

УДК 004.934

А.Н. Ручай

Модифицированный метод сегментации речевого сигнала на основе непрерывного вейвлет-преобразования

Описан модифицированный метод сегментации речевого сигнала на непересекающиеся вокализованные сегменты на основе непрерывного вейвлет-преобразования. Данный метод основан на оценке показателя Гёльдера с помощью непрерывного вейвлет-преобразования. Оцениваются точность и устойчивость модифицированного метода сегментации речевого сигнала. Модифицированный метод сегментации речевого сигнала сравнивается со стандартными методами сегментации.

Ключевые слова: автоматическая сегментация речевого сигнала, сингулярность, показатель Гёльдера, непрерывное вейвлет-преобразование.

Сегментация речевого сигнала является одной из важнейших задач в системах автоматической обработки речи [1]. Сегментация необходима при решении не только задачи распознавания речи и выделения характерных признаков голоса на определённых сегментах речевого сигнала, но и обратной задачи – восстановления формы речевого тракта по акустическому сигналу, которая может быть использована в сжатии и передаче речи, синтезе речи по произвольному тексту и автоматическом распознавании речи [2, 3].

В исследовательских целях или на этапе предварительной разработки возможно использование ручной сегментации [4]. Однако она требует значительных усилий и временных затрат. Кроме того, практически невозможно точно воспроизвести результаты ручной сегментации вследствие субъективности человеческого зрительного и слухового восприятия. Подобных проблем не возникает при автоматической сегментации, которая также не безошибочна, но даёт воспроизводимые результаты [4].

Существует два основных типа автоматической сегментации речи [4]. К первому типу относят сегментацию речи при условии, что известна последовательность фонем данной фразы. Другой тип не использует априорной информации о фразе, при этом границы сегментов определяются по степени изменения акустических характеристик сигнала. При автоматической сегментации желательно использовать только общие характеристики речевого сигнала, поскольку обычно на этом этапе нет конкретной информации о содержании речевой фразы.

В работах [5, 6] был предложен новый метод сегментации речевого сигнала. В новом предложенном методе сигнал сегментируется на непересекающиеся вокализованные сегменты, которые соответствуют не фонемам слова, а слогам, в основе которых лежат периодические гласные звуки. Предлагаемый метод с помощью непрерывного вейвлет-преобразования оценивает показатель Гёльдера, взятый в качестве показателя сингулярности сигнала.

В статье [6] показывается, что предложенный новый метод успешно разбивает речевой сигнал на непересекающиеся вокализованные сегменты и может быть применен к различным задачам. Одной из таких задач является текстозависимая верификация диктора на основе формантного метода, где с помощью предложенного метода сегментации сигнала ошибка первого рода уменьшилась на 20% при фиксированной ошибке второго рода для распознавания диктора по сравнению со стандартным методом покадровой обработки.

В данной статье были успешно сделаны попытки улучшить предлагаемый метод сегментации речевого сигнала на основе непрерывного вейвлет-преобразования, который будем называть модифицированным методом сегментации речевого сигнала на основе непрерывного вейвлет-преобразования.

Модификация метода сегментации на основе непрерывного вейвлет-преобразования. Для оценки качества работы предложенного метода сегментации речевого сигнала на вокализованные участки на основе непрерывного вейвлет-преобразования из статьи [6] был проведен эксперимент. Для этого была использована голосовая база, состоящая из 100 дикторов, каждый диктор произносил 13 раз некоторое одинаковое для всех слово, содержащее 5 гласных звуков. Ко всем фразам из базы был применен предложенный метод разбиения речевого сигнала на вокализованные сегменты.

В результате были успешно выделены вокализованные сегменты, которые полностью соответствовали 5 слогам в этом слове.

Проводились также эксперименты по оценке качества сегментации с помощью предложенного метода с голосовой базой, состоящей из 20 дикторов, каждый из которых произносит 20 раз фразу, соответствующую 5 слогам. Однако выделенные окончания срезают вокализованный сегмент, поэтому было положено в качестве одновременно и окончания, и начала взять среднее между окончанием одного сегмента и началом следующего сегмента.

