

МЕТОДЫ И АЛГОРИТМЫ ДЕТЕКТИРОВАНИЯ АКТИВНОСТИ РЕЧИ

Волченков В.А., аспирант кафедры телекоммуникаций и основ радиотехники Рязанского государственного радиотехнического университета, e-mail: volchenkov.rzn@yandex.ru

Витязев В.В., д.т.н., профессор, зав. кафедрой телекоммуникаций и основ радиотехники Рязанского государственного радиотехнического университета

Ключевые слова: детектор активности речи, обнаружение пауз, методы и алгоритмы, сравнительный анализ.

Введение

В голосовой связи речь может быть охарактеризована как прерывистый носитель информации из-за пауз, которые являются характерными особенностями данного вида сигналов по сравнению с другими мультимедийными сигналами, такими как видео, аудио и данные [1]. Интервалы, где голосовая информация присутствует, называют участками активной речи, а паузы между речевыми «всплесками» – участками пауз. Пример, иллюстрирующий участки активной речи и пауз для речевого сигнала, представлен на рис. 1 [1].

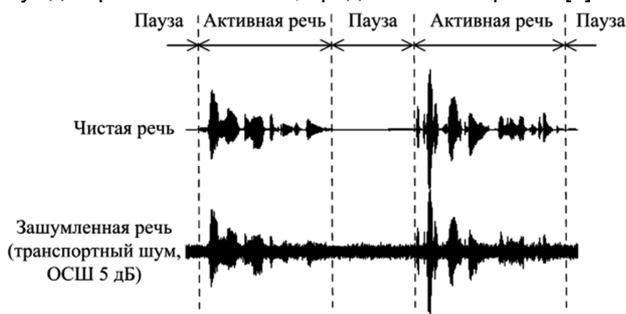


Рис. 1. Участки активной речи и паузы

Детектор активности речи (voice activity detector – VAD) – это алгоритм, предназначенный для различения интервалов активной речи и пауз. Когда система детектирует участки пауз, передача, как правило, прекращается и передается только общее описание информации о фоне. На другом конце декодер восстанавливает интервалы пауз посредством генерации комфортного шума (comfort noise generation – CNG), который дает естественный фоновый шум с плавными переходами от речевых всплесков к паузе и наоборот. Для повышения естественности генерируемого фонового сигнала анализатор фонового шума (comfort noise insertion module – CNI) регулярно передает информацию о среднем уровне фонового шума. Общая структурная схема сжатия интервалов пауз, использующая VAD, CNG и CNI, представлена на рис. 2 [3].

Системы передачи речи, которые используют VAD для сжатия участков пауз, обладают различными преимуществами, особенно полезными для каналов передачи с ограниченной пропускной способностью. Из основных преимуществ детектора активности речи можно перечислить следующие:

Проводится сравнительный анализ эффективности методов и алгоритмов детектирования участков активности речи и пауз между ними. Приведена общая информация о стандартизированных методах детектирования активности речи и оценка их производительности. Представлен новый способ обнаружения пауз в речи.

– сокращение помех от соседних каналов в сотовой связи. Уменьшая мощность передачи во время неактивных интервалов (речевые паузы), возможно подавить помехи от соседних каналов в сотовых системах беспроводной связи.

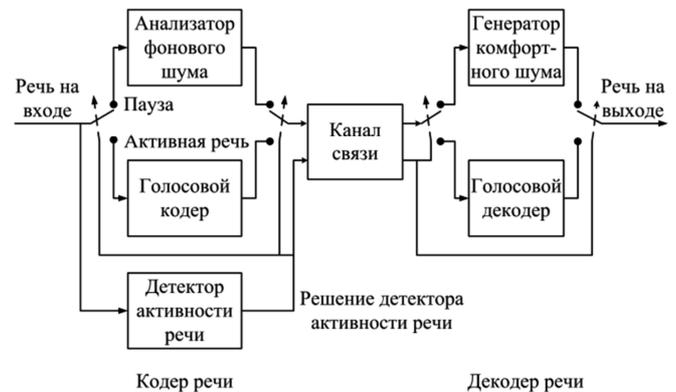


Рис. 2. Общая структурная схема кодера речи со сжатием речевых пауз

– увеличение пропускной способности канала связи в системе множественного доступа с кодовым разделением (code division multiple access – CDMA). Теоретическая пропускная способность системы CDMA, как правило, определяется возможными комбинациями расширяющего кода. Тем не менее, из-за помех от других пользователей, емкость CDMA ограничена значением меньшим, чем этот теоретический предел, то есть из-за помех от других пользователей коэффициент ошибок приема у некоторых пользователей может быть слишком высоким, чтобы иметь возможность осуществить точное декодирование. При снижении мощности передачи во время пауз интерференционные помехи в эфире могут быть уменьшены, что автоматически позволит находиться большему числу пользователей в системе, т.е. происходит увеличение емкости системы CDMA.

– энергосбережение для мобильных терминалов. Мобильные терминалы не должны передавать радиосигналы во время речевых пауз. Таким образом, срок службы батареи терминалов может быть увеличен за счет сохранения энергии во время пауз.

