

Медетов Б.Ж.,
Койшигарин А.С.,
Байгожина А.М.

**Сегментация сигнала
на квазистационарные участки
с помощью коэффициента
формы кривой**

Настоящая работа посвящена разработке нового метода сегментации сигнала на квазистационарные участки. Авторами работы предлагается метод выполнения сегментации сигнала на квазистационарные участки. В качестве сигнала рассматривается речь. Проведен обзор существующих методов автоматической сегментации слитного речевого сигнала на составные элементы в соответствии с фонетической транскрипцией языка. Проведен сравнительный анализ работы предлагаемого метода с известными другими методами. На основе данного анализа сделан вывод, что предлагаемый новый метод, использующий коэффициент формы кривой, дает наилучшие результаты.

Ключевые слова: речевой сигнал, сегментация слитной речи, коэффициент формы кривой, обработка сигналов.

Medetov B., Koyshegarin A.,
Baygozhina A.

**Segmentation of quasi-
stationary signal portions using
coefficient waveform**

This paper is devoted to the development of a new method of segmentation in the quasi-stationary signal portions. The authors of the proposed method of performing segmentation of the signal at the quasi-stationary areas. As the signal is considered speech. A review of existing methods for automatic segmentation of the fused speech signal into its component elements, in accordance with the phonetic transcription of the language. A comparative analysis of the proposed method with other methods known. On the basis of this analysis, it is concluded that the proposed new method, which uses the coefficient waveform gives the best results.

Key words: speech signal, segmentation of continuous speech, coefficient shape of the curve, signal processing.

Медетов Б.Ж.,
Койшигарин А.С.,
Байгожина А.М.

**Қыйсық пішінінің
коэффициенті көмегімен
сигналды квазитұрақты аралық
сегменттерге бөлу**

Бұл жұмыс сигналды квазитұрақты аралық сегменттерге бөлу-дің жаңа әдісін зерттеуге арналған. Жұмыс авторларымен сигналды квазитұрақты аралық сегменттерге бөлудің орындалу әдісі ұсынылған. Сигнал ретінде сөз қарастырылады. Тілдің фонетикалық транскрипциясына сәйкес біріккен сөздік сигналды құрама бөліктерге бөлуде автоматты түрде сегменттерге бөлудің қолданыстағы әдістеріне шолу жүргізілген. Аталған жаңа әдіс пен өзге белгілі әдістердің арасында салыстырмалы анализ жасалынған. Осы анализдің қортындысы бойынша, ұсынылған қыйсық пішінінің коэффициентін қолданатын жаңа әдіс ең жоғары нәтиже береді.

Түйін сөздер: сөздік сигнал, біріккен сөздік сигналды сегментациялау, қыйсық пішінінің коэффициенті, сигналдарды өңдеу.

СЕГМЕНТАЦИЯ СИГНАЛА НА КВАЗИСТАЦИОНАРНЫЕ УЧАСТКИ С ПОМОЩЬЮ КОЭФФИЦИЕНТА ФОРМЫ КРИВОЙ

Введение

Одной из важнейших задач в системах автоматической обработки речи является задача сегментации в соответствии с фонетической транскрипцией языка. Для голосовой верификации характерные признаки голоса должны вычисляться на определённых сегментах речевого сигнала. Так, частота основного тона, присущая диктору, должна вычисляться на гласноподобных участках сигнала, форма речевого тракта характеризуется формантными частотами, измеряемыми на известных гласных звуках, скорость артикуляции определяется по длительностям переходных процессов между артикуляторно-акустическими сегментами. Сегментация также необходима при решении обратной задачи – восстановления формы речевого тракта по акустическому сигналу, которая может быть использована в следующих областях: системы сжатия и передачи речи в мобильной телефонии, синтезаторы речи по произвольному тексту, системы автоматического распознавания речи, системы обучения иноязычному произношению.

