

КИРИЛЛ КОНСТАНТИНОВИЧ ТОМЧУК

ассистент кафедры бортовой радиоэлектронной аппаратуры факультета радиотехники, электроники и связи, Санкт-Петербургский государственный университет аэрокосмического приборостроения (Санкт-Петербург, Российская Федерация)

wake@inbox.ru

АНДРЕЙ ЮРЬЕВИЧ ЗИЛИНБЕРГ

кандидат технических наук, старший преподаватель кафедры бортовой радиоэлектронной аппаратуры факультета радиотехники, электроники и связи, Санкт-Петербургский государственный университет аэрокосмического приборостроения (Санкт-Петербург, Российская Федерация)

uaz13@yandex.ru

ИЕРАРХИЧЕСКАЯ МОДЕЛЬ И БАЗОВЫЕ АЛГОРИТМЫ ВРЕМЕННОЙ СЕГМЕНТАЦИИ РЕЧЕВЫХ СИГНАЛОВ

Представлен системный подход к временной сегментации речевых сигналов, основанный на разработанной иерархической модели речевых сигналов. Обосновано применение многоуровневой временной сегментации речевых сигналов и разработаны базовые алгоритмы для основных уровней сегментации: VAD-алгоритм; деление речевых сигналов на вокализованный, шумовой, взрывной сегменты; сегментация вокализованных фрагментов на периоды основного тона. Предложен возможный вариант разделения смежных вокализованных звуков с помощью анализа их структуры.

Ключевые слова: речь, модель речевого сигнала, сегментация, VAD, основной тон

ВВЕДЕНИЕ

Автоматическая обработка речи является современной и актуальной задачей, представленной широким кругом областей приложений: шумоподавление, сжатие речевых сигналов (РС), распознавание, модификация РС, кодирование и т. д. [3], [4], [5], [7], [8], [9], [10], [11], [15]. В силу недостаточной изученности процесса восприятия речи человеком, сложности структуры речевого сигнала, индивидуальных особенностей речи каждого человека, наличия огромного числа языковых культур многие из задач, стоящих перед автоматической обработкой речи, так и не были решены в достаточной степени для свободного практического применения.

Голосовой аппарат человека является акустической системой, состоящей из вибраторов, возбуждаемых потоком воздуха из легких в соответствии с законом Бернулли, и резонаторов – глотки, носовой и ротовой полостей. Механизмы возбуждения акустических колебаний связаны либо с работой гортани, либо с возникновением шумных или импульсных звуков при прохождении воздушного потока через сужения, образующиеся в определенных местах речевого тракта [13].

Все гласные звуки произносятся при активности только голосовых связок и без создания сужений в голосовом тракте. При произнесении согласных звуков в речевом аппарате возникают препятствия для свободного прохождения воздуха. В результате согласные, в том числе вокализованные ([*b*], [*n*], [*p*], ...), во-первых, имеют меньшую мощность сигнала, по сравнению с

гласными, а во-вторых, если при произнесении звука воздух с силой проталкивается через значительные сужения, в сигнале появляется шумовая компонента ([*z*], [*c*], [*sh*], [*m'*]). Кроме того, при произнесении согласного голосовые связки могут быть неактивны – тогда произносится либо чисто шумовой звук, возникающий при длительном протягивании воздуха через значительное сужение (ср. вокализованный [*z*] и шумовой [*c*]), либо так называемый взрывной звук, возникающий при силовом проходе воздуха через резко образующее отверстие ([*n*], [*m*], [*m'*], ...). Таким образом, вокализованные (звонкие) звуки формируются с участием голосовых связок, шумовые (фрикативные) – за счет прохождения воздуха через сужения голосового тракта, а взрывные – с помощью кратковременного смыкания речевого аппарата, создания в речевых полостях повышенного давления и затем резкого размыкания аппарата. Частота колебания голосовых связок при произнесении вокализованных звуков называется частотой основного тона (ОТ). Форма голосового тракта остается в основном неизменной на интервале 10–30 мс. На этом интервале речь можно рассматривать как квазистационарный случайный процесс, поэтому большинство алгоритмов предварительной обработки основываются на анализе РС на указанном интервале времени.

