УДК 004

И.И. Федоров (докт. техн. наук, доц.)
Донецкая академия автомобильного транспорта, г.Донецк
кафедра специализированных компьютерных систем
E-mail: fee75@mail.ru

МЕТОД СИНТЕЗА ВОКАЛЬНЫХ ЗВУКОВ РЕЧИ ПО ЭТАЛОННЫМ ОБРАЗЦАМ НА ОСНОВЕ САУНДЛЕТОВ

В статье изложен метод синтеза вокальных звуков речи по эталонным образцам на основе саундлетов. Использованы материнский и дочерний дискретные и 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 синтеза вокальных звуков по эталонным образцам, которые используются в интеллектуальных системах общения и позволяют сократить объем хранимой информации.
Ключевые слова: *дискретный саундлет, непрерывный саундлет, материнский саундлет, дочерний саундлет, эталонные образцы вокальных звуков, синтез звуков.*

Общая постановка проблемы

В современных условиях актуальной является разработка интеллектуальных процессоров предназначенных для распознавания речи человека, синтеза речи и др., и используемых в компьютерных системах общения.

В корне данной задачи лежит проблема построения эффективных методов, обеспечивающих большую скорость обучения модели синтеза, а также большую адекватность синтеза речевых сигналов.

Анализ исследований

Существующие системы синтеза речевых образов используют такие подходы как [1-4]: формантный синтез, синтез на основе коэффициентов линейного прогноза (КЛП-синтез) и конкатенативный синтез. Формантный синтез и КЛП-синтез опираются на модель речеобразования человека. Модель речевого тракта реализуется в виде адаптивного цифрового фильтра. Для формантного синтеза параметры адаптивного цифрового фильтра определяются формантными частотами [5-6], а для КЛП-синтеза – КЛП-коэффициентами [7]. Лучшие результаты в отношении разборчивости и натуральности звучания речи удается получить с помощью конкатенативного синтеза. Конкатенативный синтез осуществляется путем склейки нужных звуковых единиц [1,3,8,9]. В таких системах необходимо применять обработку сигнала для приведения частоты основного тона, энергии и длительности звуковых единиц к тем, которыми должна характеризоваться синтезируемая речь. В системах конкатенативного синтеза применяются три основных алгоритма: TD-PSOLA (осуществляется масштабирование звуковой единицы по времени), FD-PSOLA (осуществляется масштабирование звуковой единицы по частоте), LP-PSOLA (осуществляется масштабирование сигнала ошибки прогноза по времени с последующим применением фильтра с КЛП-коэффициентами). Недостатком конкатенативного синтеза является необходимость хранения большого числа звуковых единиц. В связи с чем возникает задача их более экономного представления [10].

Постановка задач исследования

Целью работы является разработка метода синтеза вокальных звуков речи, базирующегося на саундлетах и формируемых на их основе саундлетных отображениях.

Решение задач и результаты исследований

Для достижения поставленной цели необходимо:

1. Разработать метод создания семейства образцов вокальных звуков.
2. Сформировать семейства материнских и дочерних 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 дискретных саундлетов и саундлетных отображений.
5. Разработать модель синтеза вокальных звуков по эталонным образцам на основе семейства саундлетов и саундлетных отображений.
6. Создать критерии оценки эффективности модели.
7. Формализовать условия синтеза вокального звука по эталонным образцам на основе семейства саундлетов и саундлетных отображений для оценивания результатов синтеза.
8. Разработать логико-формальные правила для оценивания результатов синтеза по модели.

Метод создания семейства образцов вокальных звуков

Образцом вокального звука речи назовем участок вокального звука в речевом сигнале, расположенный между соседними пиковыми значениями и имеющий длину соответствующую квазипериоду.

При формировании образца экспертом вводится левая и правая границы N^l, N^r вокального звука в сигнале f.

После задания границ N^l, N^r на множестве $\{N^l,...,N^r\}$ сигнала f вычисляется функции автокорреляции, с помощью которой определяется длина периода основного тона N^{FT} вокального звука.

