К.т.н., доц. Е.И. Прохоренко, А.В. Болдышев, А.В. Эсауленко (Белгородский госуниверситет)

E.I. Prohorenko, A.V. Boldyshev, A.V. Esaulenko

МЕТОД СЖАТИЯ РЕЧЕВЫХ ДАННЫХ НА ОСНОВЕ СОСТАВНОЙ СУБПОЛОСНОЙ МАТРИЦЫ

METHOD OF COMPRESSION OF SPEECH DATA ON BASIS OF COMPOSITE SUB-BAND MATRIX

В статье изложена возможность сжатия речевых данных за счет избирательного воздействия на частотные компоненты речевых сообщений с помощью нового метода субполосного преобразования

Key words informatively-telecommunication technologies, compression of speech data, sub-band transformation, frequency interval, set parts of energy

Введение

Одной из основных тенденций развития информационно-телекоммуникационных технологий является обеспечение естественных для человека форм информационного обмена (речь, визуальные отображения действительности). Реализация информационного обмена речевыми сообщениями, включая их архивное хранение и передачу, осуществляется с помощью компьютерных технологий. При этом речевые сигналы хранятся и передаются в виде некоторых кодовых комбинаций, совокупность которых естественно называть речевыми данными. Поэтому не вызывает сомнения необходимость выбора такого способа кодирования, который обеспечивает минимум объемов битовых представлений, хранимых и передаваемых данных при сохранении приемлемого, с точки зрения пользователя, качества воспроизведения исходных речевых сообщений. Решение этой проблемы позволяет минимизировать затраты объемов компьютерной памяти для хранения данных и времени их передачи в информационно-телекоммуникационных системах (ИТС).

Можно указать достаточно много направлений и областей, для которых решение проблемы минимизации объемов битовых представлений речевых данных имеет существенное значение·

– Корпоративные информационно-телекоммуникацион-

ные системы, в которых используются средства аудио и видео конференц-связи;

– системы постоянного мониторинга речевого и визуального обмена (видеонаблюдение, в т.ч. в аэропортах, вокзалах и т.п.);

– хранение и передача речевых данных средствами Интернет (экспресс - сообщения, голосовая почта и т.д.);

– информационно - телекоммуникационные системы удаленного взаимодействия, в т.ч. системы дистанционного образования.

Таким образом, проблема уменьшения объемов битовых представлений речевых данных (сжатия) является актуальной, а её решение позволит существенно повысить эффективность использования средств ИТС при реализации современного информационного обмена на основе речевых сообщений.

Актуальность этой проблемы можно проиллюстрировать и наличием ряда исследовательских компаний и институтов, которые занимаются обширными и многоаспектными исследованиями для решения различных задач в области обработки речевых данных, например, консорциум «Российские речевые технологии», ООО «Сакрамент» (Беларусь), Институт проблем передачи информации РАН; Институт Фраунгофера и фирма Thomson (формат MP3 pro).

Для сжатия речевых данных разработаны различные процедуры обработки, основой которых служат необратимые преобразования исходных данных, например, за счет более грубого квантования по уровню. Существующие методы сжатия звуковых данных с использованием грубого квантования по уровню основываются на психо-акустической модели (например, формат MP3 pro) [1], что приводит к необходимости применения так называемых субполосных преобразований отрезков (векторов) отсчетов речевых сигналов, позволяющих получить другие векторы, подвекторы которых отражают частотные свойства исходного вектора в выбранных диапазонах оси частот. Компоненты этих подвекторов подвергаются квантованию по уровню. Для субполосного преобразования обычно используется процедура прореживания выходных последовательностей КИХ-фильтров, настроенных на соответствующие участки оси частот. Однако такая процедура субполосного преобразования не является оптимальной в смысле достижения минимальных погрешностей аппроксимации в выбранных частотных диапазонах транс-

формант Фурье исходных векторов. Это приводит к увеличению погрешностей восстановления данных по квантованным значениям и, как следствие, к ухудшению качества воспроизводимой речи.

Таким образом центральной, с точки зрения хранения и передачи, проблемой реализации в ИТС речевого обмена является создание эффективных методов сжатия, полученных на этапе регистрации речевых данных, с возможностью дальнейшего воспроизведения исходных речевых сообщений с приемлемым для пользователя качеством.