– увеличение пропускной способности канала путем статистического мультиплексирования. Канал может

быть предоставлен только на время передачи активной речи и должен быть освобожден во время пауз. Полученный пользователем канал занимается до конца участка активной речи и освобождается сразу после последнего интервала активной речи. Чтобы вновь получить доступ к каналу, устройство пользователя посылает запрос в начале следующего речевого всплеска. Таким образом, каналный ресурс может быть использован более эффективно посредством применения схемы статистического мультиплексирования, которая позволяет нескольким пользователям общаться в одно и то же время, с учетом ограниченной пропускной способности канала связи. Примечание: в статистическом мультиплексировании существует вероятность того, что когда пользователь делает запрос, свободных слотов в канале не окажется. В этом случае новому пользователю может быть отказано после тайм-аута, что может привести к потере информации и, как следствие, ухудшению качества связи.

– снижение потери пакетов при передаче голоса по сетям пакетной передачи данных. Если количество пакетов больше, чем может быть обработано, то система пакетной передачи данных будет перегружена. Перегрузку системы пакетной передачи данных голосовой связи можно уменьшить, передавая пакеты только с активной речью и вырезая пакеты с паузами.

– снижение скорости передачи. В дополнение к сокращению скорости передачи, достигаемому за счет использования методов сжатия речи, использование VAD вместе со сжатием пауз (удаление интервалов речевых пауз) дает дополнительное снижение скорости передачи независимо от применяемых кодеров речи.

Детектор активности речи обычно производит двоичное решение для обрабатываемого речевого фрейма (как правило, длиной 10-20 мс), указывая на наличие или отсутствие речи, что довольно легко при отсутствии фонового шума [1, 2]. Например, проверяя энергетический уровень входного сигнала, можно получить высокую производительность обнаружения речи / неречи (паузы). Однако, в реальных условиях входной сигнал может быть смешан с шумом, характеристики которого могут быть неизвестны и изменяться со временем. В некоторых случаях, когда уровень фонового шума очень высок, речь может быть заглушена этим шумом. Особенно глухие звуки, которые являются важными для разборчивости речи, могут быть неправильно определены в такой шумной обстановке. Рис. 1 показывает пример для зашумленного сегмента речи с шумом от транспорта, отношение сигнал-шум (ОСШ) – 5 дБ. Как видно из рисунка, некоторые части речи с малой энергией полностью погружены в шум, что делает эти речевые всплески трудно различимыми даже визуально. Неправильная классификация этих речевых всплесков может привести к обрезанию звуков и, как следствие, к значительному ухудшению качества речи. С другой стороны, увеличение ложного обнаружения «не речи» лишает потенциальной выгоды от сжатия пауз. Существует компромисс в производительности VAD – увеличение показателя правильного обнаружения активной речи при минимизации показателя ложных обнаружений интервалов пауз.

Стандартизированные методы детектирования активности речи

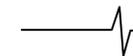
Чтобы использовать преимущества сжатия речевых пауз, было предложено множество алгоритмов VAD, некоторые из них были отобраны организациями по стандартизации, включая: Международный консультационный комитет по телефонии и телеграфии (International Telecommunication Union - Telecommunication sector – ITU-T), Европейский институт по стандартизации в области телекоммуникаций (European Telecommunications Standards Institute – ETSI), Ассоциация телекоммуникационной промышленности США (Telecommunications Industry Association – TIA) и Альянс отраслей электронной промышленности (Electronics Industries Alliance – EIA).

Международный комитет ITU-T выпустил кодеры G.729 Annex B (G.729B) [3] и G.723.1 Annex A (G.723.1A) [4] в качестве расширения для 8 кбит/с G.729 [5] и 5.3/6.3 кбит/с G.723.1 [6] кодеров речи для возможности выполнения прерывистой передачи (Discontinuous transmission – DTX). Европейский институт стандартизации ETSI рекомендовал GSM-FR, -HR и -EFR методы детектирования активности речи для европейских систем цифровой сотовой связи [7-9]. Затем ETSI представил еще два детектора активности речи: адаптивный многоскоростной VAD вариант 1 (кодер AMR1) и вариант 2 (кодер AMR2) [10], с тем, чтобы использовать их в сетях третьего поколения мобильной связи UMTS. Североамериканские организации по стандартизации TIA и EIA представили два алгоритма VAD: один для кодера IS-96 [11], а другой для кодеров IS-127 [12] и IS-733 [13] (алгоритмы VAD для IS-127 и IS-733 имеют одинаковую структуру). В таблице 1 представлены стандартизированные алгоритмы VAD, классифицированные по принципу того, что они анализируют. Главным образом – это энергии поддиапазонов и спектральная форма сигнала [1]. Например, алгоритмы VAD, предложенные TIA и EIA, используют предварительное разбиение сигнала на небольшое количество поддиапазонов, в то время как алгоритм VAD для кодера IS-96 анализирует общую энергию сигнала. С другой стороны, алгоритм VAD кодеров IS-127 и IS-733 также раскладывает сигнал, но только на два поддиапазона.

