УДК 004.934

О. Н. КАРПОВ, О. А. САВЕНКОВА (ДНУ)

СИНТЕЗ СПЕКТРАЛЬНО-ВРЕМЕННЫХ ПАРАМЕТРОВ МОДЕЛИ БЛОКА РАСПОЗНАВАНИЯ РЕЧИ В АВТОМАТИЗИРОВАННОЙ СИСТЕМЕ УПРАВЛЕНИЯ

Розглянуто алгоритм сегментно-слогового синтезу спектрально-часових параметрів моделі блока розпізнавання мови.

Рассмотрен алгоритм сегментно-слогового синтеза спектрально-временных параметров модели блока распознавания речи.

Algorithm of segment-syllabic synthesis of time-spectrum parameters for the model of speech recognition block is introduced.

Построение устройств распознавания речи для современных АСУ состоит из следующих основных задач [1; 5]:

– выбор объектов или типов речевых единиц (фонемы, слоги, слова, морфемы, фразы);

– выбор параметров описания речевых единиц и соответствующих методов интерпретации описаний;

– проектирование программных средств реализации описаний выбранных объектов и распознавания;

– встраивание разработанных проектов и программных реализаций в системные среды.

Структурная схема блока распознавания речи в АСУ, которая позволяет решить вышеперечисленные задачи, представлена на рис. 1.

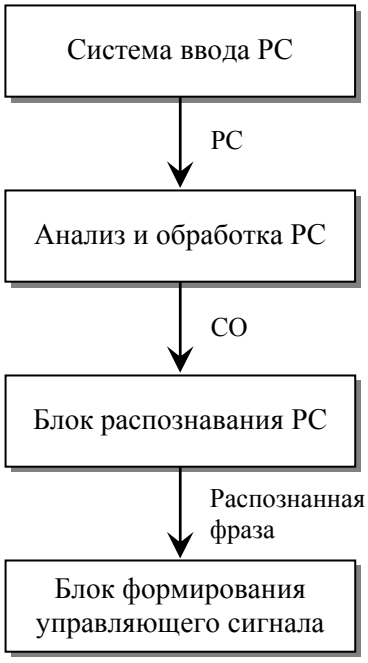

Рис. 1. Структурная схема блока распознавания речи: РС – речевой сигнал; СО – спектральный образ

Высокий уровень развития вычислительных средств позволяет решать задачи построения систем распознавания речи (СРР), с использованием большого числа параметров и методов, которые ориентированы на детальное изучение структуры речевого сигнала. Однако по прежнему осталась необходимость разрешения противоречивых требований для СРР: обеспечения высокой надежности распознавания для больших словарей объектов, что требует привлечения большого числа параметров (признаков) и, соответственно, большого времени их обработки; выполнение обработки с минимальными временными затратами [1; 2].

Общие схемы анализа и распознавания

Структурная схема блока анализа и обработки РС в существующих системах распознавания содержит следующие дополнительные этапы обработки параметров РС, которые повышают надежность распознавания: блоки выделения полезного сигнала, блоки фильтрации сигнала и его спектра, блоки сегментации РС [1; 4; 5]. На вход блока распознавания поступает сегментированная последовательность параметров РС или спектральный образ (СО). В сегментированной последовательности спектрально-временных параметров (траектории параметров) предъявленного РС сегменты находятся в некоторой зависимости от параметров предшествующих и последующих сегментов, поэтому необходимо рассматривать непрерывные траектории в терминах параметров и в терминах сегментов для решения задачи на этапе распознавания [1; 4]. Блок распознавания РС в зависимости от типов речевых единиц содержит: блоки выбора метрики и критериев сравнения; блоки выбора методов сравнения и стратегии распознавания, обеспечивающую

минимальные затраты времени на поиск наиболее подходящего эталона с максимальной надежностью в заданной системе параметров. Такая конфигурация может быть усовершенствована благодаря введению дополнительного блока аппроксимации траекторий параметров в терминах речевых единиц (РЕ), которая может быть представлена структурной компоновкой крупных речевых единиц (слов, фраз, предложений) из мелких (фонем, слогов) и выбор их наилучшего соответствия некоторой группе сегментов предъявленной реализации на каждом шаге сопоставления. Процесс укрупнения РЕ продолжается до тех пор, пока не будет найдено наилучшее соответствие для всего речевого высказывания по всей совокупности РЕ словаря для всех сегментов речевого высказывания. Наибольшей надежностью обладает пословное распознавание [1; 5]. Для РЕ, поступающих на блок аппроксимации, необходимым условием является следующий факт: РЕ должны иметь такую длину и быть подобраны в таком количестве, чтобы из них можно было бы построить любые другие слова или предложения. Этим требованиям удовлетворяют РЕ слова-слоги, которые содержат два, три символа-фонемы. Задача нахождения наилучшей траектории для предъявленной реализации РС в терминах РЕ обеспечивается перекрытием накладываемых на траекторию параметров РЕ в соответствии с алгоритмом сегментно-слогового синтеза. Процесс перебора РЕ и нахождения наилучшей траектории параметров требует значительных временных затрат. Для решения этой проблемы предлагается введение блока выбора кандидата для распознавания, который использует формализованные эвристики для исследуемой области.