В исследовательских системах и на этапе предварительной разработки возможно использование ручной сегментации. Однако она требует значительных затрат сил и времени: во-первых, в слитной речи нет пауз между словами, во-вторых, коартикуляция, возникающая и на границе последовательно производимых звуков, которая существенно облегчает правильное восприятие и понимание речи, но затрудняет задачу поиска границ сегментов. Кроме того, практически невозможно точно воспроизвести результаты ручной сегментации вследствие субъективности человеческого слухового и зрительного восприятия. Подобных проблем не возникает при автоматической сегментации, которая также не безошибочна, но даёт воспроизводимые результаты. Существует два основных типа алгоритмов сегментации речи [1]. К первому типу относятся алгоритмы, которые производят сегментацию речи при условии, что известна последовательность фонем данной фразы. Другой тип алгоритмов не использует априорной информации о фразе, и 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 речевого высказывания.

Сегментация на основе меры спектрального перехода

Метод автоматической сегментации, основанный на мере спектрального перехода (Spectral Transition Measure, STM), был предложен в работе [2].

Мера спектрального перехода определяется таким образом:

$$STM(f) = \frac{\sum_{i=1}^D a_i^2(f)}{D} \quad (2.2)$$

где D – размерность вектора признаков речевого сигнала, скорость изменения элементов которого рассчитывается в соответствии с формулой:

$$a_i(f) = \frac{\sum_{n=-I}^I c_i(f+n)*n}{\sum_{n=-I}^I n^2} \quad (2.3)$$

$$CSM(f) = \begin{cases} \frac{c_{max}(f)-c_{min}(f)}{\sum_{i=1}^{D-1} |c_{i+1}(f)-c_i(f)|}, & c_{max}(f) \neq c_{min}(f) \\ 1, & c_{max}(f) = c_{min}(f), \end{cases} \quad (2.4)$$

где f – номер фрейма; $c_i(f)$ – i -ый коэффициент MFCC-кепстра речевого сигнала, вычисленного для фрейма f ; $c_{max}(f)$ – значение максимального MFCC-коэффициента кепстра; $c_{min}(f)$ – значение минимального MFCC-коэффициента кепстра; D – размерность кепстра.

Межфонемным границам в данном случае соответствуют области изменения гладкости кепстра, то есть точки в которых кривая CSM меняет знак.

Сегментация на основе коэффициента формы кривой

В ходе работы авторами была проведена серия испытаний методов сегментации, описанных выше. Кроме того, предлагается использование другого подхода сегментации слитной речи – основанного на использовании коэффициента формы кривой.

Коэффициент формы кривой определяется как отношение действующего значения функции к среднему значению функции, взятой по абсолютной величине [4]:

$$K_{\phi} = \frac{\sqrt{\frac{1}{T} \int_0^T f^3(t) dt}}{\frac{1}{T} \int_0^T |f(t)| dt}, \quad (2.5)$$

где f – номер текущего фрейма; $c_i(f)$ – i -ый коэффициент MFCC-кепстра речевого сигнала, вычисленного для фрейма f ; I – количество соседних фреймов, слева и справа от текущего, используемых для расчета скорости изменения кепстральных коэффициентов.

Согласно работе [2] параметр STM может быть интерпретирован как модуль скорости изменения спектра (в данном случае – MFCC-кепстра) сигнала. Локальные пики кривой STM соответствуют границам между фонемами.

Недостатком данного метода является то, что он редко определяет границы между гласными и согласными, в случаях выраженной коартикуляции или «плавного» звучания речи [2].

Сегментация на основе меры гладкости кепстра
Проблему предыдущего метода предлагают решить в работе [3]. Для этого вводится мера кепстральной гладкости (Spectral Smoothness Measure, CSM):

где $f(t)$ – функция, T – общее число значений. В качестве функции нами рассматривался амплитудный спектр речевого сигнала.

Следует отметить, что все эти методы требуют проведение определенных стандартных процедур:

Речевой сигнал разбивается на неперекрывающиеся фреймы длительностью 10 мс;

Вычисляется уровень шума [3];

Фреймы с энергией, не превосходящей уровень шума, фильтруются;

В каждом фрейме, в зависимости от выбранного метода, рассчитывается один из параметров: мера спектрального перехода, мера кепстральной гладкости или коэффициент формы кривой речевого сигнала;

Для меры спектрального перехода (STM) остается определить локальные пики кривой, которые соответствуют границам;

Для меры кепстральной гладкости (CSM) и коэффициента формы (K_{ϕ}) кривой вычисляется скорость изменения функции;

Локальные пики модуля скорости изменения функции соответствуют границам между фонемами.