Были продолжены попытки улучшить предлагаемый метод сегментации на основе непрерывного вейвлет-преобразования. Идея удаления из сигнала неречевых участков с помощью оценки кратковременной энергии перед выполнением алгоритма сегментации на основе непрерывного вейвлет-преобразования не дала улучшения качества и устойчивости сегментации.

Опишем следующий модифицированный метод сегментации речевого сигнала на основе непрерывного вейвлет-преобразования, который показал лучшие результаты. Речевой сигнал грубо разбивается на сегменты, соответствующие голосовому возбуждению, методом, основанным на отношении кратковременной энергии и числа перехода через нуль, затем из каждого сегмента удаляются неречевые участки методом, основанным на оценке кратковременной энергии. После чего применяется отдельно для каждого сегмента предлагаемый метод сегментации на основе непрерывного вейвлет-преобразования.

Оценка и сравнение модифицированного метода сегментации на основе непрерывного вейвлет-преобразования. Модифицированный метод сегментации успешно справился с выделением вокализованных участков в экспериментах с вышеописанными голосовыми базами, поэтому в дальнейших экспериментах использовалась голосовая база, состоящая из 100 дикторов, каждый из которых произносил 13–15 раз стихотворение «Буря мглою небо кроет, вихри снежные крутя; то, как зверь, она завоет, то заплачет, как дитя». На рис. 1–4 показаны результаты работы данного модифицированного метода сегментации на основе непрерывного вейвлет-преобразования, откуда можно заметить, что модифицированный метод успешно справился с разбиением сигнала в соответствии с гласными звуками.

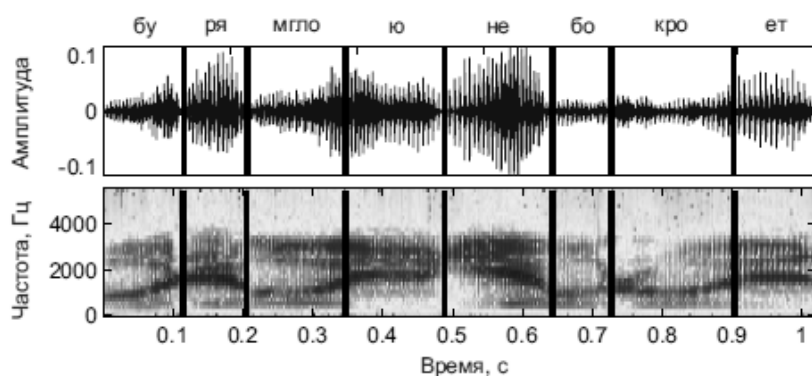


Рис. 1. Результат сегментации фразы «Буря мглою небо кроет» модифицированным методом

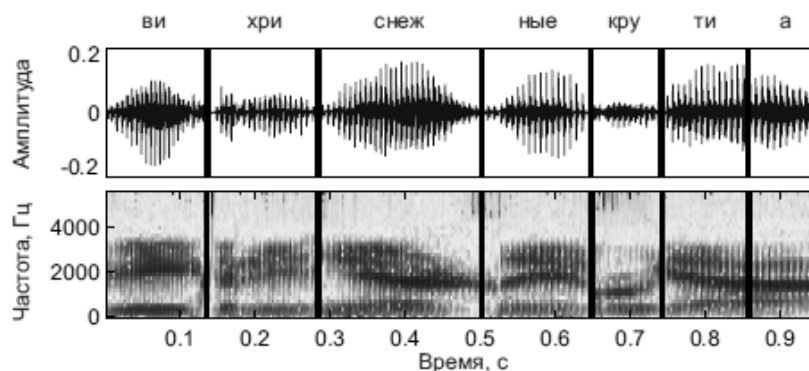
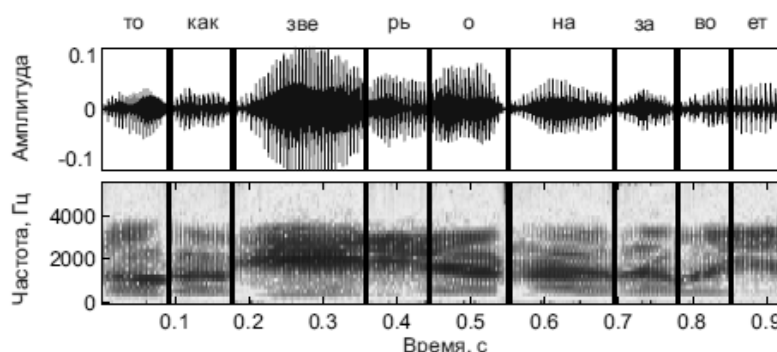