Для формирования образца как структурообразующего элемента вокального звука множество $\{N^l,...,N^r\}$ сигнала f разбивается на участки на основе вычисленной длины периода основного тона N^{FT} согласно следующему правилу

$$N_0^{\max} = \arg\max_n f(n),\ n \in \{N^l - 0.5 \cdot N^{FT},..., N^l + 0.5 \cdot N^{FT}\},$$

$$N_{i-1}^{\max} \le N^r \Rightarrow \left(N_i^{\min} = N_{i-1}^{\max}\right) \wedge \left(N_i^{\max} = \arg\max_n f(n)\right),$$

$$n \in [N_i^{\min} + 0.5 \cdot N^{FT}, N_i^{\min} + 1.5 \cdot N^{FT}].$$

На основе этого разбиения формируется конечное семейство образцов, описываемых множеством целочисленных ограниченных финитных дискретных функций $X = \{x_i \mid i \in \{1,...,I\}\}$, в виде

$$x_i(n) = \begin{cases} f(n), & n \in \{N_i^{\min},..., N_i^{\max}\} \\ 0, & n \notin \{N_i^{\min},..., N_i^{\max}\} \end{cases},\ i \in \{1,...,I\},$$

$$A_i^{\min} = \min_n f(n),\ n \in \{N_i^{\min},..., N_i^{\max}\},\ i \in \{1,...,I\},$$

$$A_i^{\max} = \max_n f(n),\ n \in \{N_i^{\min},..., N_i^{\max}\},\ i \in \{1,...,I\}.$$

Для дальнейшего сопоставления образцов между собой при формировании эталонных 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 материнский и дочерний саундлеты.

Создание семейства материнских дискретных саундлетов

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

Материнский дискретный саундлет образца вокального звука речи представлен в виде целочисленной ограниченной финитной дискретной функции

$$s^m(n) = F0x = \begin{cases} x(n+b0) - d0, & n \in \{0,...,N\} \\ 0, & n \notin \{0,...,N\} \end{cases},$$

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

где $F0$ – преобразование, переводящее образец в материнский саундлет,

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

$A^{\min}, A^{\max}$ – минимальное и максимальное значение функции x на компакте $\{N^{\min},...,N^{\max}\}$.

Таким образом, часть материнского саундлета, которая находится на компактном носителе $\{0,...,N\}$, точно вписана в амплитудно-временное окно высотой A и шириной N.

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

От материнского дискретного саундлета породим материнский непрерывный саундлет.

Создание семейства материнских непрерывных саундлетов

Материнский непрерывный саундлет ψ^m получен из материнского дискретного саундлета s^m на основе кусочно-линейного сплайна с равноотстоящими узлами.

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

$$\psi^m(t) = F1s^m = \begin{cases} \sum_{n=-1}^{N} \chi_{(t_n,t_{n+1})}(t)\left(s^m(n) + \frac{s^m(n+1)-s^m(n)}{\Delta t}(t-t_n)\right) + \sum_{n=-1}^{N+1} \chi_{\{t_n\}}(t)s^m(n), & t \in [-\Delta t, T+\Delta t] \\ 0, & t \notin [-\Delta t, T+\Delta t] \end{cases},$$

$$T = N\Delta t,\ t_n = n\Delta t,$$

$$\chi_B(t) = \begin{cases} 1, & t \in B \\ 0, & t \notin B \end{cases};$$

где Δt – шаг квантования по времени.

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 Ψ^m ограничены снизу и сверху числами 0 и A соответственно.

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

Создание семейства дочерних непрерывных саундлетов

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

Дочерний непрерывный саундлет представлен в виде вещественнозначной ограниченной финитной непрерывной функции

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

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

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

где a, c – параметры масштабирования функции ψ^m по времени и амплитуде,

b, d – параметры сдвига функции ψ^m по времени и амплитуде,

$\breve{A}^{\min}, \breve{A}^{\max}$ – заданное минимальное и максимальное значение функции ψ^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 заданное 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 Ψ^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.

От дочернего непрерывного саундлета породим дочерний дискретный саундлет.

Создание семейства дочерних дискретных саундлетов

Дочерний дискретный саундлет s^c получен из дочернего непрерывного саундлета ψ^c путем дискретизации.