Результаты сравнительного анализа

Для проведения сравнительного анализа нами были рассмотрены несколько фраз, озвученные разными дикторами – женского и мужского полов (рисунки 1 и 2). Правильность рабо-

ты каждого из методов определялась сравнением их результата с результатами ручной сегментации данных фраз. На рисунках 1 - 8 приведены графики ручной и автоматической сегментации с помощью предлагаемого метода и известных методов.

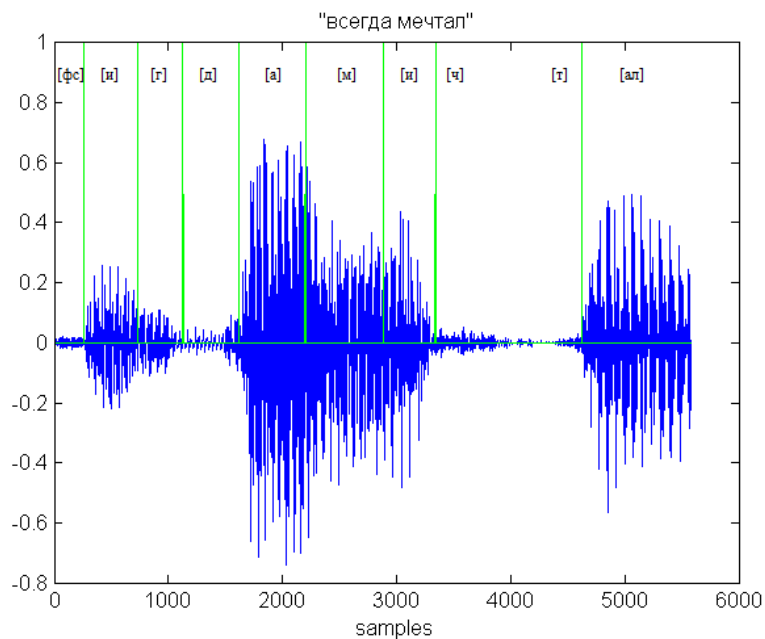


Рисунок 1 – Ручная сегментация фразы «всегда мечтал» в соответствии с фонетической транскрипцией (мужской голос). Пунктирные линии – границы, найденные ручной сегментацией.

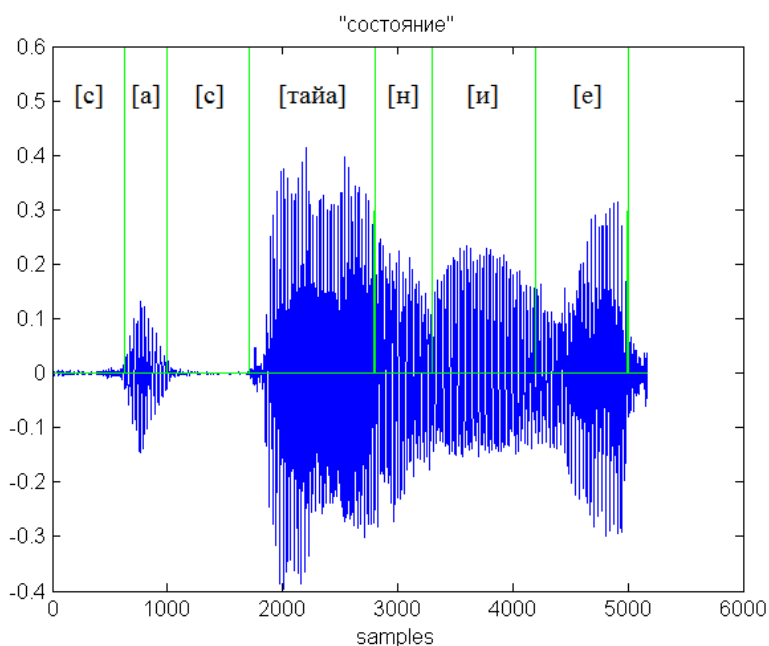


Рисунок 2 – Ручная сегментация слова «состояние» в соответствии с фонетической транскрипцией (женский голос). Пунктирные линии – границы, найденные ручной сегментацией.

