УДК 004.032.6

ПАРАМЕТРИЧЕСКИЙ АУДИОКОДЕР НА ОСНОВЕ РАЗРЕЖЕННОЙ АППРОКСИМАЦИИ С ПЕРЦЕПТУАЛЬНО-ОПТИМИЗИРОВАННЫМ СЛОВАРЕМ ЧАСТОТНО-ВРЕМЕННЫХ ФУНКЦИЙ

Ал.А. ПЕТРОВСКИЙ, В.Ю. ГЕРАСИМОВИЧ

Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, 220013, Беларусь

Поступила в редакцию 3 сентября 2014

Рассматривается метод частотно-временного преобразования сигналов, показывается преимущества совместного, частотно-временного анализа перед традиционным раздельным анализом во временной либо частотной областях. Описан алгоритм согласованной подгонки для декомпозиции сигнала в базис частотно-временных функций, а также модификация данного алгоритма с использованием принципов психоакустики для выбора наиболее важных для восприятия компонент аудиосигнала. Предлагается схема построения универсального аудиокодера на основе разреженной аппроксимации с перцептуально-оптимизированным словарем вейвлет коэффициентов.

Ключевые слова: частотно-временные преобразования, пакет дискретного вейвлет преобразования, согласованная подгонка, психоакустическая модель, аудиокодер.

Введение

На текущем этапе развития алгоритмов сжатия аудиосигналов разработано большое количество кодеров, использующих различные особенности входного сигнала. В зависимости от выделяемых особенностей кодеры можно разделить на два класса: вокодеры, эффективно сжимающие речевой сигнал и позволяющие достичь низких скоростей передачи данных при сохранении высокого качества выходного сигнала; и аудиокодеры, алгоритмы которых нацелены на работу с таким типом входных сигналов, как музыка, звуки природы и тому подобное. В двух вышеприведенных классах кодеров разработано большое количество разнообразных реализаций, которые имеют свои плюсы и минусы, однако данная ситуация ставит в затруднительное положение при выборе определенного аудиокодера, или вокодера, для решения конкретной задачи. Во-первых, необходимо заранее знать, какой тип входного сигнала будет преобладать (речь или музыка), во-вторых, необходимо изучить все разнообразие представленных алгоритмов для того, чтобы выбрать наиболее эффективный из них. Естественной задачей, вытекающей из изложенной ситуации, является построение универсального аудиокодера, способного максимально эффективно работать со всеми известными типами звукового информационного наполнения.

С точки зрения обработки аудиосигналов кодеры можно разделить на две группы: кодеры на основе преобразования (transform coders), и параметрические кодеры (parametric coders) [1]. Суть кодеров на основе преобразования заключается в том, чтобы привести входные данные к той форме, в которой компоненты сигнала будут иметь минимальную корреляцию между собой, что позволит кодировать их независимо друг от друга. Для учета особенностей восприятия звука человеком и для уменьшения перцептуальной избыточности кодируемых данных в кодерах с преобразованием используется психоакустическая модель слуха человека. Основная идея параметрического кодирования – это параметризация аудиосигнала каким-либо способом для определения значений, которые описывают важные аспекты кодируемого аудиосигнала. Оба описанных выше подхода имеют свои достоинства и

недостатки и используются в таких широко распространенных кодерах, как MPEG-1 Layer 3 (MP3), Advanced Audio Coding (AAC) и подобные им [1]. Для построения универсального аудиокодера логично выбрать параметрический подход с некоторыми особенностями, взятыми из алгоритмов кодирования с преобразованием.

Использование параметрической модели кодирования позволит уменьшить количество информации, необходимое для описания и последующего восстановления сжимаемого сигнала, что в свою очередь даст ощутимое увеличение степени сжатия целевого сигнала (и уменьшения скорости битового потока). Для более эффективной обработки и параметризации входного сигнала можно также использовать психоакустическую модель, которая позволит исключить избыточную с точки зрения восприятия человеческим ухом информацию, тем самым уменьшая необходимое для передачи и восстановления сигнала количество параметров.