Теоретические основы

Одной из особенностей звуков русской речи является сосредоточенность энергии в достаточно узких частотных диапазонах, суммарная ширина которых гораздо меньше частоты дискретизации [2,3]. Эта особенность может быть использована в различных направлениях области обработки речевых сообщений: сжатия речевых данных, обнаружения и кодирования пауз, распознавания речи, очистки от шумов. При этом необходимо точно определять, в каком количестве частотных интервалов сосредоточена необходимая доля энергии Далее эту особенность будем называть частотной концентрацией, которая определятся минимальным количеством частотных интервалов, в которых сосредоточена заданная доля энергии [4]:

$$W'_{NR} = l^{tm}_{NR} / R, \qquad (1)$$

где l^{tm}_{NR} - минимальное количество частотных интервалов, в которых сосредоточена заданная доля энергии звукового отрезка так, что имеет место

$$l^{tm}_{NR} = \min d^{tm}_{NR}, \qquad (2)$$

где: N – значение длины анализируемого отрезка;

R – количество частотных интервалов, на которые разбивается частотный диапазон;

t – обозначает один из звуков русской речи;

m – доля общей энергии, задаваемая для определения минимального количества частотных интервалов, в которых она сосредоточена.

Для правых частей (2) выполняется неравенство

$$\sum_{k=1}^{d_{NR}^{tm}} P_{(k)N} \geq m \| \vec{x}_N \|^2 = m \sum_{i=1}^{N} x_i^2, \qquad (3)$$

где: $\vec{x} = (x_1,\ldots,x_N)^T$ – анализируемый отрезок речевых данных;

T – операция транспонирования.

Индекс в скобках у слагаемых суммы слева соотношения (3) означает, что части энергий P_{kN} упорядочиваются по убыванию:

$$P_{(k)N} \in \{P_{rN}, r = 1,..,R\}; P_{(k+1)N} \leq P_{(k)N}, k = 1,..,R, \qquad (4)$$

В качестве типичного примера в табл. 1 приведено минимальное количество частотных интервалов, составляющих заданную долю энергии, для звука «А».

Таблица 1

Минимальное количество частотных интервалов, составляющих заданную долю энергии (звук «А», N=160, R=16)

m	l^{tm}_{NR}
1	16
0,98	7
0.94-0,96	6
0.92	5
0.9-0.86	4
0.84	3

При заданной доле энергии 0,98 частотная концентрация для приведенного примера составляет 0,31. Для большинства звуков русской речи величина частотной концентрации составляет порядка 0,35 и только для шумоподобных звуков – порядка 0,55-0,60 [4].

Имея сведения о номерах частотных интервалов, в которых сосредоточена заданная доля энергии, можно осуществить сжатие речевых данных за счет избирательного воздействия (например, квантования по уровню) и хранения только составляющих речевого сигнала, соответствующих этим частотным интервалам.

Одним из способов получения этих составляющих речевого сигнала, является субполосное преобразование. В настоящее время наибольшее распространение получил метод субполосного преобразования на основе КИХ-фильтров, однако, этот метод обладает ря-

дом недостатков, которые приводят к увеличению погрешностей восстановления данных [2].

В ряде публикаций [5,6,7,8] описывается новый метод субполосного преобразования, оптимальный с точки зрения минимума среднеквадратической погрешности аппроксимации трансформант Фурье исходного отрезка речевого сигнала в заданном частотном интервале. В этих работах также показаны преимущества этого метода перед современными аналогами.

В основе метода лежит новый математический аппарат с использованием субполосной матрицы:

$$A_r = \{a_{ik}^r\}, i,k = 1,..,N \tag{5}$$

с элементами вида:

$$a_{ik}^r = \{\sin(v_r(i-k)) - \sin(v_{r-1}(i-k))\}/\pi(i-k), i,k = 1,..,N \tag{6}$$

где v_r и v_{r-1} верхняя и нижняя границы частотного интервала.

Эта матрица является симметричной и неотрицательно определённой, поэтому [9] она обладает полной системой ортонормальных собственных векторов, соответствующих неотрицательным собственным числам.

На основе этих матриц можно вычислять точные значения долей энергий отрезков речевых сигналов в выбранных частотных интервалах (5):

$$P_r = \vec{x} A_r \vec{x}^T, \tag{7}$$

где $\vec{x} = (x_1,\ldots,x_N)^T$ – анализируемый отрезок речевых данных,

Для речевых сигналов можно осуществить субполосное преобразование путем формирования так называемой составной матрицы, которая вычисляется как сумма субполосных матриц, соответствующих выбранным частотным интервалам, которые составляют заданную долю энергии m:

$$A_\Sigma = \sum_{k=1}^{d_{NR}^{tm}} A_{(k)} \tag{8}$$

где $A_{(k)}$ – субполосные матрицы, соответствующие тем частотным интервалам, в которых сосредоточена заданная доля энергии m.