СИСТЕМНЫЙ ПОДХОД К МНОГОУРОВНЕВОЙ ВРЕМЕННОЙ СЕГМЕНТАЦИИ РЕЧЕВЫХ СИГНАЛОВ

На основе информации о строении речевого сигнала, об особенностях реализаций звуков

русской речи задача временной сегментации может быть разбита на несколько последовательно осуществляемых этапов, каждый из которых представляет собой соответствующий уровень сегментации РС (рис. 1).

На первом уровне сегментации (блок 1, рис. 1) происходит разбиение речевого сигнала на фрагменты активной речи и пауз. При этом в роли паузы может также выступать смычка – кратковременная пауза в пределах слова перед взрывным звуком. Вторым уровнем сегментации (блок 2, рис. 1) является разбиение активных участков речи на фрагменты, соответствующие трем основным типам звуков: вокализованные, шумовые и взрывные. Третий уровень (блок 3, рис. 1) относится к дополнительной сегментации только вокализованных фрагментов – разделение РС на отдельные периоды основного тона. Разбиение на периоды основного тона не обязательно является финальной стадией сегментации: в дальнейшем на ее основе может быть выполнено разбиение вокализованных фрагментов на отдельные аллофоны или группы из однотипных аллофонов. В данном случае становится возможным построение собственных нейронных сетей, марковских моделей и других эталонов для различных типов фонем. Например, в системе верификации, разрабатываемой компанией SPIRIT Corp. [12], процедура классификации, основывающаяся на смешанных гауссовских моделях, позволяет принимать во внимание отдельные фонемы или группы фонем и особенности их произнесения конкретным диктором.



Рис. 1. Этапы сегментации и параметризации речевого сигнала

На рис. 1 приведен алгоритм предлагаемого системного подхода к сегментации и последующей параметризации речевого сигнала.

На первом шаге обработки (блоки 1...3, рис. 1) сигнал сегментируется на фрагменты активной речи и паузы. Паузы не подвергаются дальнейшей сегментации, а информативным параметром (блок 9) участков паузы является их длительность. Над участками активной речи производится следующий, второй, уровень сегментации, на котором выделяются сегменты,

соответствующие определенным типам звуков: вокализованные (блок 4), шумные (блок 8), взрывные глухие (блок 7).

Сегментация «речь / пауза»: VAD-алгоритм (блок 1, рис. 1)

Алгоритм определения временных границ речевой активности говорящего и в зарубежной, и в отечественной литературе сокращенно обозначают как VAD-алгоритм (Voice Activity Detection). Он позволяет произвести сегментацию сигналов по двум качественным параметрам: пауза и активность говорящего.

В результате применения VAD-алгоритма речевой сигнал становится сегментированным на А- и S-сегменты: сегменты активности говорящего А (activity) и сегменты пауз S (silence).

Одной из основных проблем, возникающих перед разработчиками VAD-алгоритма, является обеспечение его надежной работы в условиях шумов, помех, наводок. При этом алгоритм должен хорошо обнаруживать границы речевой активности даже при слабых граничных (начальных и конечных) звуках. Это могут быть шумовые звуки, сравнимые по уровню с фоновыми шумами, и плавно нарастающие / убывающие по уровню вокализованные звуки. Для борьбы с этими проблемами нами был разработан новый энергетический VAD-алгоритм.

В этом алгоритме для нахождения порога (блок 5, рис. 2) и сравнения с ним используется сигнал, полученный на основе модуля отсчетов исходного сигнала (блоки 2 и 3, рис. 2), пропущенного для подавления НЧ-наводок через высокочастотный фильтр (блок 1, рис. 2). При формировании порога определения речевой активности необходимо использовать характеристику всего набора значений отсчетов фонового шума. Это можно осуществить, определяя уровень порога на гистограмме (блок 4, рис. 2) отсчетов [2] такого фильтрованного сигнала, по которой видно общее распределение как ВЧ-шума, так и НЧ-компонент речевого сигнала. Функциональная схема такого алгоритма VAD представлена на рис. 2.