Параметризация сигнала может проводиться, как правило, либо во временной области, либо в частотной. Анализ сигнала во временной области позволяет получить некоторую информацию о его природе и характеристиках, но не дает никакой информации о его частотных свойствах. Решением задачи получения частотной информации о сигнале является переход в частотную область с помощью преобразования Фурье, однако оно не позволяет произвести многомасштабный анализ, т.е. точно определить временную локализацию той или иной частоты, что весьма существенно для нестационарных сигналов. Большинство окружающих нас сигналов являются нестационарными, т.е. такими, спектральные характеристики которых быстро изменяются во времени. Для работы с такими сигналами необходимо иметь совмещенную частотно-временную картину характеристик сигнала. Этот факт определяет выбор математического аппарата для параметризации входного сигнала аудиокодера: частотно-временные преобразования сигнала.

Частотно-временные преобразования аудиосигнала

Частотно-временные преобразования подразумевают декомпозицию сигнала, то есть аппроксимацию сигнала частотно-временными функциями, полученными перемещением, модуляцией и масштабированием базисных функций, имеющих определенную временную и частотную локализацию.

Декомпозиция сигнала основана на алгоритме согласованной подгонки (matching pursuit – MP) со словарем частотно-временных функций. В данном подходе, любой сигнал $x(t)$ представляется в виде линейной комбинации частотно-временных функций (называемых атомами) $g_{\gamma_n}(t)$, выбираемых из избыточного словаря *D*. Избыточный словарь в данном контексте обозначает тот факт, что он содержит намного больше элементов, нежели минимальное необходимое количество базисных функций, покрывающих данное пространство. Любой сигнал можно разложить с помощью алгоритма частотно-временной декомпозиции следующим образом [2]:

$x(t) = \sum_{n=0}^{\infty} a_n \cdot g_{\gamma_n}(t)$, где $g_{\gamma_n}(t) = \frac{1}{\sqrt{s_n}} g(\frac{t-p_n}{s_n}) \exp(j(2\pi f_n t + \varphi_n))$, a_n – масштабирующий коэффициент, который показывает вклад атома в формирование выходного сигнала.

В большинстве приложений, где используется алгоритм согласованной подгонки, словарь частотно-временных функций подбирается с учетом специфики сигналов. Масштабирующий коэффициент s_n служит для контроля ширины оконной функции, параметр p_n необходим для определения временного расположения функции. Параметры f_n и φ_n – соответственно, частота и фаза экспоненциальной функции.

Алгоритм согласованной подгонки является «жадным» алгоритмом и подразумевает поиск локально оптимальных решений с расчетом, что глобальное решение также будет оптимальным. Выбор аппроксимирующей функции из словаря заключается в поиске такой функции, которая дает максимальное значение скалярного произведения с фреймом анализируемого сигнала. После выбора функции необходимо вычесть ее вклад в формирование сигнала, что даст остаточный сигнал. Следующая итерация алгоритма производится над

найденным остаточным сигналом. В идеальном случае, остановка алгоритма происходит когда остаточный сигнал равен нулю. Однако, как правило, фиксируется количество исполняемых алгоритмом итераций, либо вводятся определенные энергетические пороги, при достижении которых алгоритм останавливает свою работу.

При использовании алгоритма согласованной подгонки стоит учитывать два основных взаимосвязанных фактора: выбор частотно-временных функций, формирующих словарь, и вычислительные затраты при работе алгоритма. Так как словарь должен быть избыточным, он может содержать большое множество различных базисных функций, однако, с другой стороны, чрезмерно объемный словарь увеличит время поиска подходящей функции. Следовательно, стоит вопрос о выборе оптимального словаря частотно-временных функций. Одним из наилучших вариантов является формирование словаря на основе анализируемого сигнала. Для этого необходимо выполнить разложение входного сигнала по вейвлет функциям определенного семейства. На основе данных функций формируется словарь атомов. Избыточный словарь можно сформировать, применив полное дерево пакетного дискретного вейвлет преобразования (ПДВП), как это показано на рис. 1 [3, 4].

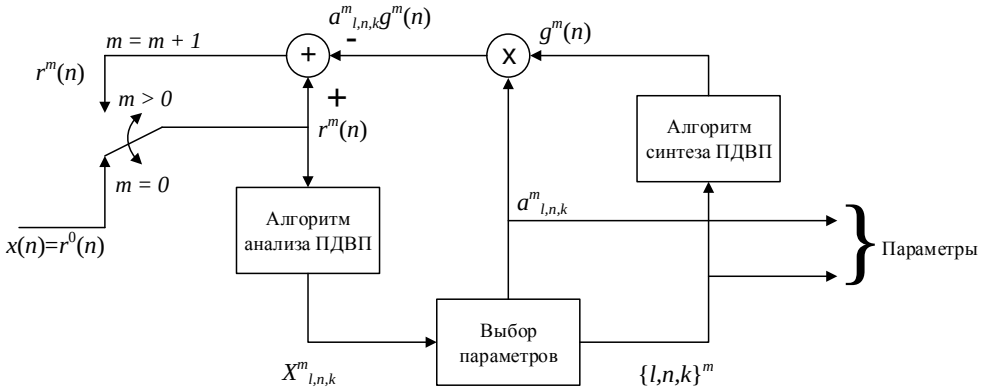

Рис. 1. Параметризация сигнала на основе полного дерева пакетного дискретного вейвлет преобразования