Таблица 1. Классификация стандартизированных методов VAD в зависимости от способа анализа сигнала (в скобках указано количество спектральных поддиапазонов)

Способ анализа	VAD
Анализ спектральной формы	GSM-FR, GSM-HR, GSM-EFR
Анализ энергии поддиапазонов	IS-96 (1), IS-127 (2), IS-733 (2), AMR1 (9), AMR2 (16)
Другое	G.729B, G.723.1A

Традиционно методы VAD Европейского института стандартизации ETSI были основаны на более точном анализе – анализе спектральной формы входного сигнала. Причиной этого является то, что энергия ошибки кодирования с предсказанием увеличивается, когда спектральные формы фона и входного сигнала не совпадают



(например, в случае активной речи). Тем не менее, в последнем стандарте AMR приняты два вида алгоритмов VAD, каждый из которых основан на анализе спектральной энергии поддиапазонов, а не на более точном анализе формы спектра. Стандартизированные методы VAD в кодерах G.729B и G.723.1A ведут обнаружение с помощью четырех различных способов, включая как анализ спектральной формы, так и анализ энергии поддиапазонов.

Алгоритмы VAD кодеров G.729B/G.723.1A

В качестве расширения к кодеру речи G.729 международный комитет ITU-T SG16 выпустил кодер G.729 Annex B с целью поддержки прерывистой передачи DTX посредством детектирования активности речи VAD, анализа фонового шума CN1 и генерации комфортного шума CNG. Кодер G.729B делит речь на интервалы по 10 мс и вырабатывает решение о наличии или отсутствии речи для каждого фрейма, оценивая при этом четыре параметра [3, 5]:

– разность энергий всего диапазона – $\Delta E_f = \overline{E_f} - E_f$,

– разность энергий диапазона НЧ – $\Delta E_l = \overline{E_l} - E_l$,

– искажение спектра – $\Delta LSF = \sum_{i=0}^9 (\overline{LSF_i} - LSF_i)^2$,

– разность частоты переходов через ноль – $\Delta ZC = \overline{ZC} - ZC$,

где E_f – энергия всего диапазона, E_l – энергия диапазона НЧ, LSF_i – i -я частота спектра сигнала и ZC – частота переходов через ноль входного сигнала, $\overline{E_f}$, $\overline{E_l}$, $\overline{LSF_i}$, \overline{ZC} – параметры, характеризующие шум и обновляемые посредством анализа фонового шума.

Блок схема алгоритма VAD кодера G.729B представлена на рис. 3 [5]. Входные параметры для анализа VAD могут быть получены из входного сигнала или из промежуточных значений речевого кодера. Затем рассчитываются параметры разницы между параметрами входного сигнала и шума ΔE_f , ΔE_l , ΔLSF и ΔZC . Решение о наличии речи получают путем анализа интервалов речевого сигнала по четырем параметрам, которые поступают на схему анализа предыдущих решений. Блок обновления параметров шума основан на схеме авторегрессии первого порядка. Они обновляются, если разница энергии всего диапазона меньше заданного фиксированного порога. Алгоритм VAD кодера G.723.1A имеет аналогичную структуру. Отличие заключается в том, что в кодер G.723.1A делит речь на интервалы по 30 мс.

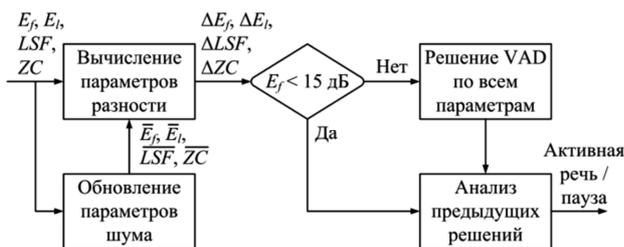


Рис. 3. Алгоритм VAD кодера G.729B

Алгоритмы VAD кодеров GSM-FR/HR/EFR

Алгоритмы VAD кодеров GSM-FR/HR/EFR, предложенные ETSI, имеют схожую структуру. Здесь предсказанная остаточная энергия сравнивается с адаптивным порогом. Предсказанная остаточная энергия вычисляется с использованием значений действительной и сглаженной автокорреляции, которые описывают спектральные характеристики сигнала. Предполагается, что если сигнал является только фоновым шумом, который обычно считается стационарным, то средняя спектральная форма будет подобна форме спектра текущего фрейма, что приведет к меньшей остаточной энергии сигнала. Порог решения алгоритма VAD обновляется только в течение интервалов «неречи», используя самые последние сигналы шума, чтобы отражать актуальные шумовые характеристики. Блок схема алгоритмов VAD кодеров GSM-FR/HR/EFR представлена на рис. 4 [9].



Рис. 4. Алгоритм VAD кодера GSM-EFR

Алгоритмы VAD кодеров AMR

Кодер AMR1 раскладывает входной сигнал на девять неравномерных поддиапазонов с помощью банка фильтров, где нижние полосы частот имеют меньшую полосу пропускания, а более высокие частотные диапазоны – большую пропускную способность. Затем он вычисляет энергию в каждом поддиапазоне, сопровождаемую ее соответствующей оценкой ОСШ. Энергия фонового шума, используемая в расчетах ОСШ, вычисляется адаптивным методом на основе авторегрессионной модели первого порядка и внутренней логической схемы VAD. Наконец, решение VAD принимается путем сравнения суммы ОСШ поддиапазонов с адаптивным порогом и поступает на схему анализа предыдущих решений. Блок схема алгоритма VAD кодера AMR1 представлена на рис. 5 [10].