Дочерний дискретный саундлет представлен в виде целочисленной ограниченной финитной дискретной функции

$$s^c(n) = F2\psi^c = round(\psi^c(n\Delta t)),\ n \in \{\breve{N}^{\min}, ..., \breve{N}^{\max}\},$$

$$\breve{N}^{\min} = \breve{T}^{\min} / \Delta t, .\breve{N}^{\max} = \breve{T}^{\max} / \Delta t,$$

где $round$ – функция, округляющая число до ближайшего целого.

Таким образом, часть дочернего саундлета, которая находится на компактном носителе $\{\breve{N}^{\min}, ..., \breve{N}^{\max}\}$, точно вписана в заданное амплитудно-временное окно высотой $\breve{A}^{\max} - \breve{A}^{\min}$ и шириной $\breve{N}^{\max} - \breve{N}^{\min}$.

Определим конечное семейство дочерних дискретных саундлетов образцов вокального звука речи как $S^c = \{s^c\}$, причем все функции S^c имеют одинаковый компактный носитель $\{\breve{N}^{\min},...,\breve{N}^{\max}\}$ и одинаковые минимальные и максимальные значения $\breve{A}^{\min}$, $\breve{A}^{\max}$ на нем.

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

Формализация саундлетных отображений

Саундлетным отображением назовем преобразование, переводящее образец в материнский дискретный саундлет, материнский дискретный саундлет в материнский непрерывный саундлет, материнский непрерывный саундлет в дочерний непрерывный саундлет, дочерний непрерывный саундлет в дочерний дискретный саундлет путем сплайновой интерполяции, сдвига и масштабирования по времени и амплитуде, дискретизации.

Преобразование $F0$, введенное в пункте 2 и осуществляющее сдвиг функции x, описывающей образец, по времени и амплитуде в левый нижний угол положительной плоскости, для получения материнского дискретного саундлета s^m, представимо в виде саундлетного отображения

$$F0 : X \to S^m.$$

Преобразование $F1$, введенное в пункте 3 и создающее из материнского дискретного саундлета s^m материнский непрерывный саундлет ψ^m (как кусочно-линейный сплайн с равноотстоящими узлами), представимо в виде саундлетного отображения

$$F1 : S^m \to \Psi^m.$$

Преобразование $G1$, введенное в пункте 4 и осуществляющее сдвиг и масштабирование материнского непрерывного саундлета ψ^m по времени и амплитуде для получения дочернего непрерывного саундлета ψ^c, представимо в виде саундлетного отображения

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

Преобразование $F2$, введенное в пункте 5 и создающее путем дискретизации из дочернего непрерывного саундлета ψ^c дочерний дискретный саундлет s^c, представимо в виде саундлетного отображения

$$F2 : \Psi^c \to S^c.$$

Композиция преобразований $F0$, $F1$, $G1$, $F2$ представлена в виде

$$F = F2G1F1F0.$$

Таким образом, преобразование, осуществляющее переход от функции x, описывающей образец, к дочернему дискретному саундлету s^c, представимо в виде саундлетного отображения

$$F : X \to S^c,$$

и обладает следующими свойствами

1. Одинаковый компактный носитель у всех дочерних саундлетов

$$\forall \hat{x} \in X \quad \forall \breve{x} \in X \quad \operatorname{supp} F\hat{x} = \operatorname{supp} F\breve{x}$$

2. Одинаковые минимальные и максимальные значения на компактном носителе у всех дочерних саундлетов

$$\forall \hat{x} \in X \quad \forall \breve{x} \in X \quad \left(\min_{n \in \operatorname{supp} F\hat{x}} (F\hat{x})(n) = \min_{n \in \operatorname{supp} F\breve{x}} (F\breve{x})(n) \right) \wedge$$

$$\wedge\left(\max_{n\in\mathrm{supp}F\hat{x}}(F\hat{x})(n)=\max_{n\in\mathrm{supp}F\breve{x}}(F\breve{x})(n)\right).$$

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.

Метод формирования эталонных образцов

Пусть дана конечная совокупность обучающих образцов вокального звука, которая описывается множеством целочисленных ограниченных финитных дискретных функций $X=\{x_i \mid i\in\{1,...,I\}\}$, причем $A_i^{\min}, A_i^{\max}$ – минимальное и максимальное значение функции x_i на компакте $\{N_i^{\min},...,N_i^{\max}\}$.