На рис. 1, $x(n)$ – входной фрейм сигнала, $r^m(n)$ – остаточный сигнал на итерации *m*, $X^m_{l,n,k}$ – вейвлет функция. Алгоритм работает по схеме «анализ через синтез» и заключается в итеративном повторении шести основных шагов.

Шаг 1. Декомпозиция сигнала-остатка (для первой итерации – это входной сигнал) полным деревом анализа ПДВП (блок алгоритм анализа ПДВП на рис. 1);

Шаг 2. Выбор наиболее значимого вейвлет-коэффициента X_γ – коэффициента с абсолютно максимальным значением весового коэффициента a_γ;

Шаг 3. Наиболее значимому вейвлет-коэффициенту ставится в соответствие атом из словаря;

Шаг 4. Формирование результирующего вектора $g^m_\gamma(n)$ декомпозиции выполняется при помощи обратного ПДВП преобразования (блок алгоритм синтеза ПДВП на рис. 1) на основе полного дерева;

Шаг 5. Получение результирующего сигнала, как умножение результирующего вектора $g^m_\gamma(n)$ на весовой коэффициент a^m_γ;

Шаг 6. Получение сигнала-остатка $r^{m+1}(n)$ путем вычитания результирующего сигнала из остаточного сигнала $r^m(n)$.

Процесс повторяется, где для последующей итерации входным сигналом является сигнал-остаток, полученный из предыдущей итерации.

Использование частотно-временного преобразования на основе алгоритма согласованной подгонки со словарем частотно временных функций позволяет получить разреженную

аппроксимацию аудиосигналов, то есть, представление сигнала наименьшим количеством ненулевых элементов. Улучшить производительность алгоритма согласованной подгонки может перцептуальная оценка узлов дерева декомпозиции для каждого фрейма индивидуально. Это позволяет получить оптимальное дерево разложения ПДВП на каждом фрейме анализируемого сигнала, что, в свою очередь, позволит добиться минимизации словаря *D*.

Алгоритм согласованной подгонки с применением психоакустической модели слуха человека

Минимизация словаря *D* реализуется в модифицированном алгоритме согласованной подгонки, где в алгоритм *MP* введена процедура расчета перцептуальной энтропии [5, 6].

Полное дерево ПДВП E_{full} содержит в себе множество структур E_j, которые представляют собой деревья с определенным набором узлов. Этот факт можно использовать для формирования структуры дерева декомпозиции, согласованное со шкалой барков (критическое дерево – рис. 2), что позволит производить расчет параметров психоакустической модели в вейвлет области [6].

Психоакустическая модель дает следующие возможности: расчет перцептуальной энтропии каждого узла дерева декомпозиции текущего фрейма для получения оптимального дерева (минимизация словаря *D*), а также расчет порогов маскирования, что позволяет оценивать перцептуальную важность каждого найденного параметра.