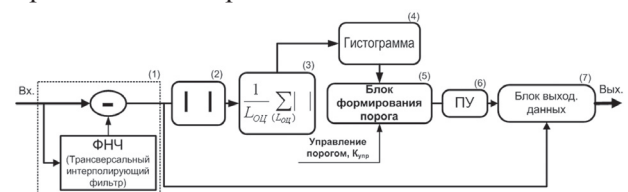


Рис. 2. Функциональная схема «энергетического» алгоритма VAD

В алгоритме VAD на интервалах оценивания текущих средних значений модулей отсчетов (длина интервала $L_{\text{оц}} = 2,5 \dots 7$ мс в зависимости от частоты дискретизации: по достаточной по шуму выборке в 50–70 отсчетов) осуществляется суммирование модулей отсчетов РС с последующей нормировкой на длину интервала (блок 3,

рис. 2). Гистограмма строится по результатам обработки отсчетов на интервале порядка 1,5...2 с, в течение которых внешний шум изменяется незначительно (стационарен). Для формирования порога активности используется мода гистограммы, характеризующая средний модуль ВЧ-сигнала на выходе фильтра (блок 1, рис. 2) в паузах.

Основным недостатком энергетического алгоритма является низкая точность определения границ слабых вокализованных звуков и – главное – высокая вероятность пропуска некоторых шумовых звуков, в особенности кратковременных шумовых звуков, расположенных обособленно от других между смычкой и паузой (например, звук [m'] на конце слова).

Для уменьшения неточности определения границ слабых вокализованных звуков необходимо вводить в VAD-алгоритм детектор вокализованных звуков, не зависящий от уровня сигнала (см. раздел «Сегментация “вокализованный / шумовой / взрывной”»).

Для отделения слабых шумовых звуков речи от фонового шума целесообразно воспользоваться текущим спектральным составом сигнала. Исследования показывают, что наиболее часто встречающийся фоновый шум (шум помещений, уличный шум) имеет более низкочастотный спектральный состав, нежели шумовые звуки активной речи. В этом случае для шумовых фрагментов речевого сигнала (как фоновых шумов, так и шумовых звуков) на спектрограмме вычисляется траектория «центра масс» спектра. Если оценка «центра масс» превышает некоторый пороговый уровень подряд на нескольких интервалах оценивания, есть основания полагать, что такой фрагмент относится к речевой активности.

На рис. 3 изображена временная функция только шумовой составляющей речевого сигнала, представляющего слово «вафли» (нешумовые составляющие вырезаны из реализации), и соответствующая спектрограмма с обозначением порога и «центра масс» спектра. На временной функции область шумов с высокой дисперсией соответствует звуку [ф].

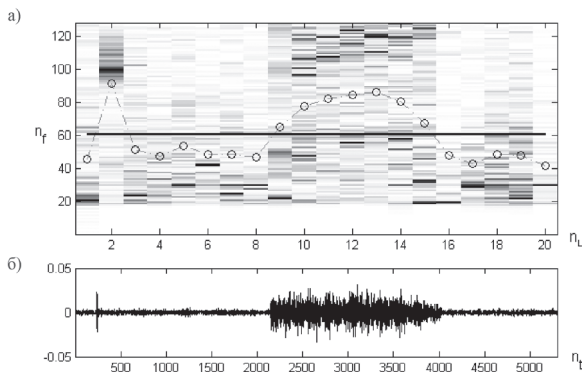


Рис. 3. Спектральное разделение речевых и фоновых шумов: а) спектрограмма сигнала с указанием траектории «центра масс» спектра и порогового уровня; б) временная функция сигнала

Сегментация «вокализованный / шумовой / взрывной» (блоки 2 и 6, рис. 1)

Разделение активных участков речи на шумовые, вокализованные и взрывные производится по выявлению характерных для определенных типов признаков. К примеру, такую сегментацию можно осуществить путем формирования выборочных корреляционных функций фильтрованной от низкочастотных наводок последовательности отсчетов на малых окнах анализа.

