breve{T}^{\min}] \\ d + c\psi^m\left(\dfrac{t-b}{a}\right), & t \in [\breve{T}^{\min}, \breve{T}^{\max}] \\ \left(d + c\psi^m\left(\dfrac{\breve{T}^{\max} - b}{a}\right)\right)\left(\dfrac{(\breve{T}^{\max} + \Delta t) - t}{\Delta t}\right), & t \in [\breve{T}^{\max}, \breve{T}^{\max} + \Delta t] \\ 0, & t \ge \breve{T}^{\max} + \Delta t \end{cases},$$

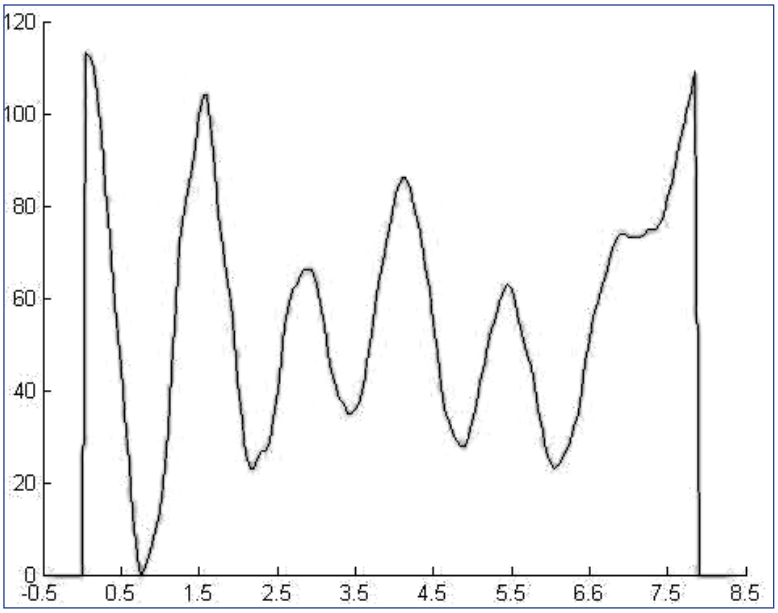


*Рис.3. Материнский саундлет вокального звука «о»*

$$a = \frac{\breve{T}^{\max} - \breve{T}^{\min}}{T}, \quad b = \breve{T}^{\min}, \quad c = \frac{\breve{A}^{\max} - \breve{A}^{\min}}{\tilde{A}^{\max} - \tilde{A}^{\min}}, \quad d = \breve{A}^{\min}$$

$$\tilde{A}^{\max} = \max_t \psi^m\left(\frac{t-b}{a}\right), \quad \tilde{A}^{\min} = \min_t \psi^m\left(\frac{t-b}{a}\right), \quad t \in [\breve{T}^{\min}, \breve{T}^{\max}],$$

где $G1$ – преобразование, переводящее материнский саундлет в дочерний саундлет,

$a$, $c$ – параметры масштабирования функции $\psi^m$ по времени и амплитуде,

$b$, $d$ – параметры сдвига функции $\psi^m$ по времени и амплитуде,

$\breve{A}^{\max}$, $\breve{A}^{\min}$ – заданное минимальное и максимальное значение функции $\psi^c$ на компакте $[\breve{T}^{\min}, \breve{T}^{\max}]$;

Таким образом, часть дочернего саундлета, которая находится на компактном носителе $[\breve{T}^{\min} - \Delta t, \breve{T}^{\max} + \Delta t]$, точно вписана в амплитудно-временное окно высотой $\breve{A}^{\max} - \breve{A}^{\min}$ и шириной $\breve{T}^{\max} - \breve{T}^{\min} + 2\Delta t$.

Определим конечное семейство дочерних непрерывных саундлетов образцов вокального звука речи как $\Psi^c = \{\psi^c\}$, причем все функции $\Psi^c$ имеют одинаковый компактный носитель $[\breve{T}^{\min} - \Delta t, \breve{T}^{\max} + \Delta t]$ и одинаковые минимальные и максимальные значения $\breve{A}^{\min}$, $\breve{A}^{\max}$ на нем.