Описание алгоритмов работы блока аппроксимации на основе решения задачи сегментно-слогового синтеза

Задачу сегментно-слогового синтеза (ССС) формулируем согласно [2]. Пусть задан словарь слогов $\{SL_k\}$ ($k = 1 \dots N$), для каждого из которых задана эталонная последовательность параметров или траектория параметров $Yk = (Yk_1, Yk_2, \dots, Yk_{mk})$, где mk – количество точек траектории параметров для k-го слога. Каждый слог SL_k содержит n_k символов-фонем α_j^k ($k = 1 \dots N$, $j = 1 \dots n_k$). Каждая траектория параметров Y_k содержит mk элементов, объединяемых в n_k сегментов $SG_j^{Y_k}$ для соответствующих символов-фонем α_j^k. Пусть задана входная последовательность параметров $X = (x_1, x_2, \dots x_{\rightleftarrows r})$, которая сегментирована на p сегментов-фонем SG_l^x ($l = 1 \dots p$), объединенных в M групп-слогов X_i ($i = 1 \dots M$). Необходимо последовательность X наилучшим образом поставить в соответствие эталонным последовательностям параметров $\{Y_k\}$, вычисляя расстояние

$$d = \sum_i \min_k \left(X_i \# Y_k \right), \qquad (1)$$

где X_i, Y_k содержат сегменты-фонемы SG_l^x, $SG_j^{Y_k}$ соответственно; # – операция сопоставления осуществляется с помощью динамического программирования. Таким образом, необходимо найти такую эталонную траекторию параметров $X*$, для которой достигнута наилучшая близость с траекторией параметров X предъявленного речевого сигнала по всей совокупности слогов, для эталонной траектории параметров результат распознавания строится как синтез соответствующих РЕ. В [3] предложен алгоритм поиска вариантов-комбинаций $\widetilde{X}*$ для эталонной траектории параметров с помощью стратегий поиска в глубину и в ширину $\widetilde{X}^* = \left(Y_1, Y_2, \dots, Y_i, \dots, Y_R\right)$, где R – количество слогов траектории параметров $\widetilde{X}*$, соответствует количеству слогов предъявленной реализации.

Предложенный алгоритм синтеза эталонной траектории параметров может быть дополнен новым уровнем обработки траекторий параметров слогов-эталонов, составляющих данную траекторию. Поскольку последовательности спектрально-временных параметров обычно искажены или зашумлены (нестационарный РС приблизительно описывается существующими ортогональными системами функций), то для получения плавно меняющейся функции параметров предлагается построение квадратичной и кубической моделей сплайн-описания траекторий параметров слогов-эталонов для синтеза спектрально-временных параметров на основе функций, которые обеспечивают непрерывную аппроксимацию значений параметров и сглаживание.

Модели сплайн-описания и сплайн-синтеза эталонных траекторий параметров

Задача настройки параметров одной траектории к другой наилучшим образом – задача

минимизации среднеквадратичного приближения модели преобразования с линейными условиями-равенствами, которые обеспечивают требуемую гладкость в точках склейки траекторий параметров слогов-эталонов (полученные эталонные траектории параметров должны быть непрерывны по нулевой и первой производным на всем временном интервале).