Задачу разделения «вокализованный / шумовой / взрывной» можно разделить на два этапа. На первом этапе производится сегментация «шумовой / нешумовой», а на втором, соответственно, сегментация нешумовых фрагментов на сегменты «вокализованный / взрывной».

Для выделения шумовых звуков пороговым параметром может являться отношение среднеквадратического отклонения (СКО) от линии нуля отсчетов речевого сигнала, пропущенного через низкочастотный фильтр, к СКО отсчетов исходного речевого сигнала:

$$K = \sigma_1 / \sigma_2; \quad \sigma_1 = \sqrt{\frac{\sum_{n=1}^N y_F^2(n)}{N}}; \quad \sigma_2 = \sqrt{\frac{\sum_{n=1}^N y^2(n)}{N}}, \quad (1)$$

где σ_1 – СКО сигнала y_F (пропущенного через НЧ-фильтр исходного сигнала y); σ_2 – СКО исходного сигнала y ; N – апертюра прямоугольного окна низкочастотного фильтра (заметим, что весовая функция трансверсального фильтра может быть и сглаженной, например, соответствовать окнам Хемминга, Блекмана; при этом полоса среза НЧ-фильтра должна составлять порядка 500 Гц).

Для шумовых участков среднеквадратическое отклонение σ_2 от нуля исходного сигнала будет больше, чем СКО σ_1 фильтрованной версии сигнала, поэтому отношение K для шумовых участков заведомо меньше единицы.

Вокализованные фрагменты, которые соответствуют вокализованным звукам, выделяются из общей картины своей квазипериодичностью. Как известно, автокорреляционная функция периодического сигнала также является периодической функцией [4]. При реализации алгоритма вычисляются взаимные корреляционные функции (ВКФ) отрезка речевого сигнала, находящегося на текущем окне анализа, и отрезков такой же длины, находящихся справа и слева от исследуемого. Такой подход позволяет более точно найти позиции начала и окончания вокализованного участка, нежели процесс вычисления АКФ от участка речевого сигнала с более широким окном анализа.

Для оценки периодичности следования экстремумов находятся максимумы и минимумы полученной ВКФ, вычисляется оценка среднего интервала их следования. Затем от этой оценки вычисляется оценка СКО интервалов между экстремумами. Для оценки относительной сте-

пени периодичности следования экстремумов используется отношение найденных оценок СКО и среднего интервала следования. Пример корреляционной функции для вокализованных участков показан на рис. 4. На рис. 4а приняты следующие обозначения: N – количество локальных максимумов и минимумов ВКФ; M – оценка среднего интервала; SD – оценка СКО интервалов от среднего значения.

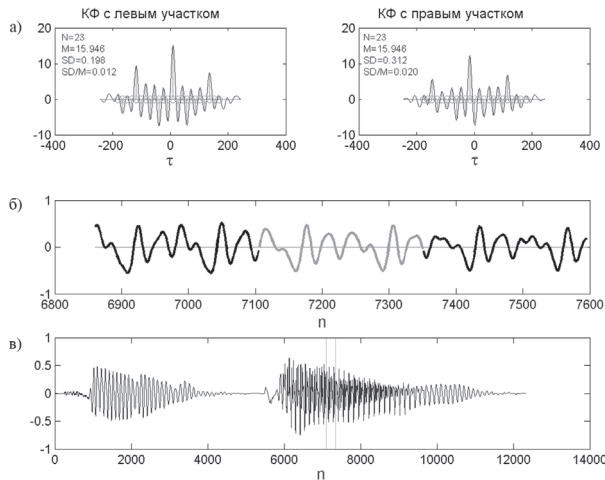


Рис. 4. Выделение вокализованных фрагментов речи: а) ВКФ исследуемого вокализованного участка речевого сигнала с двумя смежными вокализованными участками; б) временные функции исследуемого (в центре серый) участка и двух смежных; в) временная функция РС с указанием границ фрагмента, соответствующего эпюре «б» рисунка

Сегментация на периоды основного тона (блок 5, рис. 1)

В зависимости от выбранных функциональных методов обработки речевого сигнала может также потребоваться дополнительная сегментация вокализованных фрагментов на отдельные периоды основного тона, характеризующие колебания голосовых связок диктора.