Рис. 5. Алгоритм VAD кодера AMR1

Структура кодера AMR2 похожа на AMR1 в том, что детектирование речевой активности осуществляется с использованием информации об энергии в поддиапазонах вместе с энергией фонового шума. Однако алгоритм VAD кодера AMR2 трансформирует входной сигнал в частотную область, используя БПФ вместо банка фильтров, применяемого в AMR1, и затем вычисляет энергию в каждом поддиапазоне. Число полос равняется 16, ширина полос также является неравномерной, как и в AMR1. Впоследствии, по спектрам входного сигнала и фонового шума, вычисляется ОСШ для каждого поддиапазона. Энергия фонового шума для каждой полосы изменяется во время интервалов «неречи» с помощью авторегрессионной схемы первого порядка. Чтобы предотвратить возможность возникновения чрезмерной чувствительности к условиям нестационарного фонового шума, AMR2 увеличивает порог принятия окончательного решения алгоритма VAD для высокофлуктуирующих сигналов, оцениваемых по дисперсии их мгновенных межкадровых ОСШ. Кроме того, адаптация уровня шума путем измерения девиации спектра может проводиться неточно, если уровень энергий поддиапазонов меняется быстро. Таким образом, кодер AMR2 изменяет порог VAD адаптивным способом, ориентируясь также на уровень всплеска и принятые решения по прошлым интервалам. Контроль за принятыми решениями осуществляется путем измерения отношения пикового значения сигнал/шум к среднему, где среднее значение ОСШ рассчитывается при помощи авторегрессионной адаптации с увеличенным мгновенным ОСШ. Другими словами, для увеличения отношения пикового значения сигнал/шум к среднему, уменьшаются расчеты по прошлым значениям и уровням всплеска, наряду с увеличением порога алгоритма VAD. Блок-схема алгоритма VAD кодера AMR2 представлена на рис. 6 [10].

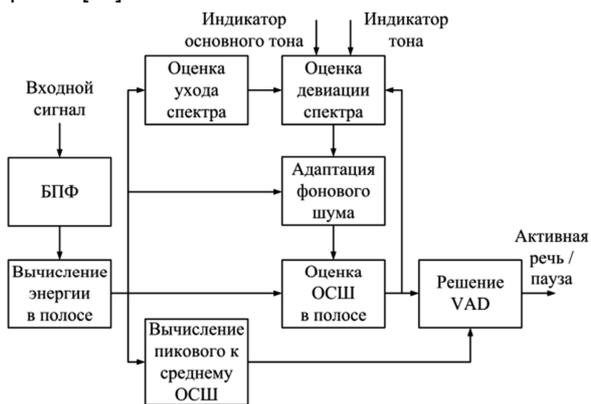


Рис. 6. Алгоритм VAD кодера AMR2

Алгоритмы VAD кодеров IS-127/133

Основанные на CDMA цифровые системы сотовой связи имеют естественную структуру для встраивания алгоритма VAD, называемого алгоритмом определения скорости (rate determination algorithm – RDA), который дает существенное увеличение канальной емкости путем контроля мощности радиопередачи с целью ослабления межканальной интерференции. Ассоциации TIA/EIA представили два вида алгоритмов RDA для стандартов кодеров IS-96 и IS-127, называемые кодером с линейным предсказанием с мультикодовым управлением Qualcomm (Qualcomm code-excited linear prediction – QCELP) 8 кбит/с и кодером с расширенной изменяющейся скоростью (enhanced variable rate codec – EVRC), соответственно. В Северной Америке стандарт CDMA RDA IS-127 поддерживает три скорости: 1, 1/2 и 1/8. Интервалы, в которых содержится речь, кодируются со скоростью 1 или 1/2, а фоновый шум со скоростью 1/8. Алгоритм RDA кодера IS-733 называют 13 кбит/с QCELP, он такой же, как и алгоритм RDA кодера IS-127.

В качестве входных параметров алгоритм RDA кодера IS-127 использует энергии двух поддиапазонов с увеличением долгосрочного предсказания. Сначала, с помощью авторегрессионной модели первого порядка, вычисляется сглаженная энергия в поддиапазонах. Затем производится адаптация энергий сигнала и шума в каждом поддиапазоне в зависимости от усиления долгосрочного предсказания. Другими словами, энергия сигнала активно адаптируется к текущему значению на входе, если коэффициент усиления предсказания относительно высок. С другой стороны, если коэффициент усиления является небольшим, это позволяет увеличить скорость адаптации шума. ОСШ в каждом из двух поддиапазонов вычисляется с использованием информации об энергиях сигнала и шума в каждой подполосе. Окончательная скорость определяется путем сравнения отношений сигнал/шум со значениями адаптивных порогов, зависящих от уровня фонового шума и ОСШ предыдущего фрейма, поступая затем на схему анализа предыдущих решений. Блок-схема RDA кодера IS-127 представлена на рис. 7 [12].