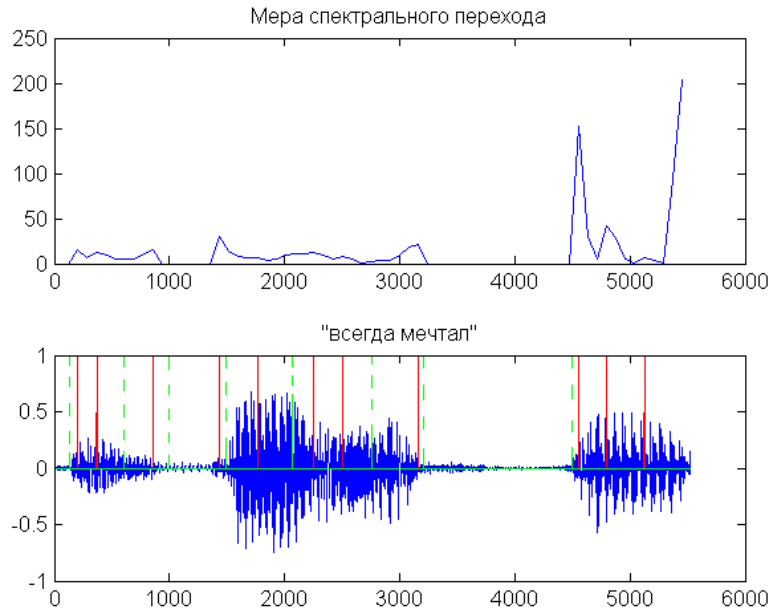


Рисунок 3 – Автоматическая сегментация фразы «всегда мечтал» на основе меры спектрального перехода. Пунктирные линии – границы, найденные ручной сегментацией.

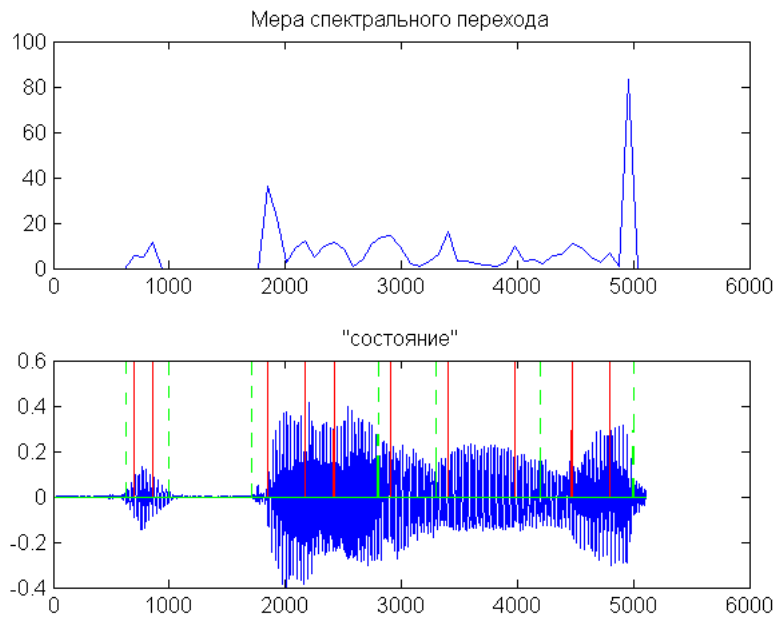


Рисунок 4 – Автоматическая сегментация слова «состояние» на основе меры спектрального перехода. Пунктирные линии – границы, найденные ручной сегментацией.