На рис. 4 представлен дочерний саундлет вокального звука «о» при $a = 0.872$, $b = 6966$, $c = 1.268$, d = –73.

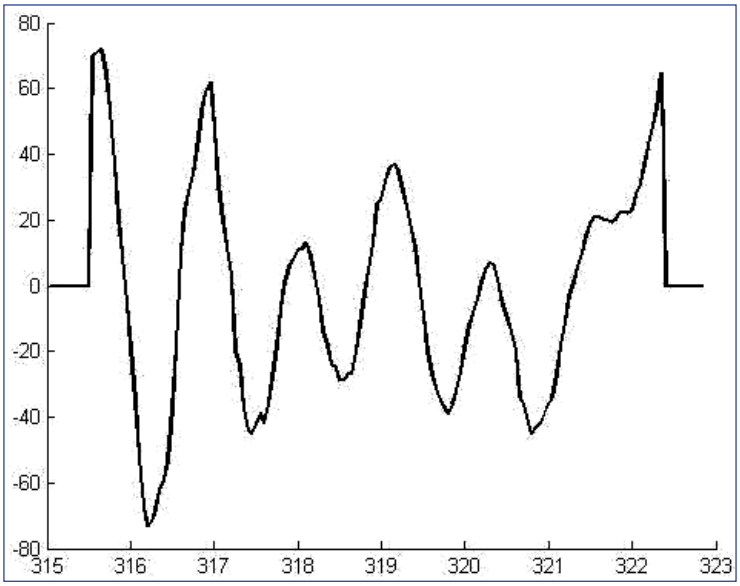


*Рис.4. Дочерний саундлет вокального звука «о»*

Для преобразования образца с целью приведения его к единообразию (одинаковому амплитудно-временному окну) формализуем отображения между образцами, материнскими саундлетами и дочерними саундлетами.

## 4. ФОРМАЛИЗАЦИЯ САУНДЛЕТНЫХ ОТОБРАЖЕНИЙ

Саундлетным отображением назовем преобразование, переводящее образец в материнский саундлет, а материнский саундлет в дочерний саундлет путем сдвига и масштабирования по времени и амплитуде.

Преобразование $G0$, введенное в пункте 2 и осуществляющее сдвиг функции $\varphi$, описывающей образец, по времени и амплитуде в левый нижний угол положительной плоскости, для получения материнского саундлета $\psi^m$, представимо в виде саундлетного отображения

$$G0 : \Phi \rightarrow \Psi^m.$$

Преобразование $G1$, введенное в пункте 3 и осуществляющее сдвиг и масштабирование материнского непрерывного саундлета $\psi^m$ по времени и амплитуде для получения дочернего саундлета $\Psi^c$, представимо в виде саундлетного отображения

$$G1 : \Psi^m \rightarrow \Psi^c.$$

Пусть метрика определена в виде

$$\rho_p(\widehat{\psi}, \breve{\psi}) = \sqrt[p]{\int_R |\widehat{\psi}(t) - \breve{\psi}(t)|^p \, dt}.$$

Если $a \leq 1$ и $c$ (или $c \leq 1$ и $a$) фиксировано, то отображение $G1$ является нерастягивающим по времени (или по амплитуде), т.е. удовлетворяет условию

$$\forall \widehat{\psi}^m, \breve{\psi}^m \in \Psi^m \; \rho_p(\widehat{\psi}^m, \breve{\psi}^m) \geq \rho_p(G1\widehat{\psi}^m, G1\breve{\psi}^m).$$

Если $a \geq 1$ и $c$ фиксировано (или $c \geq 1$ и $a$ фиксировано), то отображение $G1$ является несжимающим по времени (или по амплитуде), т.е. удовлетворяет условию

$$\forall \widehat{\psi}^m, \breve{\psi}^m \in \Psi^m \; \rho_p(\widehat{\psi}^m, \breve{\psi}^m) \leq \rho_p(G1\widehat{\psi}^m, G1\breve{\psi}^m).$$