Для сегментации вокализованных фрагментов РС периодов основного тона существует большое количество алгоритмов, работа которых основана на разных математических подходах [5].

Среди методов оценки периода ОТ можно выделить две основные подгруппы алгоритмов [3]: структурные методы оценки периода ОТ (локальные методы) и алгоритмы без привязки к временной структуре РС (интегральные методы).

При реализации корреляционного метода сегментации на периоды ОТ из речевого фрагмента берется участок длительностью не менее двух максимально возможных для человеческого голоса периодов ОТ (для человека T_{MAX} имеет значение ~ 20 мс) симметрично по времени относительно известной (или предполагаемой – на первом шаге работы алгоритма) левой границы периода ОТ.

В автокорреляционной функции этого фрагмента ищется максимум, лежащий в пределах

возможного периода основного тона человека, то есть между значениями 2,5 и 20 мс (рис. 5б). Смещение, соответствующее этому максимуму, и является оценкой среднего для текущего фрагмента периода основного тона. Так как это средняя величина, она нуждается в коррекции: правая граница периода ОТ вычисляется прибавлением к левой границе длительности оценки среднего периода ОТ с коррекцией до точки ближайшего пересечения нуля в направлении снизу вверх (для сегментации предлагается руководствоваться переходами сигнала через ноль в направлении снизу вверх) пропущенного через ФНЧ сигнала. Завершающим шагом коррекции должен стать поиск ближайшего к положению найденной границы перехода через ноль снизу вверх уже исходным сигналом. На рис. 5а левая серая черта – известная левая граница периода ОТ, правая серая черта – полученная непосредственно по оценке среднего периода основного тона правая граница периода ОТ. Черная черта – скорректированная в ноль правая граница. Таким образом, расстояние между левой серой чертой и черной чертой является оценкой текущего периода основного тона, а расстояние между серыми чертами – средним периодом основного тона исследуемого участка речевого сигнала.

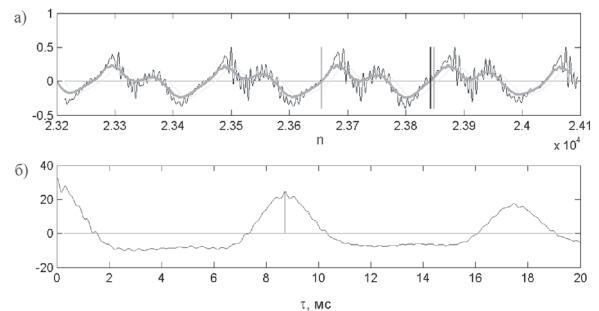


Рис. 5. Выделение периодов ОТ: а) исследуемый фрагмент сигнала; б) автокорреляционная функция исследуемого фрагмента

Сегментация вокализованных фрагментов на фонемы

В ряде функциональных приложений обработки речи может также возникнуть задача фонемной сегментации речевых сигналов. Проблема такой сегментации заключается в том, что на вокализованные звуки приходится в среднем три четверти общей длительности активной речи [14]. Поэтому различные вокализованные звуки зачастую следуют в речи непрерывно друг за другом. При этом переходы между отдельными звуками с сигнальной точки зрения могут представлять собой как плавное изменение структуры и «энергетики» колебания основного тона говорящего, так и относительно быструю их перестройку.