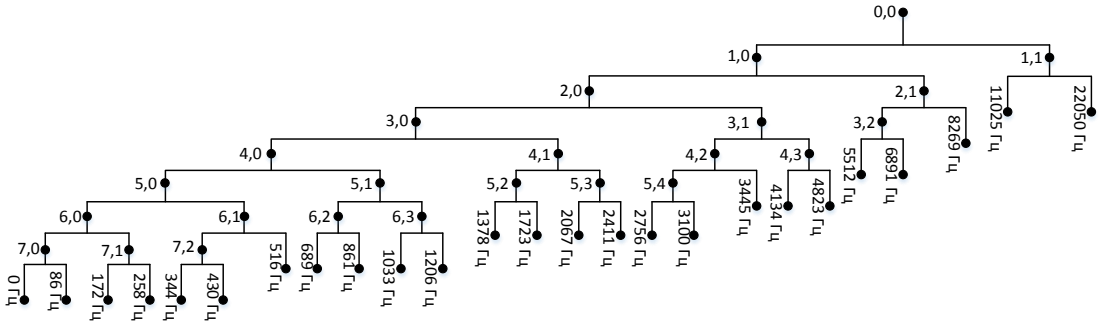

Рис. 2. Критическое дерево ПДВП, согласованное со шкалой барков

На рис. 3 отражено сравнение сходимостей алгоритма согласованной подгонки на основе трех различных словарей [5, 7]. Как видно из рис. 3, алгоритм согласованной подгонки со словарем перцептуально оптимизированных функций ПДВП обеспечивает наименьшее среднеквадратическое отклонение, по сравнению с алгоритмом согласованной подгонки на основе словаря с полным набором вейвлет-функций ПДВП или затухающих синусоид.

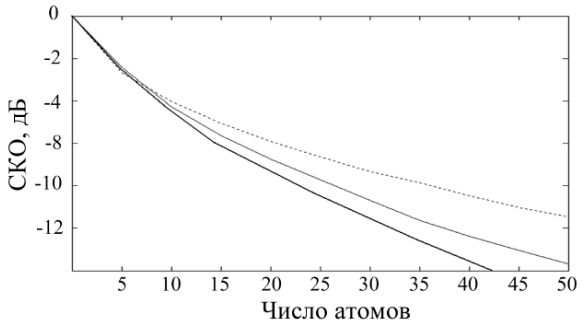

Рис. 3. Среднеквадратичное отклонение: затухающие синусоиды (пунктирная линия), вейвлет пакет с избыточным словарем (тонкая сплошная линия), вейвлет пакет, оптимизированный для каждого фрейма на основе психоакустики (толстая сплошная линия)

Модель параметрического аудиокодера на основе разреженной аппроксимации с перцептуально-оптимизированным словарем частотно-временных функций

На современном этапе развития параметрических аудиокодеров можно выделить два основных подхода. Суть первого заключается в выделении из сигнала некоторых компонент,

сходных по своей природе, и их раздельный анализ. Достоинством такого подхода является то, что появляется возможность точно смоделировать входной сигнал и подобрать для каждой компоненты подходящий математический аппарат. Однако очевидными недостатками таких алгоритмов является привязка инструментов анализа компонент к типу сигнала (речь, музыка, смешанное содержимое). Также негативным моментом является то, что при анализе нескольких компонент индивидуально, появляется большое количество параметров, необходимых для синтеза сигнала декодером. Примером таких подходов могут служить модели «синусоидальная компонента-остаток (шум)», «синусоидальная компонента-транзиенты-остаток (шум)» [7]. Смысл второго подхода заключается в том, что сигнал параметризируется без выделения каких-либо компонент, что уменьшает количество передаваемых декодеру параметров. Ниже дано краткое описание трех аудиокодеров, основанных на алгоритме согласованной подгонки, проведен их анализ и сравнение.

Алгоритм кодирования, описанный в [8], производит декомпозицию сигнала на атомы, взятые из фиксированного, заранее предопределенного избыточного словаря. После получения параметров сигнала, происходит их обработка, а именно, перцептуальная оценка параметров с помощью анализа порогов маскирования каждой полученной частотно-временной функции. На этом этапе принимается финальное решение о важности каждого из параметров. Заключительным этапом работы кодера является квантование параметров и их упаковка для передачи.

Аналогичный вышеизложенному подход представлен в [9]. В данной работе проводится разреженная аппроксимация сигнала s следующим образом: инициализируется остаток $r_0 = s$, строится избыточный словарь на основе атомов Габора. Следующий после инициализации шаг – декомпозиция сигнала с помощью алгоритма *MP*. Выходом согласованной подгонки является набор параметров. Второй этап алгоритма кодирования – перцептуальная обработка найденных атомов. Для этого они упорядочиваются по убыванию амплитуды и к ним применяется модель маскирования. Те из них, которые ниже порога маскирования, удаляются из набора.

В отличие от вышеприведенных работ, в [10] представлен аудиокодер на основе разреженной аппроксимации, но выделяющий из сигнала элементы различной природы, и работающий с ними раздельно. В данном случае под элементами разной природы понимаются гармонические (синусоидальные) составляющие сигнала, транзиенты (кратковременные высокоэнергетические всплески), микротранзиенты (острые всплески низкоэнергетических транзиент), шумовая составляющая. Первые три элемента сигнала параметризируются на основе алгоритма согласованной подгонки с различными словарями для каждой из компонент. Моделирование шумовой составляющей реализовано на основе определения спектральной огибающей сигнала алгоритмом линейного предсказания (*LP*).