Композиция преобразований $G0$ и $G1$ представлена в виде

$$G = G1G0.$$

Таким образом, преобразование, осуществляющее переход от функции $\varphi$, описывающей образец, к дочернему непрерывному саундлету $\Psi^c$, представимо в виде саундлетного отображения

$$G : \Phi \to \Psi^c.$$

На саундлетное отображение $G$ накладываются следующие ограничения:

1. Совпадение компактного носителя у всех дочерних саундлетов

$$\forall\, \widehat{\varphi} \in \Phi \cdot\ \forall\, \breve{\varphi} \in \Phi \cdot\ \operatorname{supp} G\widehat{\varphi} = \operatorname{supp} G\breve{\varphi}$$

2. Совпадение минимальных и максимальных значений на компактном носителе у всех дочерних саундлетов

$$\forall\, \widehat{\varphi} \in \Phi \cdot\ \forall\, \breve{\varphi} \in \Phi \cdot \left( \min_{t \in \operatorname{supp} G\widehat{\varphi}} (G\widehat{\varphi})(t) = \min_{t \in \operatorname{supp} G\breve{\varphi}} (G\breve{\varphi})(t) \right) \wedge$$

$$\wedge \left( \max_{t \in \operatorname{supp}\ G\widehat{\varphi}} (G\widehat{\varphi})(t) = \max_{t \in \operatorname{supp}\ G\breve{\varphi}} (G\breve{\varphi})(t) \right).$$

Ограничения 1-2 обеспечивают единое прямоугольное амплитудно-временное окно для всех полученных дочерних саундлетов, в которое точно вписана только та часть этих саундлетов, которая находится на компактном носителе.

На основе введенных семейств саундлетов и саундлетных отображений сформируем эталонные образцы вокальных звуков речи.

## 5. МЕТОД ФОРМИРОВАНИЯ ЭТАЛОННЫХ ОБРАЗЦОВ

Пусть дана конечная совокупность обучающих образцов вокального звука, которая описывается множеством вещественных ограниченных финитных непрерывных функций $\Phi = \{\varphi_i \mid i \in \{1,..., I\}\}$, причем $A_i^{\min}, A_i^{\max}$ – минимальное и максимальное значение функции $\varphi_i$ на компакте $[T_i^{\min}, T_i^{\max}]$.

Для сопоставления элементов множества $\Phi$ между собой для каждой функции $\varphi_i$, описывающей обучающий образец, формируется соответствующее ему конечное множество дочерних непрерывных саундлетов $\Psi^c$, находящихся в том же самом амплитудно-временном окне, что и эта функция, в виде

$$\forall\, \varphi_i \in \Phi\ \ \exists \Psi^c = \{\psi_r^c \mid r \in \{1,..., I\}\} : \psi_r^c = G\varphi_r.$$

Вычисляется нормированное расстояние между функцией, описывающей обучающий образец, и дочерним непрерывным саундлетом в виде

$$\forall\, i, r \in \{1,...,\ I\}\ \ d_{ir} = \frac{\rho_p(\varphi_i, \psi_r^c)}{(A_i^{\max} - A_i^{\min}) \sqrt[p]{(T_i^{\max} - T_i^{\min})}},$$

$$\rho_p(\varphi_i, \psi_r^c) = \sqrt[p]{\int_R |\varphi_i(t) - \psi_r^c(t)|^p\ dt}.$$

Осуществляется выбор множества функций $\Gamma$, описывающих эталонные образцы, из множества функций $\Phi$, описывающих обучающие образцы, на основе матрицы нормированных расстояний $[d_{ir}]$. Для этого в статье предложена следующая процедура.