Для характеристики изменений структуры кластеров основного тона (ОТ-кластеров) можно ввести метрики, учитывающие разницу

«энергетик» и структур двух сравниваемых кластеров [14]. При этом плавные переходы внутри одного звука и между звуками являются трендами, а сравнительно резкие переходы между звуками – разладами структуры РС. Рассмотрим два вида метрик: амплитудно-структурную и структурную.

Амплитудно-структурные тренды выражаются как в амплитудных, так и в структурных изменениях периодов ОТ:

$$M_{амп-стр}(j) = \frac{1}{T_{от}(j)A_{cp}(j)} \sum_{i=1}^{T_{от}(j)} |X(i, j) - X(i, j-1)|, \quad (2)$$

где $M(j)$ – последовательность метрик, полученная на последовательности периодов ОТ $T_{от}(j-2) \Rightarrow T_{от}(j-1) \Rightarrow T_{от}(j) \Rightarrow T_{от}(j+1) \Rightarrow T_{от}(j+2) \Rightarrow \dots$; $T_{от}(j)$ – длительность j -го периода ОТ; $A_{cp}(j)$ – средняя амплитуда группы периодов ОТ (группа состоит из 2–3 периодов), вычисленная на скользящем временном интервале для j -го периода ОТ.

Структурные тренды выражаются в структурных изменениях периодов ОТ:

$$M_{стр}(j) = \frac{1}{T_{от}(j)} \sum_{i=1}^{T_{от}(j)} \left| \frac{X(i, j)}{X_{max}(j)} - \frac{X(i, j-1)}{X_{max}(j-1)} \right|, \quad (3)$$

где $X_{max}(j)$ – максимальный отсчет на интервале j -го ОТ-кластера.

По формулам (2) и (3) видно, что для вычисления предложенных метрик длины векторов

отсчетов речевого сигнала, соответствующих периодам ОТ, должны быть равны. Однако в общем случае в реальном РС длины двух смежных векторов ОТ немного отличаются, поэтому перед вычислением метрик необходимо уравнивать длины двух исследуемых векторов путем векторной интерполяции.

В практических приложениях могут быть более полезны модифицированные методы анализа трендов и разладок. Например, для исключения случайных изменений структур и длительностей от одного периода ОТ к другому следует вычислять метрики не смежных периодов, а разнесенных друг от друга на несколько промежуточных периодов. В этом случае значения метрик в трендах станут несколько выше по величине (так как в тренде дальше отстоящие друг от друга периоды ОТ будут сильнее отличаться и, следовательно, иметь большую метрику), а огибающая метрик станет «плавнее».

ЗАКЛЮЧЕНИЕ

Системный подход к сегментации речевых сигналов позволяет получать необходимую для реализации речевых приложений детальность сегментации вплоть до выделения и параметрического описания отдельных периодов основного тона диктора. Приведенные в статье алгоритмы многоуровневой временной сегментации относятся к классу технологических и могут использоваться в широком спектре речевых приложений.

