

АЛГОРИТМЫ СЕГМЕНТАЦИИ С ПРИМЕНЕНИЕМ БЫСТРОГО ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ

Т. Ермоленко, В. Шевчук

Аннотация: в статье описаны два алгоритма сегментации речевого сигнала с использованием вейвлет-преобразования. Даны их сравнительные характеристики.

Ключевые слова: вейвлет-преобразование, детализация, аппроксимация, сегментация.

Введение

Как известно, речевой сигнал состоит из квазистационарных участков, соответствующих голосовым и шипящим фонемам, перемежаемых участками со сравнительно быстрыми изменениями спектральных характеристик сигнала (межфонемные переходы, взрывные и смычные фонемы, внутрисловные переходы речь-пауза). В пределах стационарных участков значительную роль для анализа речевого сигнала играют спектральные особенности сигнала, определяемые передаточной характеристикой речевого тракта, изменяющейся в процессе артикуляции. Можно сказать, что речевой сигнал характеризуется нелинейными флуктуациями различных масштабов. Поэтому весьма эффективным для анализа речевого сигнала представляется кратномасштабный анализ и вейвлет – преобразование.

Сегментация речевого сигнала (РС) подразумевает выделение участков сигнала, соответствующим отдельным структурным единицам РС. Если в качестве таких единиц рассматривать фонемы, то задача сегментации сводится к обнаружению межфонемных переходов. В рамках традиционных подходов решение этой задачи весьма проблематично. Однако ВП позволяет решить эту проблему по крайней мере для фонем, соответствующих сравнительно протяженным квазистационарным участкам РС. Дело в том, что на межфонемных переходах сигнал претерпевает значительные изменения сразу на многих масштабах исследования, и, соответственно, характеризуется возрастанием вейвлет-коэффициентов для многих уровней детализации, в то время, как на стационарных участках фонем вейвлет-коэффициенты оказываются сгруппированными вблизи определенных масштабов. Таким образом, отыскание межфонемных границ может быть сведено к отысканию моментов увеличения вейвлет-коэффициентов на значительном количестве уровней масштабирования. При этом существенным является выбор вейвлетного базиса, который должен позволять описывать стационарный РС со сравнительно малым числом ненулевых коэффициентов. Возможно использование нескольких вейвлетных базисов для поиска межфонемных переходов в каждом из них с последующим объединением результатов.

Алгоритм 1

Разложение по вейвлетам РС длины N отсчетов представляет собой сумму

$$f(t) = \sum_{k=0}^{\frac{N}{2^n}-1} s_{nk} \varphi_{nk} + \sum_{j=1}^n \sum_{k=0}^{\frac{N}{2^j}-1} d_{jk} \psi_{jk} \quad (1)$$

при этом n является уровнем детализации, s_{jk}, d_{jk} - коэффициенты вейвлет-разложения, называемые средними и разностями соответственно (в этой работе будем именовать их коэффициентами аппроксимации и детализации), $\varphi_{jk} = 2^{j/2} \varphi(2^j t - k), j, k \in Z$, φ - скейлинг-функция или масштабная функция, $\psi_{jk} = 2^{j/2} \psi(2^j t - k), j, k \in Z$, ψ - базисный или «материнский» вейвлет.

В своих исследованиях мы использовали быстрое вейвлет-преобразование РС, который раскладывали по 6 уровням, полагая $s_{0,k}$ равными отсчетам исходного сигнала. РС, оцифрованный с частотой дискретизации 22 050 Гц, разбивается на перекрывающиеся окна размером 20 мс с половинным перекрытием окна.

Экспериментальным путем был получен следующий критерий для выбора наиболее информативного уровня разложения. Для J -го уровня разложения, начиная с 3-го, проверяется выполнение неравенства:

$$\frac{E}{N} < \frac{E_j}{N_j}, \quad (2)$$

где N_j - количество коэффициентов детализации на J -том уровне, больших, чем 0,5;

$$E = \sum_{i=0}^{N-1} s_{0,i}^2;$$

$$E_j = \sum_{i=0}^{\frac{N}{2^j}-1} d_{j,i}^2.$$

Первый уровень, для которого выполняется условие (2), является наиболее информативным.