## 6. ПРОЦЕДУРА ВЫБОРА ПОДМНОЖЕСТВА ЭТАЛОННЫХ ОБРАЗЦОВ ИЗ МНОЖЕСТВА ОБУЧАЮЩИХ ОБРАЗЦОВ

Приведем этапы процедуры выбора подмножества эталонных образцов из множества обучающих образцов на основе матрицы $[d_{ir}]$

1. Создать точечно конечное покрытие $C$ множества номеров обучающих образцов $B0 = \{1,..., I\}$ в виде

1.1. $C = \{C_i\},\ C_i = \{r \mid d_{ir} < \varepsilon, r \in B0\},\ i \in B0,\ 0 < \varepsilon \le 1$.

1.2. $\forall i \in B0\ \ |C_i| < \delta \Rightarrow C_i = \{i\},\ 1 < \delta < I$,

причем $\varepsilon$, $\delta$ задаются экспертом.

2. Создать множество $B1$ из номеров элементов покрытия в виде

$$B1 = \{i \parallel C_i \mid > 1\}.$$

3. Создать множество $B2$ из элементов множества $B1$ в виде

$$B2 = \left\{ i(i \in B1) \wedge \left( \wedge_{m \in B1,\ i \ne m} (C_i \ne C_m \wedge C_i \not\subset C_m) \vee \left( C_i = C_m \wedge \sum_{z \in C_i} d_{iz} > \sum_{z \in C_m} d_{mz} \right) \right) \right\}.$$

4. Создать множество $B3$ из элементов множества $B2$

4.1. $i = 2, E1 = \varnothing$.

4.2. Если $j = b2_i \wedge C_j \subset \bigcup_{m \in V} C_m$, то $E1 = E1 \cup \{j\}$, где $V = (B2 \cap E1) \setminus \{j\}$.

4.3. Если $i < |B2|$, то $i = i + 1$, переход к шагу 2.4.2, иначе $B3 = B2 \setminus E1$.

5. Создать конечное множество эталонных образцов $\Gamma$

$$\Gamma = \{\gamma_k \mid \gamma_k = \varphi_{i_k},\ i_k \in B4\},\ B4 = B3 \cup \left( B0 \setminus \bigcup_{m \in B3} C_m \right),$$

6. Создать подпокрытие $\widetilde{C}$

$$\widetilde{C} = \{\widetilde{C}_k \mid \widetilde{C}_k = C_{i_k},\ i_k \in B4\},\ \widetilde{C} \subset C.$$

7. Создать вектор весов эталонных образцов

$$w = (w_1,..., w_k,..., w_{|\widetilde{C}|}),\ w_k = |\widetilde{C}_k| \Big/ \sum_{k=1}^{|\widetilde{C}|} |\widetilde{C}_k|,\ k \in \{1,..., |\widetilde{C}|\}.$$

На основе введенных семейств саундлетов и саундлетных отображений и сформированного множества эталонных образцов создадим модель метрического распознавания вокальных звуков.

## 7. МОДЕЛЬ МЕТРИЧЕСКОГО РАСПОЗНАВАНИЯ ВОКАЛЬНЫХ ЗВУКОВ

Дадим общую математическую постановку задачи распознавания, которая может служить основой для построения моделей метрического распознавания. Пусть $\varphi$ – функция, описывающая подлежащий распознаванию образец, $y$ – номер класса образца (номер класса вокального звука речи). Задача заключается в том, чтобы по значению $\varphi$ определить значение величины $y$. Тогда построение модели метрического распознавания сводится к определению зависимости между номером класса образцов $y$ от значения $x$ на основе метрики.

Модель метрического распознавания вокальных звуков представлена в виде

$$y = \arg\min_j \theta(\varphi, \Gamma_j), \quad \tilde{y} = \min_j \theta(\varphi, \Gamma_j), \quad j \in \{1,..., J\}, \quad \Gamma_j = \{\gamma_{jk}\},$$

$$\theta(\varphi, \Gamma_j) = \min_k \frac{(1 - w_{jk})\rho_p(\gamma_{jk}, G\varphi)}{(A_{jk}^{\max} - A_{jk}^{\min})\sqrt[p]{(T_{jk}^{\max} - T_{jk}^{\min})}}, \quad k \in \{1,..., K_j\}, \quad j \in \{1,..., J\},$$

$$\rho_p(\gamma_{jk}, G\varphi) = \sqrt[p]{\int_R |\gamma_{jk}(t) - (G\varphi)(t)|^p \, dt},$$