Далее, для полученной составной матрицы вычисляются

собственные векторы и соответствующие им собственные числа:

$$Q_\Sigma = \{\vec{q}_{\Sigma 1}, \vec{q}_{\Sigma 2}, ..., \vec{q}_{\Sigma N}\} \quad (9)$$

$$L_\Sigma = diag(\lambda_{\Sigma 1}, .., \lambda_{\Sigma N}) \quad (10)$$

Собственные числа численно равны сосредоточенным в выбранных частотных интервалах долям энергий соответствующих собственных векторов и удовлетворяют условию:

$$0 \le \lambda_{\Sigma k} \le 1,\ k=1,...,N \quad (11)$$

Субполосное преобразование можно вычислить как:

$$\vec{y}_\Sigma = Q_\Sigma^T \vec{x} \quad (12)$$

При этом должно выполняться равенство:

$$\|\vec{x}\|^2 = \|\vec{y}\|^2 \quad (13)$$

Энергию отрезка речевого сигнала можно определить как:

$$P_\Sigma = \sum_{i=1}^{N} \lambda_{i\Sigma} y_i^2 \quad (14)$$

С точки зрения сжатия речевых данных, можно поставить задачу нахождения количества собственных значений составной матрицы, при оставлении которых будет достигаться максимальная степень сжатия при минимальной погрешности восстановления исходного отрезка речевых данных, т.е. будет обеспечиваться высокое качество воспроизведения исходного речевого сообщения.

$$P_\Sigma = \sum_{i=1}^{N} \lambda_{i\Sigma} y_i^2 \approx \sum_{i=1}^{J} \lambda_{i\Sigma} y_i^2 \quad (15)$$

где J - количество собственных значений, необходимое для восстановления с минимальной погрешностью.

Для решения этой задачи были проведены экспериментальные исследования по определению количества собственных чисел, величина которых значительно больше нуля в зависимости от параметра m. На рис. 1 приведены результаты этих исследования для звука «А» при длительности окна анализа N=160 отсчетов и количестве частотных интервалов R=16.

Было оценено количество собственных чисел, величина которых значительно больше нуля ($L>0.01$). Результаты представлены в табл. 2.

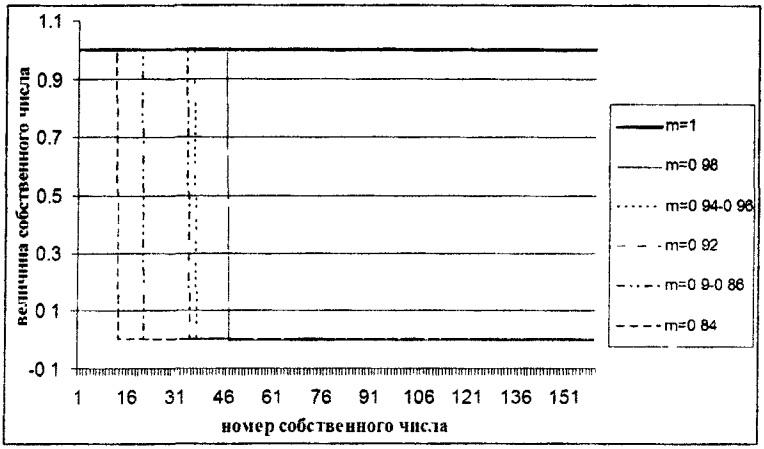

Рис. 1
Значения собственных чисел при различных значениях параметра *m*

Таблица 2

Количество собственных чисел, величина которых значительно больше нуля при различных значениях параметра *m* (звук «А», *N*=160 и *R*=16)

m	**L>0.01**	**L<0.01**
1	160	0
0,98	81	79
0.94-0,96	71	89
0.92	57	103
0.9-0.86	49	111
0.84	38	122

Как видно из табл. 2 уменьшение доли общей энергии *m* влияет на количество собственных чисел, величина которых значительно больше нуля. Так, например, при *m*=0.92 их количество составляет 57. Эта величина принимает значения от 40 до 80 собственных чисел для разных звуков русской речи.