Рис. 2. Результат сегментации фразы «Вихри снежные крутя» модифицированным методом

Модифицированный метод сегментации на основе непрерывного вейвлет-преобразования успешно разбил фразы на вокализованные сегменты.

Однако из рис. 1–4 видно, что в ударном положении буква «я» была разбита на два звука, в безударном – букве соответствовал только один сегмент. Кроме того, пары гласных не были разделены, например «ою» и «ые», стоящие сразу после согласного. Также на рисунке 3 можно увидеть, что звук [р'] был выделен в отдельный сегмент. Все фразы всех дикторов успешно прошли проверку на соответствие границ с помощью ручной сегментации.

Рис. 3. Результат сегментации фразы «То, как зверь, она завоет» модифицированным методом



Проводились также эксперименты по оценке устойчивости к шумам модифицированного метода сегментации на основе непрерывного вейвлет преобразования. Используя модифицированный метод сегментации, были получены значения границ вокализованных участков, то есть эталонные значения границ. Затем на речевые сигналы накладывался аддитивный белый гауссовский шум с разным уровнем отношения сигнала к шуму SNR, и затем оценивалось среднее отклонение вычисленных значений границ от эталонных. Эксперимент проводился 100 раз, значения отклонений усреднялись по всем фразам дикторов.

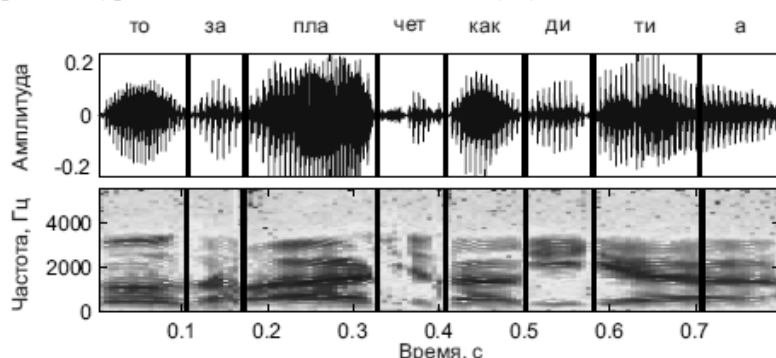


Рис. 4. Результат сегментации фразы «То заплачет, как дитя» модифицированным методом

В таблице представлены результаты экспериментов в зависимости от SNR. На основании результатов можно сделать вывод, что приемлемые результаты средних отклонений можно ожидать только при SNR, равном 30 дБ. Помимо этого, при SNR ниже 30 дБ изменяется количество сегментов. Из чего следует, что модифицированный метод сегментации устойчив к уровню шума и помех с SNR более 30 дБ.