Подходы, описанные выше, имеют основной недостаток: словарь атомов является фиксированным, либо предопределенным для каждого сигнала, что не может являться оптимально эффективным в силу того, что тип сжимаемых данных может сильно отличаться в различных сигналах. Для решения этой проблемы можно использовать большее количество элементов в словаре, однако это приведет к резкому возрастанию времени поиска по словарю и увеличению вычислительных затрат. Также этот вариант может не дать эффективного решения в силу нестационарности реальных аудиосигналов и их большого разнообразия.

Анализируя вышесказанное можно выделить ряд требований, предъявляемых к разрабатываемому кодеру: масштабируемость, низкая скорость битового потока, высокое качество восстановленного сигнала. В качестве математической модели описания сигнала решено было выбрать модель разреженной аппроксимации сигнала. Данная модель позволяет описать входной аудиосигнал минимальным количеством параметров. В отличие от рассмотренных выше подходов, словарь атомов будет формироваться из самого сигнала для каждого входного фрейма индивидуально. Это позволит добиться максимальной гибкости и оптимальной эффективности алгоритма. Еще одним шагом оптимизации работы кодера будет являться тот факт, что словарь вейвлет коэффициентов будет формироваться на основе перцептуально-оптимизированного ПДВД вместо использования полного дерева декомпозиции. Это позволит исключить маскируемые компоненты, не воспринимаемые при прослушивании восстановленного аудиосигнала. Общая схема разрабатываемого кодера

представлена на рис. 4. Основу кодера составляют адаптивный ПДВП (область, отмеченная штриховкой) и блок подбора параметров (область, выделенная штрихпунктирной линией). В блоке адаптивного ПДВП для каждого фрейма сигнала выполняется подбор оптимальной структуры дерева. Параметры реконфигурации дерева рассчитываются на основе данных из блоков «Перцептуальная энтропия» и «Энтропия».

Блок подбора параметров отвечает за формирование параметров входного фрейма. Основу составляет блок «Согласованная подгонка», на вход которому поступают рассчитанная перцептуальная энтропия для каждого узла дерева и полученные в блоке «ПДВП» вейвлет коэффициенты. На основе отобранных компонент выполняется синтез сигнала в блоке «Обратный ПДВП», который затем вычитается из сигнала. Остаточный сигнал анализируется в блоке «ПДВП» и весь процесс повторяется.

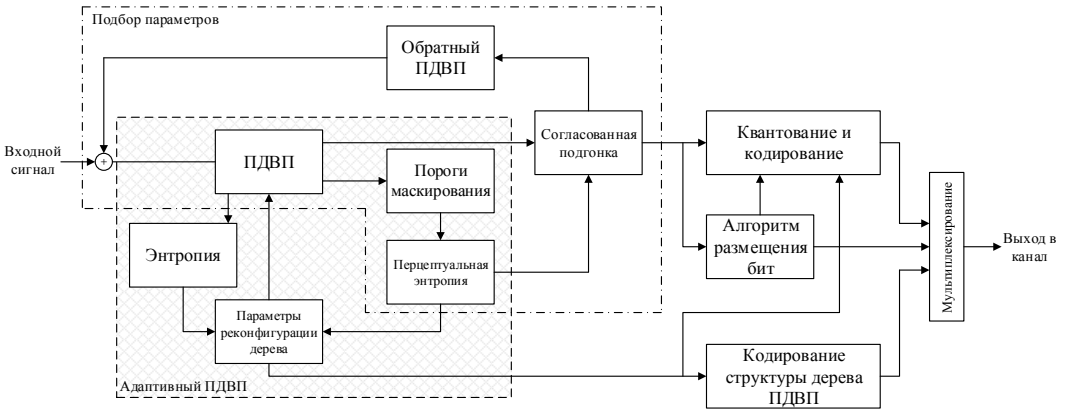

Рис. 4. Общая схема разрабатываемого аудиокодера

Отобранные параметры входного фрейма сигнала на последней стадии работы кодера квантуются, кодируется структура дерева ПДВП. Эта информация затем передается в канал. Работа декодера показана на рис. 5.