где $y$ – номер звука,

$\tilde{y}$ – расстояние между тестовым образцом $\varphi$ и множеством эталонных образцов всех вокальных звуков $\{\Gamma j\}$,

$\varphi$ – вещественнозначная ограниченная финитная непрерывная функция, описывающая тестовый образец непрерывного речевого сигнала,

$\gamma_{jk}$ – вещественнозначная ограниченная финитная непрерывная функция, описывающая $k$-й эталонный образец $j$-го звука,

$A^{\min}$, $A^{\max}$ – минимальное и максимальное значение функции $\varphi$ на компакте $[T^{\min}, T^{\max}]$,

$A_{jk}^{\min}$, $A_{jk}^{\max}$ – минимальное и максимальное значение функции $\gamma_{jk}$ на компакте $[T_{jk}^{\min}, T_{jk}^{\max}]$,

$J$ – количество звуков,

$K_j$ – количество эталонных образцов $j$-го звука,

$w_{jk}$ – вес $k$-го эталонного образца $j$-го звука, $w_{jk} \in [0,1]$. Если вес не учитывается, то $w_{jk} = 0$.

Для созданной модели сформулируем критерии эффективности.

## 8. КРИТЕРИИ ОЦЕНКИ ЭФФЕКТИВНОСТИ МОДЕЛИ

1. *Критерий скорости распознавания* означает выбор из заданного набора метрик такой метрики, которая *на стадии обучения модели* требует наименьшего количества эталонных образцов

$$F = T \to \min_p.$$

2. *Критерий оценки пороговой вероятности распознавания* означает выбор такого множества эталонных образцов *на стадии опытной эксплуатации модели*, чтобы для тестового образца номер звука, вычисленный по модели, совпадал с тестовым номером звука этого тестового образца

$$F = \frac{1}{I}\sum_{i=1}^{I} \phi(y_i^{model}, y_i^{test}) \to \max_{\{\Gamma_j\}},$$

$$\phi(a,b) = \begin{cases} 1, & a = b, \\ 0, & a \neq b \end{cases},$$

$$y_i^{model} = \arg\min_j \theta(\varphi_i, \Gamma_j), \; j \in \{1,..., J\},$$

где $\varphi_i$ – $i$-е тестовые образцы,

$y_i^{test}$ – тестовый номер звука для $i$-го тестового образца,

$I$ – количество тестовых образцов.

3. *Для оценки готовности модели к эксплуатации* используется *критерий адекватности модели*, основанный на минимуме среднеквадратичной ошибки

$$F = \frac{1}{I}\sum_{i=1}^{I}(\tilde{y}_i^{model} - \tilde{y}_i^{test})^2 \to \min_{\{\Gamma_j\}},$$

$$\tilde{y}_i^{model} = \min_j \theta(\varphi_i, \Gamma_j),\ j \in \{1,...,\ J\},$$

где $\varphi_i$ – $i$-е тестовый образец,

$y_i^{test}$ – тестовое расстояние для $i$-го тестового образца,

$I$ – количество тестовых образцов.

Для оценивания результатов распознавания вокальных звуков необходимо сформулировать условия их распознавания.

## 9. УСЛОВИЯ РАСПОЗНАВАНИЯ ТЕСТОВОГО ОБРАЗЦА ВОКАЛЬНОГО ЗВУКА ПО ЭТАЛОННЫМ ОБРАЗЦАМ

Пусть дан тестовый образец вокального звука, который описывается вещественнозначной ограниченной финитной непрерывной функцией $\varphi$.

Пусть для каждого $j$-го вокального звука дано множество эталонных образцов $\Gamma_j,\ j \in \{1,...,\ J\}$.

Пусть для каждого $j$-го вокального звука вычислено расстояние $\theta(\varphi, \Gamma_j)$ между функцией $x$, описывающей тестовый образец, и множеством функций $\Gamma_j$, описывающих эталонные образцы $j$-го звука.