Рассмотрим следующую модель сплайн-описания параметров траекторий эталонов Y_k ($k = 1 \ldots R$), которые входят в синтезированную траекторию параметров $\widetilde{X}^*$:

$$\widetilde{X}^* = \begin{cases} \widetilde{Y}_{1i}, & N_0 \le i < N_1, \\ \widetilde{Y}_{2i}, & N_1 \le i < N_2, \\ \ldots \\ \widetilde{Y}_{ki}, & N_{k-1} \le i < N_k, \\ \ldots \\ \widetilde{Y}_{Ri}, & N_{R-1} \le i < N_R, \end{cases} \qquad (2)$$

- квадратичная модель преобразования траекторий эталонов имеет следующий вид

$$\widetilde{Y} = a_1 \cdot X^2 + a_2 \cdot X + a_3,$$

где a_1, a_2, a_3 – параметры квадратичной модели преобразования;

- кубическая модель преобразования траекторий эталонов имеет вид

$$\widetilde{Y} = a_1 \cdot X^3 + a_2 \cdot X^2 + a_3 \cdot X + a_4,$$

где a_1, a_2, a_3, a_4 – параметры кубической модели преобразования.

Для нахождения неизвестных коэффициентов/параметров моделей решается задача минимизации среднеквадратичного приближения

$$\sigma^2 = \sum_{k=1}^{R} \left(\sum_{i=N_{k-1}}^{N_k} \left| \widetilde{Y}_{ki} - Y_{ki} \right|^2 \right) \to \min \qquad (3)$$

с линейными условиями-равенствами в точках склейки траекторий параметров слогов-эталонов T_j ($j = k-1$, $k = 1 \ldots R$):

а) равенство значений параметров склеиваемых траекторий

$$\widetilde{Y}_k(T_j) = \widetilde{Y}_{k+1}(T_j); \qquad (4)$$

б) равенство значений производных функций параметров траекторий в точке склейки

$$\widetilde{Y}'_{k-1}(T_j) = \widetilde{Y}'_k(T_j). \qquad (5)$$

Формализация факторов эвристической функции для поиска оптимальных решений сегментно-слогового синтеза

Для нахождения эталонной траектории (ЭТП) параметров необходимо сопоставить все возможные комбинации траекторий параметров (ТП) $\widetilde{X}$, составленные из ТП имеющихся в словаре слогов-эталонов, с ТП X, что требует огромных временных затрат. Сокращение рассматриваемых вариантов, а соответственно временной и пространственной сложности, может быть достигнуто благодаря использованию базовых стратегий поиска в глубину и в ширину. Решения, найденные с помощью базовых стратегий, не всегда оптимальны в смысле наилучшей близости, так как при раскрытии узлов в пространстве поиска не используется информация о данной проблемной области [3]. Использование эвристик предполагает: выявление факторов для оценки состояний и степени значимости каждого фактора; определение эвристических оценок для узлов на графе синтеза ЭТП, определяющих перспективность рассматриваемого узла с точки зрения достижения целевого состояния.

Оценочная функция (ОФ) сводится к виду, в котором формализованы наиболее значимые характеристики сегментно-слогового представления состояния: вложенность слогов, наличие групповых признаков сегментов (тон, шум, пауза), величина отклонения слогов предъявленной реализации и эталонной. Таким образом, для каждого узла n на графе синтеза ЭТП определяется ОФ вида

$$f(n) = g(n) + h(n),$$

где $g(n)$ – стоимость пути к узлу n, а $h(n)$ – оценка достижения целевого состояния из узла n. ОФ может быть представлена также в виде логической связки предикатов или факторов выбора. Для оценки качества поиска с помощью ОФ вычисляется величина целенаправленности поиска (показывает, в какой мере поиск идет в направлении к цели) $P = L/T$, где L – длина найденного пути к цели, T – общее число вершин, раскрытых в процессе поиска.

Экспериментальные исследования

Для проведения исследований была модифицирована система распознавания Speach. Программа работает в реальном времени на компьютере типа IBM PC с процессором Intel Celeron 700 МГц. Основные функции, реали-

зуемые распознающей системой: система акустического ввода и вывода информации (ввод речевого сигнала с микрофона, wav-файлы; воспроизведение РС и визуальное отображение формы РС во временной области; выделение полезного сигнала от шумов окружающей среды; визуальное отображение спектра РС; блоки распознавания: обучение (создание новых словарей эталонов, дополнение существующих словарей); распознавание речевых команд на основе алгоритма ССС; блок принятия решения и оценки решения о распознавании и формирования управляющего сигнала.

