

<http://like-money.ru/stati/259-metody-avtomaticheskoy-rubrikaczii-i-oczenka-ix-kachestva>.

3. Построение семантической сети из разнородных данных [Электронный ресурс] / Панченко А. – Режим доступа: [http://it-claim.ru/Persons/Panchenko/presentation\\_2010\\_sept\\_final.pdf](http://it-claim.ru/Persons/Panchenko/presentation_2010_sept_final.pdf).

**Крамаренко А.В.**

**Науч. руководитель к.т.н. Жук А.В.**

*Донецкий национальный технический университет*

**Автоматизированное наполнение звуковой базы  
синтезатора речи с использованием методов  
кратномасштабного анализа**

Одним из развиваемых способов организации взаимодействия человека с компьютером является речь. В связи с этим данную задачу разделяют на два направления – распознавание и синтез речи.

При проектировании и создании синтезаторов речи часто используют звуковую базу [1], содержащую либо отдельные фонемы, либо их сочетания, либо полноценные записи слов, либо словосочетания в зависимости от подхода к синтезу. На первых этапах создания синтезаторов наполнение базы может производиться вручную. Однако такой подход чреват временными затратами на поиск необходимых сегментов, также его недостатками являются субъективное человеческое восприятие звукового сигнала, сложность воспроизведения результатов. В связи с этим встает вопрос автоматизации данного процесса с использованием методов сегментации речи.

Существует множество методов сегментации, основанных на различных математических аппаратах и позволяющих получать воспроизводимые результаты сегментации с разным качеством и точностью.

Целью данной работы является разработка алгоритма сегментации речи, основанного на кратномасштабном анализе, для автоматизированного наполнения звуковой базы синтезатора речи. Особым требованием, выдвигаемым к разрабатываемому алгоритму, является его дикторнезависимость, поскольку данное свойство позволяет, не изменяя параметры алгоритма под отдельного диктора, получить базу звуков для различных голосов.

В работе предлагается производить сегментацию записей целых слов, произнесенных различными дикторами, основываясь на априорном знании транскрипции этих слов.

Как известно, речевой сигнал состоит из квазистационарных участков, соответствующих голосовым и шипящим фонемам, перемежаемым участками со сравнительно быстрыми изменениями спектральных характеристик сигнала [2]. В пределах стационарных участков значительную роль для анализа речи играют спектральные особенности сигнала, изменяющиеся в процессе артикуляции. Поэтому весьма эффективным для анализа речевого сигнала представляется кратномасштабный анализ и вейвлет-преобразование [3].

При использовании метода кратномасштабного анализа [4] на записях целых слов было определено, что возможно появление в результатах меток границ, не соответствующих ни одной из позиций транскрипции. Такая ситуация происходит из-за неверно подобранных параметров сегментации  $\eta$  и  $g_{\text{пор}}$ .

С целью усовершенствования алгоритма предлагается следующая модификация:

1. Подсчет количества необходимых меток сегментирования  $N$ , равного количеству фонем в транскрипции слова минус один.

2. Изменение параметров  $\eta$  и  $g_{\text{пор}}$  в диапазонах  $[2.5;4]$  и  $[0.4;1]$ , соответственно, для формирования вектора  $T$ , задающего всевозможные уникальные метки границ;

3. Проверка выполнения условия (1) в векторе  $T$ :

$$T[i] - T[i+1] < 512, \quad (1)$$

где  $T[i]$  – значение границы в  $i$  позиции вектора;

Если условие (1) выполняется, то граница  $T[i]$  полагается равной следующему значению:

$$T[i] = \frac{T[i] + T[i+1]}{2}, \quad (2)$$

а значение метки границы  $T[i+1]$  удаляется из вектора  $T$ .

4. Если размер вектора  $T$  больше чем  $N$ , то формируется вектор  $P$ , описывающий скорость изменения мощности спектра на границах:

$$P[i] = |F_{\text{max}}(j) - F_{\text{max}}(j-1)| + |F_{\text{max}}(j) - F_{\text{max}}(j+1)|, \quad j = T[i]/256, \quad (3)$$

где  $j$  – это номер окна в 256 отсчетов звукового сигнала для границы  $T[i]$ ;  $F_{\text{max}}(j)$  – максимальное значение спектра звукового сигнала в  $j$ -м окне.

5. До тех пор, пока размер вектора  $T$  больше  $N$ , из вектора  $T$  удаляются границы, соответствующий которым элемент вектора  $P$  является минимальным.

Было выполнено сравнение результатов работы модифицированного и исходного алгоритмов на выборке записей слов, содержащих различные типы звукоочетаний. Результаты обоих алгоритмов сравнивались с эталонными метками межфонемных переходов проставленных вручную, и рассчитывалась величина погрешности  $E$  равная:

$$E = \sum |V[i] - W[i]| + Q, \quad (4)$$

$$Q = 500 * ||W| - |V||,$$

$$i = [1; n],$$

где:  $W$  – вектор эталонных меток границ;  $V$  – вектор полученных меток границ;  $Q$  – штраф за найденное неправильное количество границ;  $n$  – наименьший размер вектора между  $W$  и  $V$ .

По результатам тестов в подавляющем большинстве случаев модифицированный алгоритм производит более качественное сегментирование.

Выводы: предложенный модифицированный алгоритм сегментации на основе алгоритма кратномасштабного анализа позволяет находить метки границ межфонемных переходов в речевых сигналах различных дикторов. Эффективность алгоритма по сравнению с исходным была подтверждена подсчетом ошибки определения положения метки границы. Данный модифицированный алгоритм в дальнейшем может быть использован для автоматизированного наполнения звуковой базы синтезаторов речи.

#### Литература.

1. Сорокин В.К. Синтез речи / Сорокин В.К. – М. : Наука, 1992. – 392 с.
2. Сорокин В.Н. Сегментация и распознавание гласных / В.Н. Сорокин, А.И. Цыплихин // Информационные процессы. – 2004. – Т. 4, №2. – С. 202-204.
3. Дремин И.М. Вейвлеты и их использование / И.М. Дремин, О.В. Иванов, В.А. Нечитайло // Успехи физических наук. – 2001. – Т. 171, № 5. – С. 465-500.
4. Ермоленко Т.В. Алгоритмы сегментации с применением быстрого вейвлет-преобразования [Электронный ресурс]. –

Режим доступа : <http://www.dialog-21.ru/Archive/2003/Ermolenko.htm>