Для сопоставления элементов множества X между собой для каждой функции x_i, описывающей обучающий образец, формируется соответствующее ему конечное множество дочерних дискретных саундлетов S^c, находящихся в том же самом амплитудно-временном окне, что и эта функция, в виде

$$\forall x_i\in X\ \exists S^c=\{s_r^c \mid r\in\{1,...,I\}\}\ :\ s_r^c=Fx_r.$$

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

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

$$\rho_p(x_i,s_r^c)=\sqrt[p]{\sum_m|x_i(m)-s_r^c(m)|^p}\ .$$

Осуществляется выбор множества функций H, описывающих эталонные образцы, из множества функций X, описывающих обучающие образцы, на основе матрицы нормированных расстояний $[d_{ir}]$.

На основе введенных семейств саундлетов и саундлетных отображений и сформированного множества эталонных образцов и подпокрытия $\tilde{C}$ создадим модель синтеза вокального звука по эталонным образцам.

Модель синтеза вокального звука по эталонным образцам

Модель синтеза вокального звука по эталонным образцам создается на основе детерминированного конечного автомата. Детерминированный конечный автомат, синтезирующий некоторый вокальный звук с учетом его квазипериодической структуры, множества эталонных образцов H и подпокрытия $\tilde{C}$, представлен в виде графа на рис.1.

Чтобы повысить эффективность синтеза необходимо определить точное количество повторений u каждого состояния s_k, которое соответствует эталонному образцу h_k, для чего в статье предложена следующая процедура.

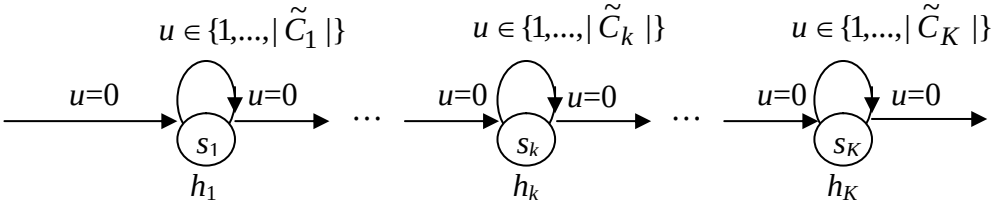

Рисунок 1 – Детерминированный конечный автомат, синтезирующий вокальный звук

Процедура установления соответствия между обучающим образцом и эталонным образцом

Вектор $\bar{k} = (k_1,...,k_i,...,k_I)$, содержащий номера эталонных образцов, которые соответствуют обучающим образцам, формируется согласно следующей процедуре

$$k_i = \arg\min_k \theta(Fx_i, h_k),\ k \in \{1,...,K\},\ i \in \{1,...,I\},$$

$$\theta(Fx_i, h_k) = \frac{\rho_p(Fx_i, h_k)}{(A_k^{\max} - A_k^{\min})\sqrt[p]{(N_k^{\max} - N_k^{\min} + 1)}},\ k \in \{1,...,K\},\ i \in \{1,...,I\},$$

$$\rho_p(Fx_i, h_k) = \sqrt[p]{\sum_m |(Fx_i)(m) - h_k(m)|^p},$$

где x_i – целочисленная ограниченная финитная дискретная функция, описывающая k-й эталонный образец i-й обучающий образец вокального звука,

h_k – целочисленная ограниченная финитная дискретная функция, описывающая k-й эталонный образец вокального звука,

k_i – номер эталонного образца, соответствующий i-му обучающему образцу,

$A_k^{\min}, A_k^{\max}$ – минимальное и максимальное значение функции h_k на компакте $\{N_k^{\min},..., N_k^{\max}\}$,

K – количество эталонных образцов вокального звука,

I – количество обучающих образцов вокального звука.

Для созданной модели сформулируем критерии эффективности.

Критерии оценки эффективности модели

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

$$Q = T \to \min_p.$$

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

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

$$\tilde{y}_i^{model} = \theta(Fx_i, h_{k_i}),$$

где x_i – целочисленная ограниченная финитная дискретная функция, описывающая k-й эталонный образец i-й обучающий образец вокального звука,

$\tilde{y}_i^{test}$ – тестовое расстояние для i-го образца,

h_k – целочисленная ограниченная финитная дискретная функция, описывающая k-й эталонный образец вокального звука,

k_i – номер эталонного образца, соответствующий i-му образцу,

I – количество образцов вокального звука.