Рассмотрим задачу построения эталонной траектории параметров X^*, которые представлены в виде системы энергетических спектров $\{S_{ik}\}$, распределенных по частотным группам, для РС, который поступает на вход системы распознавания, из некоторого множества слогов-эталонов. Для предъявленного РС определены границы сегментов одним из методов сегментации [1]. Множество доступных слогов включает словари $\{E^i\}$ $(i = 2, 3, 4)$, которые состоят из двух-, трех- и четырехсегментных слогов (содержат два, три символа-фонемы) для заданного набора слов. Для построения словарей слогов были предварительно проанализированы слова – цифры от нуля до ста. Для предъявленной траектории параметров X необходимо найти такую синтезированную эталонную траекторию параметров X^*, для которой расстояние (1) минимально по всей совокупности слогов. Поиск одного из возможных вариантов-комбинаций для X^* осуществлен на графе синтеза эталонных траекторий параметров с помощью базовых стратегий поиска (поиск в глубину, поиск в ширину) [3] и использованием эвристической функции для выбора очередного кандидата-эталона. Для нахождения оптимального решения для этого варианта-комбинации решена задача настройки параметров траекторий слогов-эталонов относительно параметров соответствующих слогов предъявленной реализации на основе квадратичной и кубической моделей сплайн-описания спектрально-временных траекторий параметров с одним узлом в точке склеивания смежных слогов-эталонов.

Результат синтеза оптимальных траекторий параметров Y_model2 и Y_model3 в одной из частотных полос на основе квадратичной и кубической моделей для слова «адин» из траекторий параметров двух слогов-эталонов «ад» и «ин» представлены на рис. 2, 3.

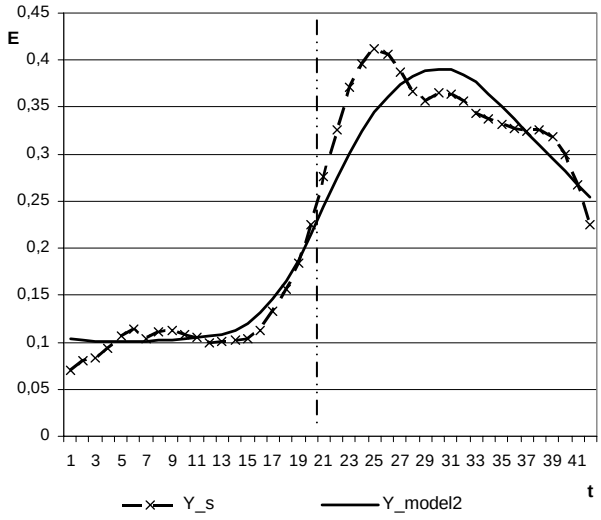

Рис. 2. Эталонная траектория параметров на основе квадратичной модели с одним узлом

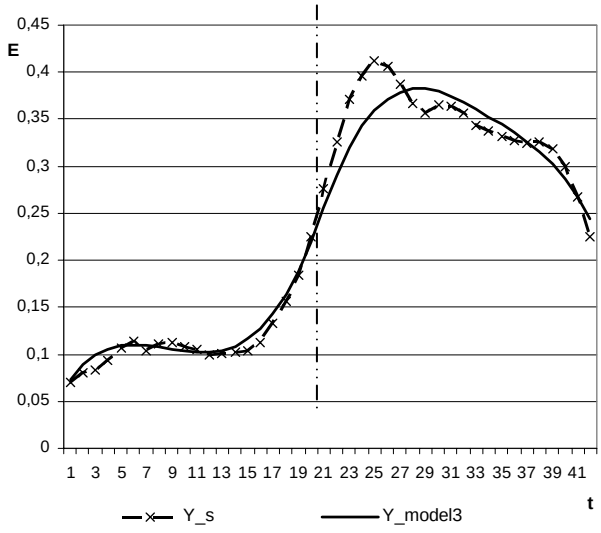

Рис. 3. Эталонная траектория параметров на основе кубической модели с одним узлом

Была исследована модификация алгоритма для блока аппроксимации траекторий параметров, которая заключается в следующем: в качестве узлов для сплайн-описания и сплайн-синтеза рассматриваются точки сегментации каждого слога и точки склеивания смежных слогов-эталонов. Траектории параметров в каждой частотной полосе между границами сегментации имеют простой вид, что позволяет их аппроксимировать полиномами невысокой степени на каждом таком интервале.

На рис. 4, 5 представлены эталонные траектории параметров Y_model2 и Y_model3, построенные на основе квадратичной и кубической моделей с узлами в точках сегментации и в точках склеивания смежных слогов-эталонов.