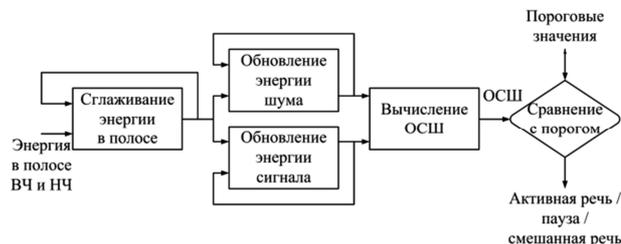
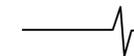


Рис. 7. Алгоритм RDA кодера IS-127

Сравнение производительности алгоритмов VAD, используемых в различных стандартах

В [1] была проведена сравнительная оценка пяти стандартизированных алгоритмов VAD на предмет количества ошибок обнаружения для интервалов активной речи и паузы. Тестовым сигналом была речь длительностью 96 секунд, преобразованная с помощью модифицированной системы промежуточного отклика, а затем смешанная с транспортным шумом с ОСШ: 5, 10, 15, 20 и 25 дБ. Интервалы активной речи и паузы были отмечены вручную. Пропорции между неактивными и активными участками речи были 0,43 и 0,57, соответственно. Решение алгоритмом VAD для кодеров G.729B и AMR2 принимается каждые 10 мс, для кодеров GSM-EFR, AMR1 и IS-127 – каждые 20 мс. При незначительной модификации исходного кода AMR2, результаты можно получать каждые 10 мс, т.к. в своей основе AMR2 принимает решения каждые 10 мс и затем возвращает решение по интервалу длительностью 20 мс, анализируя логическую комбинацию решений по двум интервалам длительностью 10 мс. Что касается многоскоростного кодера IS-127, две верх-



них скорости (1 и 1/2) применяются для кодирования активной речи, а нижняя скорость (1/8) используется для кодирования «неречи».

Производительность работы алгоритмов в среде транспортного шума показана на рис. 8 и 9 [1]. Осциллограммы тестового сигнала и результатов работы детекторов для ОСШ 15 дБ представлены на рис. 10 [1].

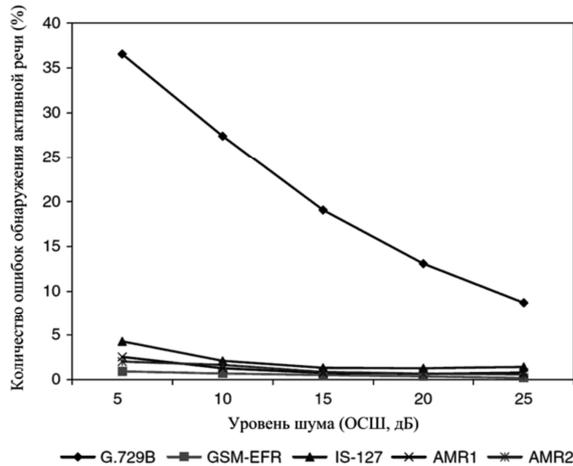


Рис. 8. Количество ошибок обнаружения активной речи по отношению к уровню транспортного шума

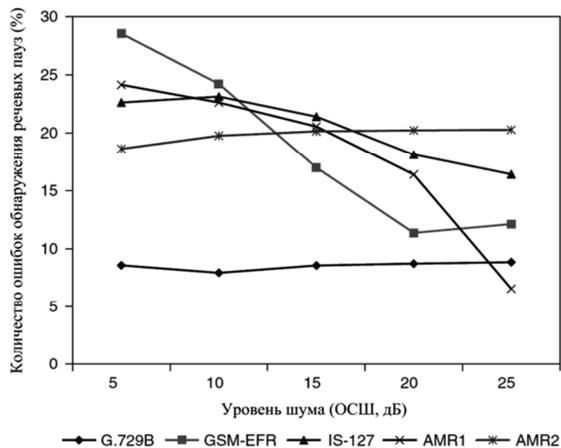


Рис. 9. Количество ошибок обнаружения речевых пауз по отношению к уровню транспортного шума

Кодер G.729B показывает хороший результат по сравнению с другими методами при обнаружении речевых пауз. Однако он демонстрирует очень высокий уровень ошибок детектирования речи, что может повлечь резкое обрезание сигнала речи. Кодер IS-127 показывает относительно высокое количество ошибок при детектировании активной речи в сравнении с алгоритмами VAD, предложенными ETSI. Методы VAD Европейского института стандартизации ETSI, т.е. алгоритмы VAD кодеров GSM-EFR, AMR1 и AMR2, показывают довольно схожие результаты в обнаружении интервалов активной речи, в то время как результаты детектирования интервалов пауз сильно разнятся. Кодер GSM-EFR демонстрирует наилучшие результаты для сравнительно высокого уровня ОСШ (больше 15 дБ). Тем не менее, количество ошибок обнаружения интервалов пауз при уменьшении уровня ОСШ существенно растет. Кодер AMR2 показывает относительно последовательные результаты, не смотря на изменение уровня шума, при

обнаружении участков пауз для речевого сигнала с транспортным шумом. По производительности AMR1 находится между GSM-EFR и AMR2.