Для созданной модели сформулируем критерии эффективности.

Условия синтеза вокального звука по эталонным образцам

Пусть дан вокальный звук, который представлен конечным семейством обучающих образцов, описываемых множеством целочисленных ограниченных финитных дискретных функций $X = \{x_i \mid i \in \{1,...,I\}\}$.

Пусть для этого вокального звука дано множество эталонных образцов, описываемых множеством целочисленных ограниченных финитных дискретных функций $H = \{h_k \mid k \in \{1,...,K\}\}$, вектор $\bar{k} = (k_1,...,k_i,...,k_I)$, содержащий номера эталонных образцов, которые соответствуют обучающим образцам, и множество $\{(N_i^{\min}, N_i^{\max}, A_i^{\min}, A_i^{\max}) \mid i \in \{1,...,I\}\}$, характеризующее положение амплитудно-временного окна для каждого i-го образца.

Пусть для каждого обучающего образца вокального звука вычислено расстояние $\theta(Fh_{k_i}, x_i)$ между функцией x_i, описывающей i-й обучающий образец, и функцией h_{k_i}, описывающей k_i-й эталонный образец вокального звука в виде

$$\theta(Fh_{k_i}, x_i) = \frac{\rho_p(Fh_{k_i}, x_i)}{(A_i^{\max} - A_i^{\min})\sqrt[p]{(N_i^{\max} - N_i^{\min} + 1)}},\ i \in \{1,...,I\},$$

$$\rho_p(Fh_{k_i}, x_i) = \sqrt[p]{\sum_m |(Fh_{k_i})(m) - x_i(m)|^p}\ .$$

Необходимое условие синтеза вокального звука. Вокальный звук синтезирован, если

$$\sum_{i=1}^{I} \theta(Fh_{k_i}, x_i) < \tilde{\varepsilon},$$

где $\tilde{\varepsilon}$ – заданная точность синтеза.

Достаточное условие синтеза вокального звука. Вокальный звук синтезирован, если

$$\sum_{i=1}^{I} \theta(Fh_{k_i}, x_i) = 0.$$

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

Логико-формальные правила оценивания результата синтеза

Для оценивания результатов синтеза формируются следующие логико-формальные правила

если $\sum_{i=1}^{I} \theta(Fh_{k_i}, x_i) < \tilde{\varepsilon}$, то $q = 1$,

если $\sum_{i=1}^{I} \theta(Fh_{k_i}, x_i) \geq \tilde{\varepsilon}$, то $q = 0$,

где $q = 1$ – успешный синтез, $q = 0$ – неуспешный синтез,

$\tilde{\varepsilon}$ – заданная точность синтеза.

Численное исследование метода синтеза вокальных звуков

В табл. 1 приведено сравнение метрик, используемых при формировании эталонных образцов на основе базы данных TIMIT. исследовались все вокальные звуки. Отношение I/K представляет собой отношение общего количества образцов, содержащих вокальные звуки, к количеству эталонных образцов, 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 метрика ρ_2 обеспечивает наименьшее количество эталонных образцов (позволяет сократить количество хранимых образцов вокального звука в 4 раза) и, следовательно, обладает наибольшей обобщающей способностью.

Таблица 1

Оценка метрики

Метрики	Среднее значение I / K
ρ_1	3.0
ρ_2	4.0
ρ_3	3.0
ρ_∞	3.4

Выводы

Новизна. В работе усовершенствован подход к синтезу вокальных звуков, который отличается тем, что позволяет обобщать образцы одного звука различной длины и различным размахом амплитуд, что повышает эффективность синтеза вокальных звуков речи. Получил дальнейшее развитие метод создания множества эталонных образцов, который отличается тем, что основан на семействах саундлетов и саундлетных отображений, что повышает эффективность процедуры формирования эталонных образцов. На основе семейств саундлетов и саундлетных отображений усовершенствована модель синтеза вокальных звуков, которая отличается тем, что позволяет сопоставлять образцы различной длины и использовать адаптивный нормированный порог в логико-формальных правилах, что повышает эффективность синтеза полезных звуков.