Анализировали поведение коэффициентов детализации следующим образом: на выбранном уровне разложения J строилась числовая последовательность $\{e_{ij}\}_{i=1}^{N/256}$:

$$e_{ij} = 10 \lg \sum_{k=0}^{n_j-1} d_{j,i+k}^2,$$

где i - номер скользящего окна, $n_j = \frac{n}{2^j}$ - размер скользящего окна на J -том уровне, n - размер окна в исходном сигнале (512 отсчетов).

Границы предполагаемых сегментов проставлялись между окнами с номерами i и $i+1$, для которых $|e_{i+1,j} - e_{i,j}| \geq 3.5$.

Итоги тестирования работы алгоритма таковы.

1. Взрывные согласные, стоящие перед голосовым, делятся на 2 сегмента, соответствующие паузе и взрыву
2. Переход от голосовых к шипящим выделяется в отдельный сегмент;
3. Хорошо выделяются звуки -р-, -ц-;
4. В ударном и безударном -а- может выделяться лишний сегмент;
5. Не разделяются два голосовых звука, стоящие рядом, например, -оа-, -ао-, -ио- в словах *коала*, *аорта*, *миллион*; характерное окончание -ия-, например, *квалификация*, *аппроксимация*, также не сегментируется, что объяснимо, поскольку -ия- нивелируется и звучит, как безударный -а-. В таких словах, как *миля*, *миллион*, -ми- является одним сегментом.
6. В случае сочетания -к- с шипящим взрыв от -к- объединяется с шипящим.

На рисунке 1 показан результат работы алгоритма на слове *акселерация*. Как видно из рисунка, сегмент 1 соответствует безударному -а-. Достаточно четко отделена пауза от речи. Сегмент 2 содержит паузу звука -к-. В 3-ем сегменте объединены взрыв от -к- и звук -с-. В 4-ом - безударный -е- и -л-, что в какой-то мере объяснимо: голосовой звук редуцируется, имеет короткую длительность и теряет свои

качества. Второй безударный звук –е- хорошо отделен границами 5-го сегмента, очевидно, это связано с тем, что он является первым предупредарным гласным. Сегмент 6 соответствует –р-, 7-й – ударному –а-. 9-й сегмент содержит звук –ц-, а сегмент 8 - паузу, предшествующую этому звуку, что полностью согласуется с правилами фонетики. 10-й сегмент заключает в себе безударное окончание –ия-.

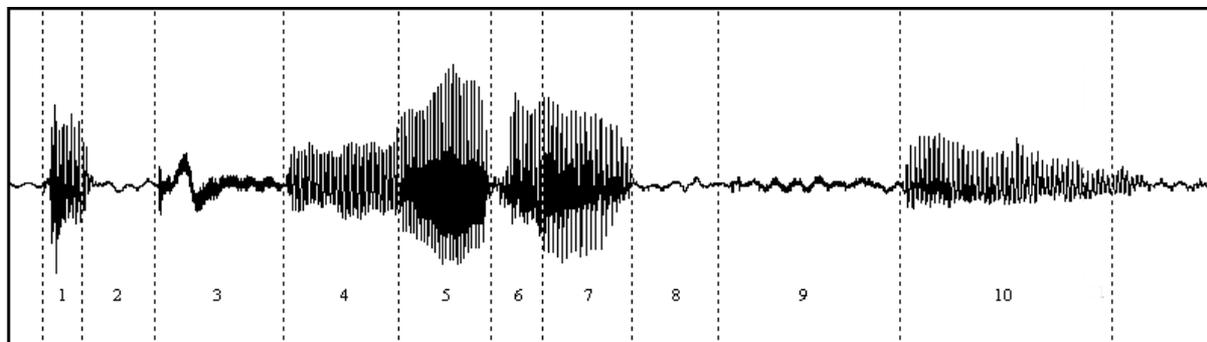


Рисунок 1. Сегментация слова *акселерация* по первому алгоритму.

Алгоритм 2

На каждом уровне детализации вычисляем

$$E_j = \sum_{i=0}^{\frac{N}{2^j}-1} d_{j,i}^2, \text{ где } j - \text{ номер уровня разложения.}$$

Строим числовую последовательность $\{\tilde{e}_{ij}\}_{i=1}^{N/256}$:

$$\tilde{e}_{ij} = \frac{\sum_{k=0}^{n_j-1} d_{j,i+k}^2}{E_j},$$

где i - номер скользящего окна, $n_j = \frac{n}{2^j}$ - размер скользящего окна на j -том уровне.