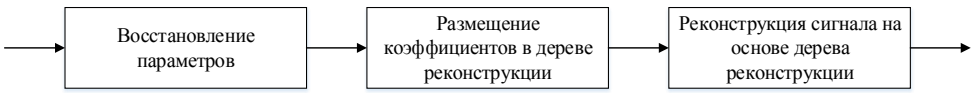

Рис. 5. Общая схема работы декодера

На стороне декодера принятая информация восстанавливается, затем декодированные параметры размещаются на соответствующие позиции в структуре дерева. Реконструкция сигнала выполняется на основе обратного ПДВП в соответствии со структурой полного дерева ПДВП.

Результаты экспериментальных исследований [3] параметрического аудиокодера на основе разреженной аппроксимации с перцептуально-оптимизированным словарем частотно-временных функций сведены в таблицу.

Оценки реконструированного сигнала параметрическим аудиокодером

Тестовый материал	*MP3*			*AAC*			Параметрический кодер		
	Степень сжатия	Субъективное отличие	*MBSD*	Степень сжатия	Субъективное отличие	*MBSD*	Степень сжатия	Субъективное отличие	*MBSD*
Rock	7,7	-0,2	1,45	9,5	-0,067	1,17	18,3	-0,3	1,55
Classic	16,6	-0,067	1,02	9,6	-0,2	1,32	20,2	-0,4	1,67
Pop	8,4	-0,045	1,13	11,5	-0,067	1,16	14,7	-0,2	1,34

Заключение

Приведено краткое описание нескольких схем параметрических аудиокодеров и алгоритмов их работы, показаны особенности и недостатки. На основе этой информации

сформулировано направление исследования и разработки универсального масштабируемого аудиокодера. Дальнейшая работа должна заключаться в исследовании и обосновании выбора оптимального семейства вейвлет функций для декомпозиции входного сигнала на основе ПДВП, разработке эффективной схемы квантования полученных параметров сигнала и их упаковке, увеличению быстродействия и уменьшению вычислительных затрат алгоритма.

Стоит отметить, что выбранный математический аппарат позволяет решать некоторые другие задачи помимо сжатия: классификация аудиосигналов, защита информации (добавление «водяных знаков» в сигнал), конверсия голоса.

PARAMETRIC AUDIO CODER BASED ON SPARSE APPROXIMATION WITH FRAME-BASED PSYCHOACOUSTIC OPTIMIZED WAVELET PACKET DICTIONARY

Al.A. PETROVSKY, V.Y. HERASIMOVICH

Abstract

Time-frequency signal transform is considered. Advantages of joint time-frequency analysis over traditional separate analysis in time or frequency domains are shown. Matching pursuit algorithm and modified psychoacoustic matching pursuit is described. Universal audiocoder scheme based on sparse approximation with frame-based psychoacoustic optimized wavelet packet dictionary is proposed.

Список литературы

1. *Painter T., Spanias A.* // Proceedings of the IEEE. 2000. Vol. 88, № 4. P. 451–513.
2. *Mallat S.G., Zhang Z.* // IEEE Transactions on signal processing. 1993. Vol. 41, № 12. P. 3397–3415.
3. *Petrovsky Al., Petrovsky A.* // Electronica. Konstrukcje, technologie, zastosowania. 2008. № 4. P. 74–80.
4. *Veras-Candeas V., Ruiz-Reyes N., Rosa-Zurera M, et. al.* // IEEE Proceedings on Vision, Image and Signal Processing. February, 2004. Vol. 151, Iss. 1. P. 21–28.
5. Анализаторы речевых и звуковых сигналов: методы, алгоритмы и практика. // Под ред. А.А Петровского. Минск, 2009.
6. *Петровский Ал.А.* // Речевые технологии. 2008. № 4. С. 61–71.
7. *Petrovsky Al., Azarov E., Petrovsky A.* // Elsiver, Signal Processing, Special Issue «Fourier Related Transforms for Non-Stationary Signals». June 2011. Vol. 91, Iss. 6. P. 1489–1504.
8. *Umapathy K., Ghoraani B., Krishnan S.* // EURASIP Journal on Advances in Signal Processing. 2010. Vol. 2010. P. 1–28.
9. *Gilles C., Thibaud N., Peter B.* // ICASSP. 2014. P. 3126–3130.
10. *Ruiz Reyes N., Vera Candeas P.* // IEEE Transactions on audio, speech and language processing. 2010. Vol. 18, № 3. P. 447–460.