Среднее отклонение границ сегментов в мс и ошибка числа сегментов в % от SNR в дБ

SNR (дБ)	50	40	30	20	15
Среднее отклонение (мс)	7,54	11,78	16,38	25,27	43,35
Ошибка числа сегментов (%)	0,538	1,769	3,615	9,461	15,153

Для сравнения модифицированного метода сегментации со стандартными были реализованы два метода: первый метод основан на отношении квадратного корня из кратковременной энергии к функции перехода через ноль, который назовем стандартным, второй метод основан на вычисление автокорреляционной функции. На основании полученных предварительных результатов можно сделать вывод, что второй метод сегментирует не корректно. Поэтому сравнение дальше производилось только со стандартным методом.

Стандартный метод сегментации показал ряд некорректно сегментируемых участков. Например, слова «буря», «мглою», «небо» не были разделены на отдельные сегменты, и, наоборот, в отдельные сегменты были выделены звуки [х], [ж], [р'] и [т].

Полученные результаты позволяют говорить о хороших результатах работы модифицированного метода сегментации на основе непрерывного вейвлет-преобразования по сравнению со стандартными методами. Однако стандартные методы более устойчивы к шумам, устойчивость сохраняется даже при SNR, равным 20 дБ.

Заключение. Разработан и модифицирован метод сегментации речевого сигнала на основе непрерывного вейвлет преобразования, который дал лучшие результаты. На основании полученных оценок точности и устойчивости модифицированного метода сегментации можно сделать вывод, что данный метод корректно выделяет локализованные сегменты речевого сигнала, однако могут быть выделены некоторые специфические сегменты, которые могут быть специальным образом обработаны и учтены, что требует дальнейших улучшений предложенного алгоритма сегментации.

Экспериментально установлено, что модифицированный метод сегментации устойчив только к небольшому уровню шума и помех. Модифицированный метод показал высокие результаты сегментации в сравнении со стандартными методами, хотя стандартные методы более устойчивы к шуму.

Также стоит отметить, что могут быть продолжены попытки модифицировать метод сегментации на основе непрерывного вейвлет-преобразования для выделения участков, содержащих отдельные фонемы, что также требует дальнейших исследований.

Литература

1. Рабинер Л.Р. Цифровая обработка речевых сигналов / Л.Р. Рабинер, Р.В. Шафер. – М.: Радио и связь, 1981. – 496 с.
2. Сорокин В.Н. Сегментация и распознавание гласных / В.Н. Сорокин, А.И. Цыплихин // Информационные процессы. – 2004. – Т. 4, № 2. – С. 202–220.
3. Цыплихин А.И. Анализ и автоматическая сегментация речевого сигнала: дис. ... канд. тех. наук / ИППИ РАН. – М., 2006. – 149 с.
4. Вишнякова О.А. Автоматическая сегментация речевого сигнала на базе дискретного вейвлет-преобразования / О.А. Вишнякова, Д.Н. Лавров // Математические структуры и моделирование. – 2011. – Вып. 23. – С. 43–48.
5. Ручай А.Н. Текстозависимая верификация диктора: математическая модель, статистические исследования, комплекс программ. – Saarbrücken: LAP LAMBERT Academic Publishing, 2012. – 144 с.
6. Ручай А.Н. Улучшение надежности формантного метода текстозависимой верификации диктора с помощью нового метода сегментации сигнала // Доклады ТУСУРа. – №2(24). – 2011. – С. 241–246.

Ручай Алексей Николаевич

Канд. физ.-мат. наук, ст. преподаватель каф. компьютерной безопасности и прикладной алгебры
Челябинского государственного университета
Тел.: 8 (351) 977-92-92
Эл. почта: ruchai@pochta.ru

Ruchay A.N.

Modified method of segmentation of speech signal by continuous wavelet transform

This article describes a modified method of segmentation of the speech signal into disjoint voiced segments by the continuous wavelet transform. This method is based on an assessment of the Holder exponent using a continuous wavelet transform. The accuracy and stability of the modified method of segmentation of the speech signal are evaluated. The modified method for the segmentation of the speech signal is compared with the standard methods of segmentation.

Keywords: automatic segmentation of speech signal, singularity, Holder exponent, continuous wavelet transform.