Далее производим усреднение по всем уровням детализации:

$$\tilde{e}_i = \frac{\sum_{j=1}^6 \tilde{e}_{ij}}{6}.$$

Сегментация по этому алгоритму происходит в 2 этапа.

Этап первый – определение шипящих и пауз. Для этого вводится порог

$$p = \max_i(\tilde{e}_i) * \text{Average}(\tilde{e}_i) / 10 + \min_i(\tilde{e}_i) * 10, \text{ где } \text{Average}(\tilde{e}_i) - \text{ среднее}$$

арифметическое последовательности $\{\tilde{e}_i\}_{i=1}^{N/256}$. Граница между шипящей (паузой) и остальными звуками является правой границей i -того окна, для которого выполняется условие

$$(\tilde{e}_i - p)(\tilde{e}_{i+1} - p) < 0.$$

На втором этапе находятся границы на участках сигнала, которым соответствуют $\tilde{e}_i > p$.

Ищутся локальные минимумы и максимумы последовательности $\{\tilde{e}_i\}$ в окрестности двух точек на

исследуемых участках. Правые границы окон, отвечающих этим максимумам (минимумам) являются границами предполагаемых сегментов.

Рисунок 2 иллюстрирует работу этого алгоритма на примере слова *акселерация*.

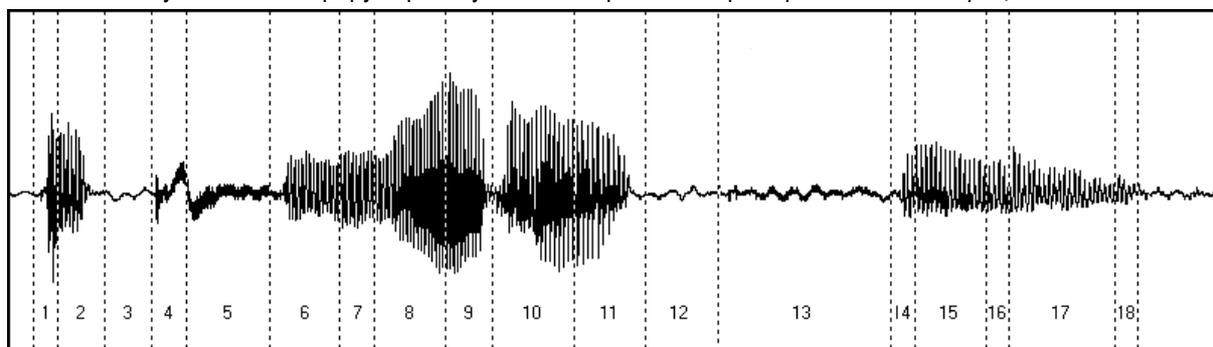


Рисунок 2. Сегментация слова *акселерация* по второму алгоритму.

Как видно из рисунка, второй алгоритм более чувствителен, позволяет выделять кратковременные изменения в сигнале, такие как взрывная часть взрывных звонких и глухих согласных (сегмент 4), и переходные участки между звуками (сегменты 14, 16, 18).

Выводы

Описанные выше алгоритмы дают несколько отличные результаты и могут быть использованы как два независимых критерия выделения границ. Границы, совпадающие с точностью до некоторой заранее определенной окрестности, следует считать правомерными. Для оценки правомерности остальных границ следует учитывать относительную амплитуду участка, частотные признаки, длительность сегмента и аналогичные признаки ближайших соседних участков.

Литература

1. Дремин И. М., Иванов О. В., Нечитайло В. А. Вейвлеты и их использование. //Успехи физических наук, т. 171, №5 с. 465-500, 2002г.
2. Астафьева Н. М. Вейвлет-анализ: основы теории и примеры применения. //Успехи физических наук, т. 166, №11 с.1145-1170, 1996г.
3. Walker J. Fourier Analysis and Wavelet Analysis. //Notices of the AMS, vol. 44 №6, p. 658-670, 1997
4. Daubechies I. Ten lectures on wavelets. // Philadelphia: SIAM, 1991