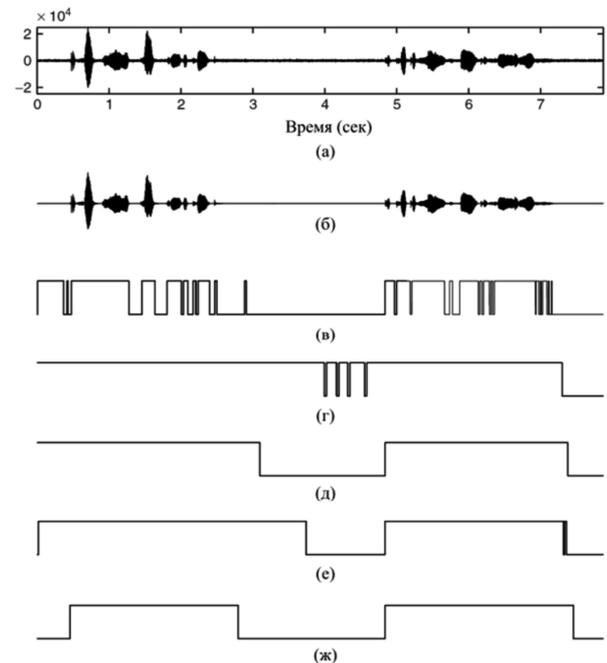


Рис. 10. Сравнение результатов работы алгоритмов VAD для речевого сигнала с транспортным шумом (ОСШ 15 дБ): а) входной зашумленный речевой сигнал; б) чистая речь; в) G.729B; г) IS-127; д) GSM-EFR; е) AMR1; ж) AMR2

Как видно из приведенных графиков, ни один метод не демонстрирует высокой достоверности и точности обнаружения активной речи и речевых пауз одновременно. Таким образом, встает проблема разделения речевого сигнала на интервалы активной речи и речевых пауз с высокой степенью достоверности и наименьшими потерями.

Новый способ обнаружения пауз в речевых сигналах

В настоящей работе предложен детектор активности речи, обеспечивающий существенное повышение точности правильного разделения речевых сигналов на периоды активной речи и паузы [14].

Структурная схема детектора активности речи изображена на рис. 11. Речевой сигнал с выхода электроакустического преобразователя усиливается селективным усилителем и подается на вход сумматора. На второй вход сумматора подается сигнал с выхода генератора измерительного сигнала. Суммарный сигнал с выхода сумматора поступает на вход усилителя-ограничителя 1, где происходит усиление сигнала, а затем ограничение по амплитуде. Аналогичная операция проводится над сигналом, поступающим с выхода селективного усилителя на вход усилителя-ограничителя 2. Сигнал с выхода усилителя-ограничителя 1 подается на первый вход перемножителя. На второй вход перемножителя подается сигнал с выхода усилителя-ограничителя 2. Сигнал с выхода перемножителя поступает на вход накопителя-усреднителя (интервал накопления 10 мс), где происходит вычисление сигнала, по амплитуде которого принимают решение о наличии периода активного речевого сигнала или паузы в пороговом устройстве.