*Необходимое условие распознавания тестового образца*. Тестовый образец распознан, если

$$\forall n \in \{1,...,\ J\}\ \forall m \in \{1,...,\ J\}$$

$$\left(\theta(\varphi, \Gamma_n) = \min_j \theta(\varphi, \Gamma_j)\right) \wedge \left(\theta(\varphi, \Gamma_m) = \min_j \theta(\varphi, \Gamma_j)\right) \to (n = m) \wedge (\theta(\varphi, \Gamma_n) < \tilde{\varepsilon}),\ j \in \{1,...,\ J\},$$

где $\tilde{\varepsilon}$ – заданная точность распознавания, $0 < \tilde{\varepsilon} \le 1$.

*Достаточное условие распознавания тестового образца*. Тестовый образец распознан, если

$$\forall n \in \{1,...,\ J\}\ \forall m \in \{1,...,\ J\}$$

$$\left(\theta(\varphi, \Gamma_n) = \min_j \theta(\varphi, \Gamma_j)\right) \wedge \left(\theta(\varphi, \Gamma_m) = \min_j \theta(\varphi, \Gamma_j)\right) \to (n = m) \wedge (\theta(\varphi, \Gamma_n) = 0),\ j \in \{1,...,\ J\}.$$

На основе полученных условий возможно сформировать логико-формальные правила оценивания результатов распознавания.

## 10. ЛОГИКО-ФОРМАЛЬНЫЕ ПРАВИЛА ОЦЕНИВАНИЯ РЕЗУЛЬТАТА МЕТРИЧЕСКОГО РАСПОЗНАВАНИЯ

Для оценивания результатов распознавания формируются следующие логико-формальные правила

Если $\tilde{y} < \tilde{\varepsilon}$, то $q = y$,

Если $\tilde{y} \ge \tilde{\varepsilon}$, то $q = 0$,

где $q$ – номер звука,

$\tilde{\varepsilon}$ – это численно вычисленное расстояние между множеством функций, описывающих эталонные образцы вокальных звуков, и множеством порожденных непрерывных саундлетов тестовых образцов невокальных звуков.

## 11. ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ МЕТРИЧЕСКОГО МЕТОДА РАСПОЗНАВАНИЯ ВОКАЛЬНЫХ ЗВУКОВ

В табл. 1 приведено сравнение предложенного метода и существующих метрических методов на основе базы данных TIMIT, при этом для авторского метода непрерывный сигнал создавался из дискретного на основе линейного сплайна с равноотстоящими узлами. Распознаванию подлежали все вокальные звуки. В неавторских методах в качестве образцов брались вектора мел-частотные кепстральные коэффициенты (MFCC), вычисленные на участках равной длины, т.е. фреймах. Ошибка распознавания представляет собой отношение количества правильно распознанных образцов, содержащих вокальные звуки, к их общему количеству в процентах, при этом образцы, содержащие конец первого вокальные звука и начало вокальные второго звука, не учитывались. Приведенные в табл.1 стандартные метрические методы были реализованы автором статьи, посредством пакета Matlab. Исследование позволяет сделать вывод, что авторский метод обеспечивает высокую вероятность распознавания.

*Таблица 1*

Оценка метрических методов распознавания

| Методы метрического распознавания | Ошибка распознавания (%) |
|---|---|
| на основе кодовой книги | 30 |
| на основе алгоритма DTW | 8 |
| авторский метод | 5 |

## ВЫВОДЫ

*Новизна.* В работе впервые излагаются теоретические основы саундлетов и саундлетных отображений. Усовершенствован метрический подход к распознаванию вокальных звуков, который отличается тем, что позволяет учитывать квазипериодическую структуру вокальных звуков и обобщать образцы одного звука различной длины и различным размахом амплитуд, что повышает эффективность распознавания вокальных звуков речи. Получил дальнейшее развитие метод создания множества эталонных образцов, который отличается тем, что основан на семейства непрерывных саундлетов и саундлетных отображений, что повышает эффективность процедуры формирования эталонных образцов. В рамках предложенных саундлетов и саундлетных отображений усовершенствована модель распознавания вокальных звуков, которая отличается тем, что позволяет сопоставлять образцы различной длины и использовать адаптивный нормированный порог в логико-формальных правилах, что повышает вероятность распознавания полезных звуков.