Практическое значение. Разработан метод построения модели синтеза вокальных звуков по эталонным образцам на основе семейств саундлетов и саундлетных отображений, что позволяет сократить количество эталонных образцов. Предложен адаптивный нормированный порог для логико-формальных правил оценивания синтеза речевых сигналов, который позволяет с большей вероятностью выделять полезные синтезированные звуки. В результате численного исследования было установлено, что алгоритм синтеза вокальных звуков на основе семейств саундлетов и саундлетных отображений позволяет сократить количество хранимых образцов вокального звука в 4 раза. Созданные алгоритмы могут использоваться в информационных системах для решения задач, связанных с конкатенативным синтезом речи. Предложенные в статье саундлеты и саундлетные отображения могут использоваться в системах распознавания речевых образов

Список использованной литературы

1. Бондарев В.Н., Аде Ф.Г. Искусственный интеллект / В.Н. Бондарев, Ф.Г. Аде. – Севастополь: Изд-во СевНТУ, 2002. – 615 с.
2. Потапова Р.К. Речь: коммуникация, информация, кибернетика / Р.К. Потапова. М.: Радио и Связь, 1997. 528 с.
3. Dutoit T. An introduction to text-to-speech synthesis / T. Dutoit. – Dodrecht: Kluwer Academic Publishers, 1997. – 285 p.
4. Allen J. From text to speech, the MITALK system / J. Allen, S. Hunnicut, D. Klatt. – Cambridge: Cambridge University Press, 1987. – 340 p.
5. Рабинер Л.Р. Цифровая обработка речевых сигналов / Л.Р. Рабинер, Р.В. Шафер. – М.: Радио и связь, 1981. – 495 с.
6. Bailly G. A text-to-speech system for French using formant synthesis / G. Bailly, G. Murillo, O. Dakkak, B. Guerin // Proc. of SPEECH' 88. – Edinburgh, 1988. – P. 255-260.
7. Rabiner L.R. Fundamentals of speech recognition / L.R. Rabiner, B.H. Jang. – Englewood Cliffs: Prentice Hall PTR, 1993. –507 p.
8. Hunt A.J. Unit selection in a concatenative speech synthesis system using a large speech database / A.J. Hunt, A. Black // Proc. of the ICASSP 96. – Atlanta, 1996. – C. 11-14.

9. Hamon C. A diphone system based on time-domain prosodic modifications of speech / C. Hamon, E. Moulines, F. Charpentier // Proc. of ICASP 89. – Edinburgh, 1989. – P. 238-241.
10. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов / Т.К. Винцюк. – К.: Наук. думка, 1987. – 261 с.
11. Федоров Е.Е. Методология создания мультиагентной системы речевого управления / Е.Е. Федоров. – Донецк: изд-во «Ноулидж», 2011. – 356 с.

References

1. Bondarev, V.N., and Ade F.G. *Iskusstvenniy intellekt* (2002), SevNTU, Sevastopol, Ukraine.
2. Potapova, R.K. *Rech: kommunikatsia, informatsia, kibernetika* (1997), Radio i sviaz, Moskva, Russia.
3. Dutoit, T. *An introduction to text-to-speech synthesis* (1997), Kluwer Academic Publishers, Dodrecht, Netherlands.
4. Allen, J., Hunnicut, S., and Klatt, D. *From text to speech, the MITALK system* (1987), Cambridge University Press, Cambridge, UK.
5. Rabiner, L.R., and Shafer, R.V. *Tsifrovaia obraborka rechevykh signalov* (1981), Radio i sviaz, Moskva, USSR.
6. Bailly, G., Murillo, G., Dakkak, O., and Guerin, B. (1988), "A text-to-speech system for French using formant synthesis", *Proc. of SPEECH' 88*, Edinburgh, pp. 255-260.
7. Rabiner, L.R., and Jang, B.H. *Fundamentals of speech recognition* (1993), Prentice Hall PTR, Englewood Cliffs, USA.
8. Hunt, A.J., and Black, A. (1996), "Unit selection in a concatenative speech synthesis system using a large speech database", *ICASSP 96*, Atlanta, pp. 11-14.
9. Hamon, C., Moulines, E., Charpentier, F. (1989), "A diphone system based on time-domain prosodic modifications of speech", *Proc. of ICASP 89*, Edinburgh, pp. 238-241.
10. Vintsiuk, T.K. *Analiz, raspoznavanie i interpretatsia rechevikh signalov* (1987), Naukova dumka, Kiev, USSR.
11. Fedorov, E.E. *Metodologia sozdania multiagentnoi sistemy rechevogo upravlenia* (2011), Noulidzh, Donetsk, Ukraine.