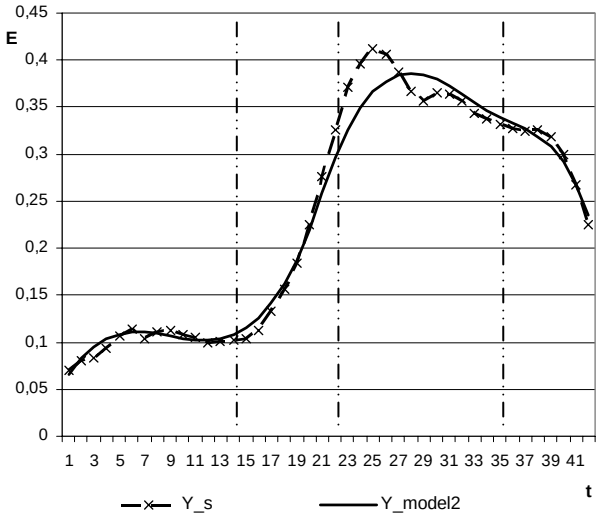

Рис. 4. Эталонная траектория параметров
на основе квадратичной модели с узлами
в точках сегментации и в точках склеивания

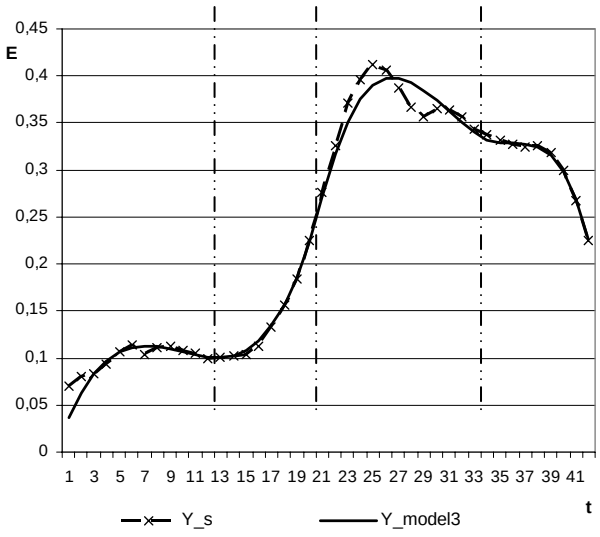

Рис. 5. Эталонная траектория параметров
на основе кубической модели с узлами
в точках сегментации и в точках склеивания

Анализ экспериментов по распознаванию эталонных траекторий параметров на тестовом наборе речевых образцов показал, что использование модифицированного алгоритма позволяет получить в среднем на 5 % лучшие результаты распознавания, вследствие уменьшения погрешности аппроксимации.

Выводы

Представленный алгоритм для моделирования блока распознавания речи является развитием алгоритма, предложенного в [3], и добавляет новый уровень обработки траекторий параметров слогов-эталонов, позволяющий повысить надежность распознавания. Достоинства предложенных моделей настройки траекторий параметров: модели зависят от малого числа линейных параметров; построение таких моделей основано на применении стандартных быстродействующих алгоритмов. Применение квадратичной и кубической моделей настройки параметров эталонных траекторий позволило увеличить надежность распознавания на 3 % по сравнению с базовым алгоритмом ССС. Использование эвристической функции выбора кандидата-эталона позволило сократить время распознавания в среднем в 10 раз. Полученные результаты свидетельствуют о том, что добавление блока аппроксимации траекторий параметров в общую схему блока распознавания улучшает характеристики системы распознавания речи в целом.

Дальнейшая работа проводится в направлении исследований модифицированного алгоритма для модели 3-го порядка сплайн-описания и сплайн-синтеза эталонных траекторий параметров.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Карпов О. Н. Технология построения устройств распознавания речи: Моногр. – Д.: Изд-во Днепропетр. ун-та, 2001. – 184 с.
2. Карпов О. Н. Некоторые эксперименты по повышению надежности распознавания слов заданного словаря / О. Н. Карпов, О. А. Савенкова // Системные технологии. Вып. 6 (35), – Д., 2004, С. 60–66.
3. Карпов О. Н. Распознавание речи на основе сегментно-слогового синтеза в терминах пространства состояний / О. Н. Карпов, О. А. Савенкова // Искусственный интеллект. – 2006. – № 3. – С. 532–536.
4. Kopeček I. Speech recognition and syllable segments. // http://www.fi.muni.cz/~kopecek/
5. Ronzhin A. L. Survey of Russian Recognition Systems. // R. M. Yusupov, I. V. Li, A. B. Leontieva. In Proc. of Int. Conf. SPECOM'2006, St. Petersburg, 2006, pp. 54–60.

Поступила в редколлегию 29.03.07.