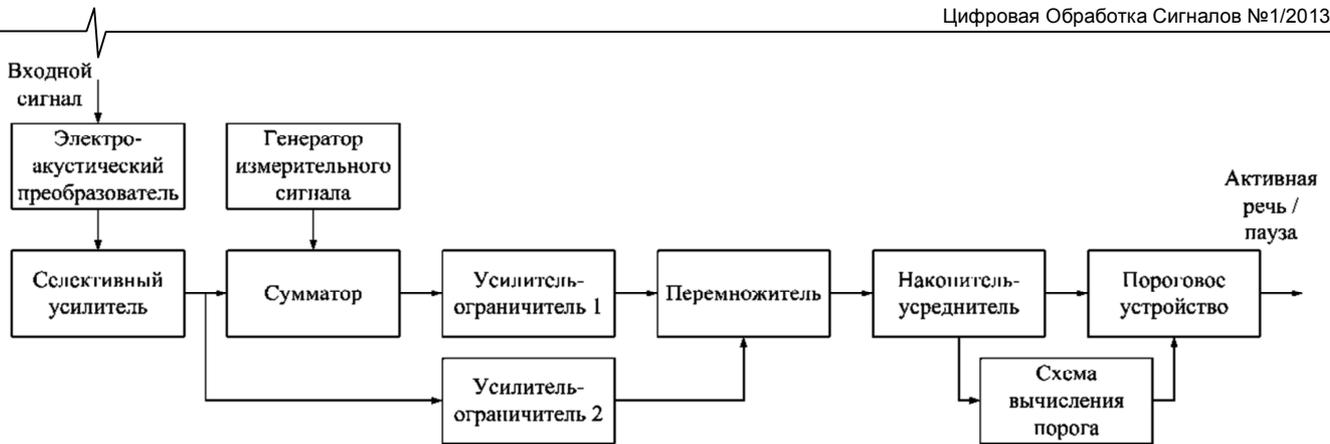


Рис. 11. Структурная схема детектора активности речи

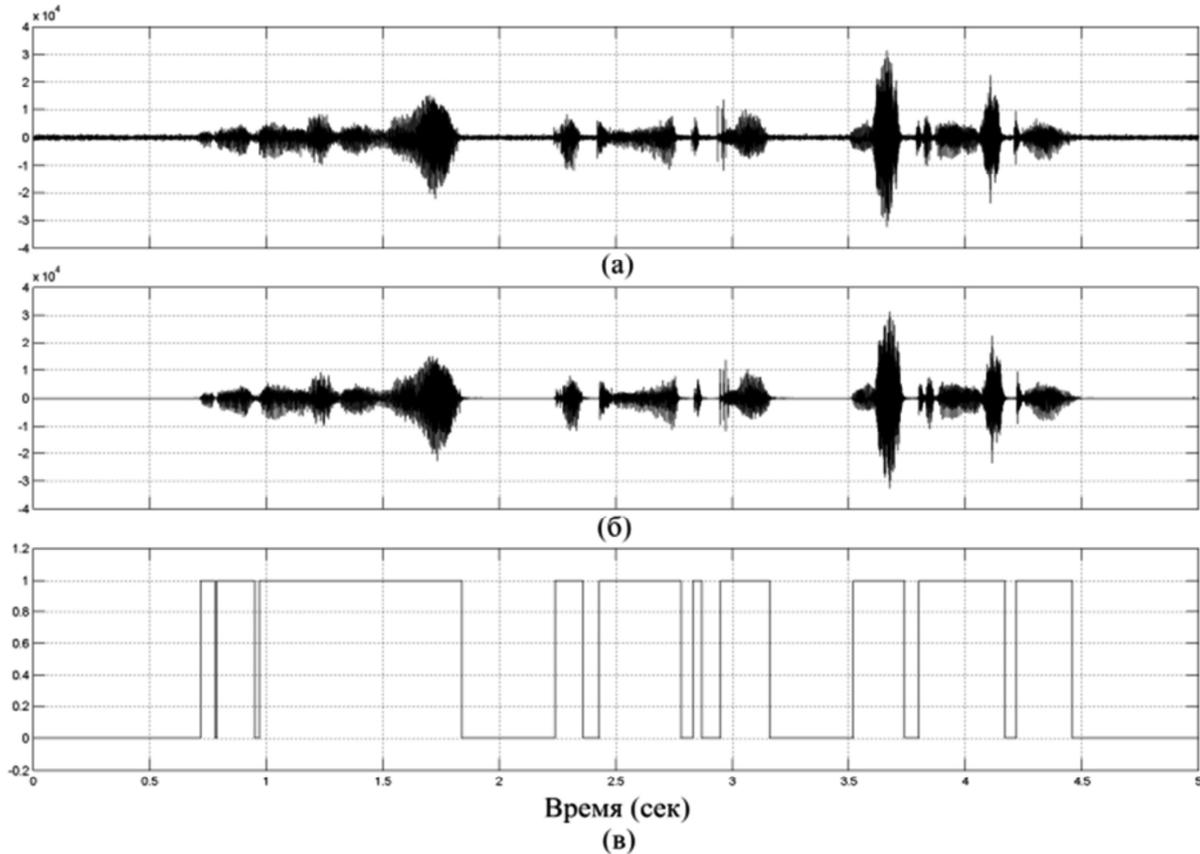


Рис. 12. Осциллограмма тестовой фразы с добавлением транспортного шума (ОСШ 15 дБ), чистого речевого сигнала и сигнала с выхода обнаружителя

Значение порога вычисляется в схеме вычисления порога путем анализа первых 150 мс от начала анализа, т.к. на этом интервале речь обычно отсутствует.

Для проведения исследования была выбрана тестовая фраза: «Продолжение отладки устройства». На рис. 12 представлена осциллограмма данной фразы с добавлением транспортного шума (ОСШ 15 дБ), чистого речевого сигнала и сигнала с выхода обнаружителя.

Общее время записи речевого сигнала указанной фразы составило 5 с, а суммарное время активных периодов – 2,64 с, что составляет 52,8 % времени выбранного речевого сигнала.

Количество ошибок обнаружения активной речи составило 3,56 %, количество ошибок обнаружения речевых пауз – 1,61 %.

Для проверки восприятия речи на слух после обра-

ботки была произведена запись речевого сигнала с помощью блока аудиозаписи, управляемого командами с обнаружителя пауз. Интервалы, в которых присутствовала речь, оставались без изменения. В интервалы, в которых находились паузы, записывался нуль. Качество полученного речевого сигнала практически не отличалось от исходного. Слова были хорошо различимы, речь легко воспринималась на слух.

Заключение

Таким образом, можно сделать вывод о том, что разработан новый способ обнаружения пауз в речи, который позволяет существенно повысить точность разделения на активные участки речи и паузы. При этом качество восстанавливаемой речи остается практически на том же уровне.



Предметом последующего исследования является сравнительная оценка количества ошибок и вероятности правильного обнаружения участков активной речи и паузы при различных уровнях ОСШ.