Надійшла до редакції:
18.02.2014 р.

Рецензент:
докт. техн. наук, проф. Н.І. Чичикало

Є.Є. Федоров
Донецька академія автомобільного транспорту

Метод синтезу вокальних звуків мовлення по еталонним зразкам на підставі саундлетів. *У статті викладений метод синтезу вокальних звуків мови по еталонних зразках на основі саундлетів. Використано материнський і дочірній дискретні й безперервні саундлети та досліджені властивості саундлетних відображень, які дозволяють ураховувати структуру квазиперіодического сигналу й зіставляти зразки вокальних звуків мови різної довжини. На основі саундлетов і саундлетних відображень розроблений метод створення зразків, метод формування еталонних зразків і модель синтезу вокальних звуків по еталонних зразках, які використаються в інтелектуальних системах спілкування й дозволяють скоротити обсяг збереженої інформації.*

Ключові слова: *дискретний саундлет, безперервний саундлет, материнський саундлет, дочірній саундлет, еталонні зразки вокальних звуків, синтез звуків.*

E.E. Fedorov
Donetsk Academy of Automobile transport

Method of synthesis of vocal sounds of speech on reference samples on a basis саундлетов. *Existing systems of synthesis of speech patterns use such approaches as: formant synthesis,*

synthesis on the basis of coefficients of the linear prediction (LPC-synthesis) and concatenative synthesis (methods PSOLA and unit selection), and the most effective is concatenative synthesis. However this approach for storage of speech units demands great volumes of a database (especially a method unit selection) in this connection there is a problem of effective representation of speech units. For the decision of this problem in article the method of synthesis of vocal sounds of speech on reference patterns on a basis soundlets is stated. The author enters concept mother soundlet (the pattern of a vocal sound of the speech, shifted on time and amplitude into the left bottom corner of a positive plane) and child soundlet (shifted and scaled on time and amplitude mother soundlet) both their discrete and continuous versions. The author enters the maps operating between families discrete and continuous mother and child soundlets. In work the approach to synthesis of vocal sounds which differs that allows to generalise patterns of one sound of various length and various scope of amplitudes that raises efficiency of synthesis of vocal sounds of speech is improved. Has received the further development a method of creation of set of reference patterns which differs that is based on families soundlets and soundlets maps that raises efficiency of procedure of formation of reference patterns. On the basis of families soundlets and soundlets maps the model of synthesis of vocal sounds which differs that allows to compare samples of various length is improved and to use an adaptive normalised threshold in logic-formal rules that raises efficiency of synthesis of useful sounds. The method of construction of model of synthesis of vocal sounds on reference patterns on the basis of families soundlets and soundlets maps that allows to reduce quantity of reference patterns is developed. The adaptive normalised threshold for logic-formal rules evaluated synthesis of speech signals which allows with big probability to allocate the useful synthesised sounds is offered. As a result of numerical research it has been established, that the synthesis algorithm of vocal sounds on the basis of families soundlets and soundlets maps allows to reduce quantity storage samples of a vocal sound in 4 times. The created algorithms can be used for the decision of the problems connected with concatenative by synthesis of speech in information systems, and also in systems of reproduction of texts. Offered in article soundlets and soundlets maps can be used in systems of recognition of speech patterns.

Keywords: *discrete soundlet, continuous soundlet, mother soundlet, child soundlet, reference patterns of vocal sounds, synthesis of sounds.*

Федоров Евгений Евгениевич, Украина, закончил Донецкий национальный технический университет, докт. тех. наук, доцент, проректор по научной работе Донецкой академии автомобильного транспорта (просп. Дзержинского, 7, г. Донецк, 83086, Украина). Основное направление научной деятельности – методы идентификации и верификации диктора; распознавания и синтеза речи; методы анализа и синтеза естественно-языковых объектов.