СПИСОК ЛИТЕРАТУРЫ

- Архипов И. О., Гитлин В. Г. Оценка точности выделения основного тона методом GS // Современные речевые технологии: Сборник трудов IX сессии Российского акустического общества. М.: ГЕОС, 1999. С. 38–42.
- Зилинберг А. Ю., Корнеев Ю. А. Разработка и исследование временных и спектральных алгоритмов VAD (Voice Activity Detection) // Российская школа-конференция «Мобильные системы передачи данных». М.: МИЭТ, 2006. С. 58–70.
- Кипяткова И. С. Комплекс программных средств обработки и распознавания разговорной русской речи // Информационно-управляющие системы. 2011. № 4. С. 53–59.
- Кипяткова И. С., Карпов А. А. Автоматическая обработка и статистический анализ новостного текстового корпуса для модели языка системы распознавания русской речи // Информационно-управляющие системы. 2010. № 4. С. 2–8.
- Маркел Дж. Д., Грэй А. Х. Линейное предсказание речи: Пер. с англ. М.: Связь, 1980. 308 с.
- Михайлов В. Г., Златоустова Л. В. Измерение параметров речи. М.: Радио и связь, 1987. 168 с.
- Ронжин А. Л., Глазков С. В. Метод автоматического распознавания голосовых команд и неречевых акустических событий // Информационно-управляющие системы. 2012. № 4. С. 74–77.
- Ронжин А. Л., Карпов А. А., Кагиров И. А. Особенности дистанционной записи и обработки речи в автоматах самообслуживания // Информационно-управляющие системы. 2009. № 5. С. 32–38.
- Ронжин А. Л., Карпов А. А., Лобанов Б. М., Цирульников Л. И., Йокиш О. Фонетико-морфологическая разметка речевых корпусов для распознавания и синтеза русской речи // Информационно-управляющие системы. 2006. № 6. С. 24–34.
- Савченко В. В., Акатьев Д. Ю. Результаты экспериментальных исследований методики формирования фонетической базы данных диктора из непрерывного потока его разговорной речи // Информационно-управляющие системы. 2012. № 6. С. 38–42.
- Савченко В. В., Савченко А. В. Метод фонетического декодирования слов в информационной метрике Кульбака – Лейблера для систем автоматического анализа и распознавания речи с повышенным быстродействием // Информационно-управляющие системы. 2013. № 2. С. 7–12.
- Свириденко В. А. Аутентификация личности по голосу // Мобильные системы. 2004. № 2 [Электронный ресурс]. Режим доступа: http://www.spirit.ru/articles/svi_mb.html
- Старченко И. Б., Вишневецкий В. Ю. Практикум по курсу «Математическое моделирование биологических процессов и систем». Таганрог: Изд-во ТТИ ЮФУ, 2010. 36 с.
- Томчук К. К., Зилинберг А. Ю., Корнеев Ю. А. Анализ трендов и разладок структуры ОТ-кластеров вокализованных сегментов речи // Сборник докладов Научной сессии ГУАП. СПб.: ГУАП, 2010. С. 70–73.
- Хованова Н. А., Хованов И. А. Методы анализа временных рядов. Саратов: ГосУНЦ «Колледж», 2001. 120 с.

Tomchuk K. K., Saint-Petersburg State University of Aerospace Instrumentation (St. Petersburg, Russian Federation)
Zilinson A. Yu., Saint-Petersburg State University of Aerospace Instrumentation (St. Petersburg, Russian Federation)

HIERARCHICALLY STRUCTURED MODEL AND BASIC ALGORITHMS OF TIME DOMAIN SEGMENTATION OF SPEECH SIGNALS

This paper introduces a systematic approach to speech signal (SS) segmentation based on developed hierarchically structured SS-model. Ground is given to application of multilevel time voice sound record segmentation, and basic algorithms of principal segmentation stages are developed: VAD algorithm; “voiced / noise / plosive” segmentation; voiced fragments’ segmentation into separate pitch periods. The possible variant of splitting adjacent voiced sounds using analysis of their structure has been proposed.