Литература

1. Kondo A.M. Digital Speech. Coding for Low Bit Rate Communication Systems. – John Wiley & Sons, Ltd. 2004. – 442 p.
2. Шелухин О.И. Цифровая обработка и передача речи / О.И. Шелухин, В.Г. Лукьянцев; Под ред. О.И. Шелухина. – М.: Радио и связь, 2000. – 456 с.: ил.
3. ITU-T (1996) A silence compression scheme for G.729 optimised for terminals conforming to ITU-T V.70, ITU-T Rec. G.729 Annex B.
4. ITU-T (1996) Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s. Annex A: Silence compression scheme, ITU-T Rec. G.723.1 Annex A.
5. ITU-T (1996) Coding of speech at 8 kbit/s using conjugate-structure algebraic code excited linear prediction (CS-ACELP), ITU-T Rec. G.729.
6. ITU-T (1996) Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s, ITU-T Rec. G.723.1.
7. ETSI (1998) Digital cellular telecommunications system (phase 2+); Voice activity detector (VAD) for full rate speech traffic channels, GSM 06.32 (ETSI EN 300 965 v7.0.1).
8. ETSI (1999) Digital cellular telecommunications system (phase 2+); Voice activity detector (VAD) for full rate speech traffic channels, GSM 06.42 (draft ETSI EN 300 973 v8.0.0).
9. ETSI (1997) Digital cellular telecommunications system; Voice activity detector (VAD) for enhanced full rate (EFR) speech traffic channels, GSM 06.82 (ETS 300 730), March.
10. ETSI (1998) Digital cellular telecommunications system (phase 2+); Voice activity detector (VAD) for adaptive multi-rate (AMR) speech traffic channels, GSM 06.94 v7.1.1 (ETSI EN 301 708).
11. DeJaco P., Gardner W., and C. Lee (1993) 'QCELP: The North American CDMA digital cellular variable rate speech coding standard', in IEEE Workshop on Speech Coding for Telecom, pp. 5–6.
12. TIA/EIA (1997) Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems, IS-127.
13. TIA/EIA (1998) High rate speech service option 17 for wideband spread spectrum communication systems, IS-733.
14. Пат. 2436173 Российская Федерация, МПК G10L 15/00, G10L 11/02, Способ обнаружения пауз в речевых сигналах и устройство его реализующее / Витязев В.В., Розов В.И., Волченков В.А.; заявитель и патентообладатель Рязанский государственный радиотехнический университет. – № 2010124342/08, заяв. 15.06.10; опубл. 10.12.11, Бюл. 34.

VOICE ACTIVITY DETECTION METHODS

Volchenkov V.A., Vityazev V.V

Problems of accuracy increase in voice activity detection (VAD) are considered. General information about standard VAD methods and performance comparison of VADs are given. A new method of voice activity detection is offered.

Уважаемые коллеги!

Приглашаем Вас принять участие в формировании тематических выпусков журнала «Цифровая обработка сигналов» и размещению рекламы продукции (услуг) Вашей организации на его страницах. В случае положительного решения просим представить в редакцию журнала Ваши предложения по плановому размещению информационных материалов и макет рекламы продукции (услуг) с указанием желаемого её месторасположения: обложка (2-я, 3-я или 4-я стр.), цветная внутренняя полоса (объем полосы).

Журнал «Цифровая обработка сигналов» издается с 1999 года. Выходит ежеквартально, тиражом – 700 экз. Распространяется по подписке через агентство «Роспечать» в России (индекс 82185), СНГ и странах Балтии (индекс 20630), а также на Конференции: «Цифровая обработка сигналов и ее применение – DSPA'» и Выставках: «ExpoElectronica», «СвязьЭкспокомм», «ЭЛЕКТРОНИКА: компоненты, оборудование, технологии» (г. Москва) и др.

Научно-технический журнал «Цифровая обработка сигналов» включен в Список изданий, рекомендуемый ВАК РФ для публикации результатов научных исследований соискателями ученой степени доктора и кандидата технических наук в области радиотехники, связи, вычислительной техники, электроники, информационно-измерительных и управляющих систем.

Планируемые сроки издания отдельных номеров журнала:

№ 2 – июнь 2013 г. Тематический выпуск: «Теория и методы цифровой обработки сигналов» (по материалам международной научно-технической конференции «Цифровая обработка сигналов и ее применение – DSPA' 2013».

№ 3 – сентябрь 2013 г. Тематический выпуск: «Цифровая обработка изображений».

№ 4 – декабрь 2013 г. Тематический выпуск: «ЦОС в радиотехнике и системах телекоммуникаций».

Ориентировочная стоимость рекламных услуг:

4-я (внешняя) страница цветной обложки – 25 тысяч рублей.

2-я и 3-я (внутренние) страницы цветной обложки – 15 тысяч рублей.

1\2 цветной внутренней полосы – 8 тысяч рублей.

Ждем Ваших предложений

С наилучшими пожеланиями, зам. главного редактора д.т.н., профессор Витязев Владимир Викторович, телефон 8-903-834-81-81

Предложения прошу направлять по адресу: E-mail: vityazev.v.v@rsreu.ru или info@dspa.ru