*Практическое значение.* Разработан метод построения модели метрического распознавания вокальных звуков на основе семейства непрерывных саундлетов и саундлетных отображений, что позволяет сократить количество эталонных образцов. Предложен адаптивный нормированный порог для логико-формальных правил оценивания распознавания речевых сигналов, который позволяет с большей вероятностью выделять полезные звуки. В результате численного исследования было установлено, что алгоритм метрического распознавания вокальных звуков на основе семейства непрерывных саундлетов и саундлетных отображений дает вероятность распознавания 0.95. Созданные алгоритмы могут использоваться для решения задач, связанных с распознаванием речи оператора, синтезом речи, анализом вибрационного сигнала.

## СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. *Rabiner L.R.* Fundamentals of speech recognition / *L.R. Rabiner, B.H. Jang.* – Englewood Cliffs, NJ: Prentice Hall PTR, 1993. –507 p.

2. *Винцюк Т.К.* Анализ, распознавание и интерпретация речевых сигналов / Т.К. Винцюк. – К.: Наук. думка, 1987. – 261 с.

3. *Потапова Р.К.* Речь: коммуникация, информация, кибернетика / Р.К. Потапова. М.: Радио и Связь, 1997. 528 с.

4. *Осовский С.* Нейронные сети для обработки информации / С. Осовский. – М.: Финансы и статистика, 2002. – 344 с.

5. *Хайкин С.* Нейронные сети: полный курс / С. Хайкин. – М.: Издательский дом «Вильямс», 2006. – 1104 с.

6. *Каллан Р.* Основные концепции нейронных сетей / Р. Каллан. – М.: Издательский дом «Вильямс», 2001. – 288 с.

7. *Федоров Е.Е.* Методология создания мультиагентной системы речевого управления: монография / Е.Е. Федоров. – Донецк: изд-во «Ноулидж», 2011. – 356 с.

### Сведения об авторах

***Фёдоров Евгений Евгениевич,***
заведующий кафедрой специализированных компьютерных систем Донецкой академии автомобильного транспорта, профессор кафедры автоматизированных систем управления Донецкого национального технического университета, доцент.
В 2012 году защитил докторскую диссертацию в Национальном авиационном университете г. Киева. Автор свыше 110 научных публикаций, в том числе 10 монографий, посвящённых: моделям и методам преобразования и распознавания речевых образов; моделям и методам преобразования и распознавания зрительных образов; моделям и методам анализа и синтеза естественно-языковых объектов; моделям и методам вибрационной, шумовой и медицинской диагностики; интеллектуальным технологиям в логистике, метаэвристикам.
Основная область интересов: методы идентификации и верификации диктора; распознавания и синтеза речи; методы анализа и синтеза естественно-языковых объектов; методы распознавания лица человека; методы диагностики состояния электромеханических объектов по вибрационному и шумовому сигналам; методы диагностики состояния пациентов по электрограммам; эвристические и метаэвристические методы решения оптимизационных задач транспортной логистики; метаэвристические методы оценивания значений параметров моделей распознавания, диагностики и прогноза.

***Эгле Слесорайтите,***
старший преподаватель Вильнюсского университета.
Автор нескольких десятков научных публикаций, посвящённых: моделям и методам преобразования и распознавания речевых образов; моделям и методам преобразования и распознавания зрительных образов; интеллектуальным технологиям в логистике, метаэвристикам.
Основная область интересов: идентификация и верификация диктора, распознавание и синтез речи, распознавание лица человека, интеллектуальные технологии в транспортной логистике (поиск оптимального маршрута и мультиагентное взаимодействие), оптимизация числовых функций и комбинаторная оптимизация на основе метаэвристик.