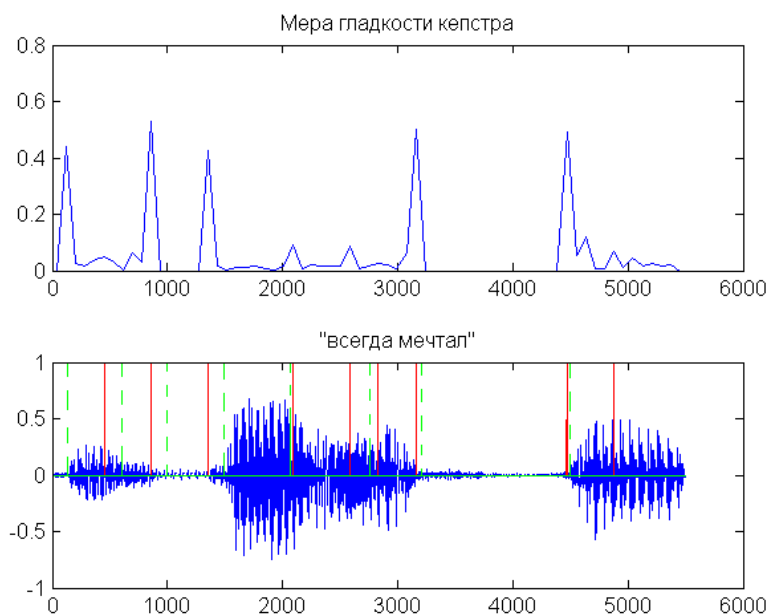


Рисунок 5 – Автоматическая сегментация фразы «всегда мечтал» на основе меры гладкости кепстра. Пунктирные линии – границы, найденные ручной сегментацией.

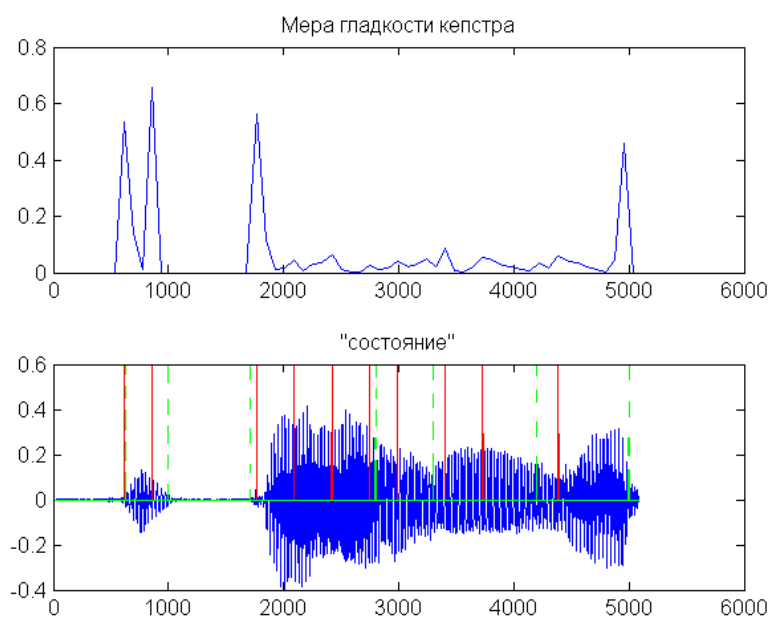


Рисунок 6 – Автоматическая сегментация слова «состояние» на основе меры гладкости кепстра. Пунктирные линии – границы, найденные ручной сегментацией.