Key words: speech, model of speech signal, segmentation, VAD, pitch

REFERENCES

1. Arkhipov I. O., Gitlin V. G. Evaluation of the accuracy of pitch by GS metod [Otsenka tochnosti vydeleniya osnovnogo tona metodom GS]. *Sovremennye rechevye tekhnologii: Sbornik trudov IX sessii Rossiyskogo akusticheskogo obshchestva*. Moscow, GEOS Publ., 1999. P. 38–42.
2. Zilinson A. Yu., Korneev Yu. A. Development and research of temporal and spectral algorithms VAD [Razrabotka i issledovanie vremennykh i spektral'nykh algoritmov VAD (Voice Activity Detection)]. *Rossiyskaya shkola-konferentsiya “Mobil'nye sistemy peredachi dannykh”*. Moscow, MIJeT, 2006. P. 58–70.
3. Kipyatkova I. S. Software complex of processing and recognition of conversational Russian speech [Kompleks programnykh sredstv obrabotki i raspoznavaniya razgovornoy russkoy rechi]. *Informatsionno-upravlyayushchie systemy*. 2011. № 4. P. 53–59.
4. Kipyatkova I. S., Karpov A. A. Automatic processing and statistical analysis of a news text for the language model of Russian speech recognition [Avtomaticheskaya obrabotka i statisticheskiy analiz novostnogo tekstovogo korpusa dlya modeli yazyka sistemy raspoznavaniya russkoy rechi]. *Informatsionno-upravlyayushchie systemy*. 2010. № 4. P. 2–8.
5. Markel J. D., Grey A. H. *Lineynoe predskazanie rechi* [Linear prediction of speech]. Moscow, Svyaz' Publ., 1980. 308 p.
6. Mihaylov V. G., Zlatoustova L. V. *Izmerenie parametrov rechi* [Measuring speech parameters]. Moscow, Radio i svyaz' Publ., 1987. 168 p.
7. Ronzhin A. L., Glazkov S. V. The method of automatic recognition of voice commands and non-speech acoustic events [Metod avtomaticheskogo raspoznavaniya golosovykh komand i nerechevykh akusticheskikh sobyitiy]. *Informatsionno-upravlyayushchie systemy*. 2012. № 4. P. 74–77.
8. Ronzhin A. L., Karpov A. A., Kagiroy I. A. Particular qualities of remote recording and processing of speech in self-service machine [Osobennosti distantsionnoy zapisi i obrabotki rechi v avtomatakh samoobslyuzhivaniya]. *Informatsionno-upravlyayushchie systemy*. 2009. № 5. P. 32–38.
9. Ronzhin A. L., Karpov A. A., Lobanov B. M., Tsurul'nik L. I., Jokish O. Phonetic and morphological marking of speech cases for recognition and synthesis of Russian speech [Fonetiko-morfologicheskaya razmetka rechevykh korpusov dlya raspoznavaniya i sinteza russkoy rechi]. *Informatsionno-upravlyayushchie systemy*. 2006. № 6. P. 24–34.
10. Savchenko V. V., Akat'ev D. Yu. Results of experimental research of method of forming a phonetic database speaker of continuous flow of his speech [Rezultaty eksperimental'nykh issledovaniy metodiki formirovaniya foneticheskoy bazy dannykh diktora iz nepreryvnogo potoka ego razgovornoy rechi]. *Informatsionno-upravlyayushchie systemy*. 2012. № 6. P. 38–42.
11. Savchenko V. V., Savchenko A. V. The method of phonetic decoding of words in information metric of Kullback-Leibler for automatic analysis and speech recognition with high performance [Metod foneticheskogo dekodirovaniya slov v informatsionnoy metrike Kul'baka – Leyblera dlya sistem avtomaticheskogo analiza i raspoznavaniya rechi s povyshennym bystrodeystviem]. *Informatsionno-upravlyayushchie systemy*. 2013. № 2. P. 7–12.
12. Sviridenko V. A. Authenticate of personality on the voice [Autentifikatsiya lichnosti po golosu]. *Mobil'nye sistemy*. 2004. № 2. Available at: http://www.spirit.ru/articles/svi_mb.html
13. Starchenko I. B., Vishnevetskii V. Yu. Workshop on the course “Mathematical modeling of biological processes and systems” [Praktikum po kursu “Matematicheskoe modelirovanie biologicheskikh protsessov i system”]. Taganrog, TTI YFU, 2010. 36 p.
14. Tomchuk K. K., Zilinson A. Yu., Korneev Yu. A. The analysis of trend and disorders in the structure of clusters containing voiced speech segments [Analiz trendov i razladok struktury OT-klastrov vokalizovannykh segmentov rechi]. *Sbornik dokladov Nauchnoy sessii GUAP*. St. Petersburg, GUAP, 2010. P. 70–73.
15. Khovanova N. A., Khovanov I. A. Methods of time series analysis [Metody analiza vremennykh ryadov]. Saratov, GosUNC “Kolledzh”, 2001. 120 p.

Поступила в редакцию 24.01.2013