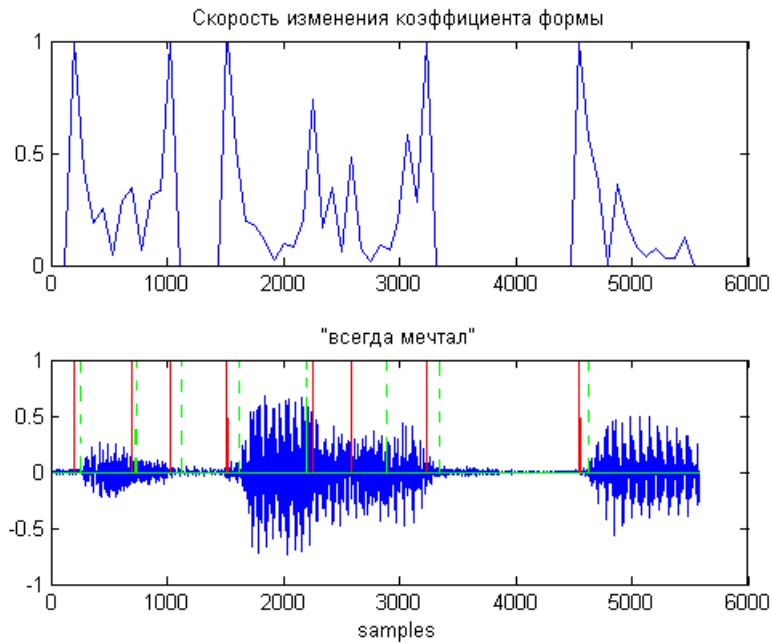


Рисунок 7 – Автоматическая сегментация фразы «всегда мечтал» на основе коэффициента формы кривой. Пунктирные линии – границы, найденные ручной сегментацией.

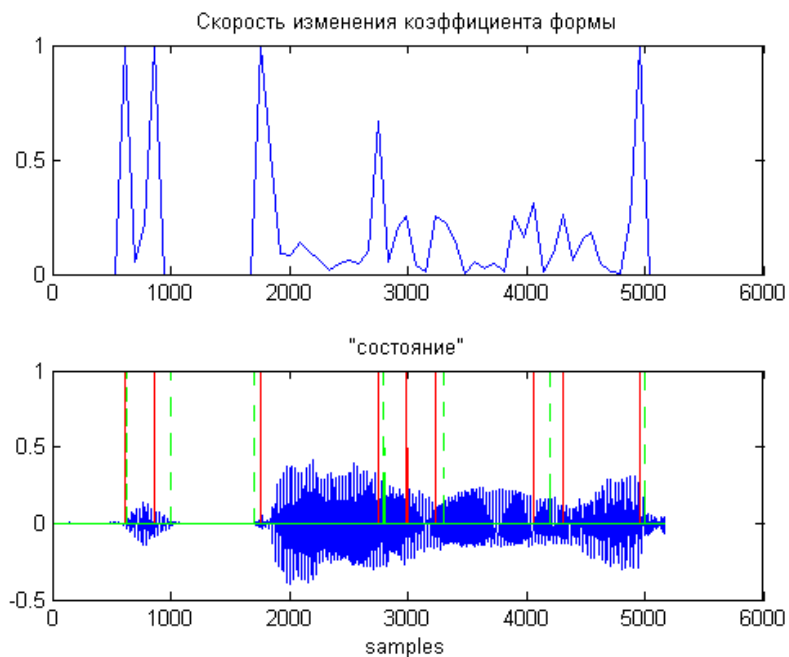


Рисунок 8 – Автоматическая сегментация фразы «всегда мечтал» на основе коэффициента формы кривой. Пунктирные линии – границы, найденные ручной сегментацией.

Заключение и обсуждение результатов

В ходе работы, экспериментально было обнаружено, что методы STM и CSM довольно часто ошибаются, находя ложные границы фонем. Это естественным образом влияет на дальнейшую обработку речи, так как из-за неверно обнаруженных границ увеличивается вероятность ошибочной работы всей системы, в

которой могут быть использованы данные методы сегментации речи. Для таких систем более предпочтителен метод сегментации на основе коэффициента формы кривой, так как он обладает меньшей вероятностью неверного обнаружения границ между фонемами. Предлагаемый метод может быть использован в задачах распознавания, синтеза речи, сжатия речевых сигналов на основе вычисления фонем языка и т.д.

References

1. Rabiner L.R., Shafer R.V. Tsifrovaya obrabotka rechevykh signalov. Perevod s angliyskogo pod red. Nazarova M.V. i Prokhorova YU.N. -M: Radio i svyaz', 1981. – 496 c.
2. Dusan S., Rabiner L.R. On the Relation Between Maximum Spectral Transition Positions and Phone Boundaries. //Proceedings of ICSLP'06 - Pittsburgh, 2006. – S.17–21.
3. Shariy T.V. Ob odnom metode avtomaticheskoy segmentatsii rechevykh signalov. //Bionika intellekta - Khar'kov.– 2009. – №2 (71). - S. 61-65.
4. Atabekov G.I. Osnovy teoriiitsepey: Uchebник. 2-ye izd., ispr. SPb.: Izdatel'stvo «Lan'», 2006. – 432 s.