Так как собственные числа численно равны сосредоточенным в выбранных частотных интервалах долям энергий соответствующих собственных векторов, то при осуществлении субполосного преобразования можно пренебречь теми собственными вектора-

ми, собственные числа которых близки к 0 ($L<0.01$). Таким образом, можно уменьшить длину вектора субполосного преобразования, а, следовательно, и объем данных, необходимых для хранения или передачи в 3-4 раза для большинства звуков русской речи.

Субполосное преобразование будет осуществляется по формуле:

$$\vec{y}_{\Sigma} = Q^{T}_{1\Sigma}\vec{x} \qquad (16)$$

где: $\vec{x}=(x_1,\ldots,x_N)^T$ – анализируемый отрезок речевых данных;

$Q^{T}_{1\Sigma}$ – матрица собственных векторов, собственные числа которых значительно больше нуля ($L>0.01$);

На основе вышеизложенных теоретических положений была разработана схема системы обработки речевых данных, представленная на рис. 2.

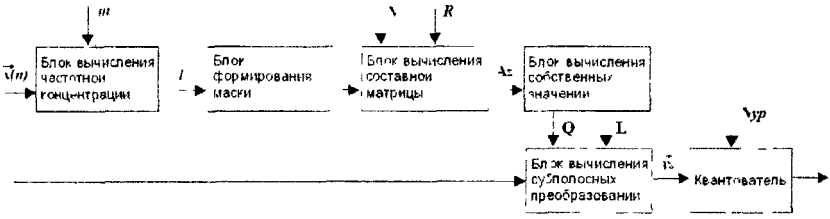

Рис 2

Функциональная схема процедуры сжатия речевых данных

Вектор исходных значений, представляющий собой N отсчетов речевого сигнала, поступает на вход блока вычисления частотной концентрации, где выполняются процедуры вычисления долей энергии, сосредоточенных в каждом из R частотных интервалов (выражение (7)), сортировки полученных значений (4) и определения минимального количества частотных интервалов, в которых сосредоточена заданная доля энергии звукового отрезка согласно выражениям (1), (2). Номера найденных частотных интервалов сохраняются и используются при генерации вектора из R значений (маски) в блоке формирования маски. Согласно данным о номерах частотных интервалов по выражению (8) вычисляются значения составной матрицы, затем ее собственные значения (9), (10). В блоке вычисления субполосных преобразований определяется минимальное количество собственных векторов, достаточное для восстанов-

ления сигнала с малой погрешностью в соответствии с условиями (11 и 15) и собственно субполосное преобразование (выражение (16)). Для увеличения степени сжатия полученные вектора субполосного преобразования могут быть проквантованы по заданному количеству уровней квантования *Nур*.

Обратное субполосное преобразование будет осуществляться по формуле:

$$\hat{\hat{x}} = Q_{1\Sigma}\hat{\hat{y}}_{\Sigma}, \qquad (15)$$

где $\hat{\hat{y}}_{\Sigma}$ - вектор субполосного преобразования, восстановленный по квантованным значениям.

Вычислительные эксперименты

Экспериментальные исследования показали высокую эффективность разработанного метода субполосного преобразования речевых данных с позиции сжатия. В качестве исходных были использованы речевые данные, полученные в результате записи естественной речи различных дикторов, из которых выделялись отрезки, соответствующие определенным звукосочетаниям.

На рис. 3-4 представлены типичные результаты вычислительных экспериментов.

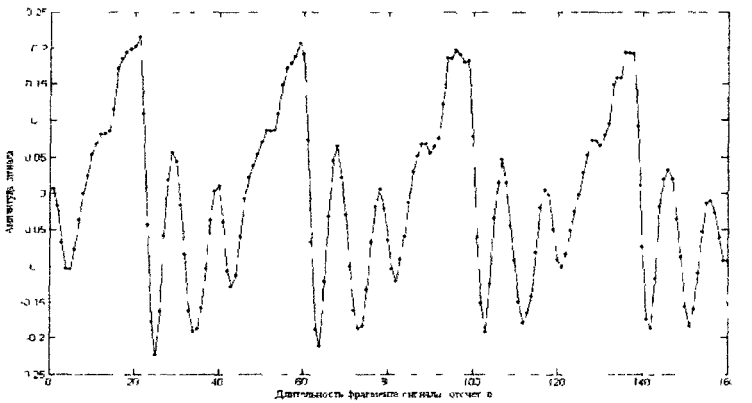

Рис. 3

Отрезок речевого сигнала, соответствующий звуку «А» при $m = 1$. Сплошная линия – исходный сигнал; пунктир (маркер точка) – восстановленный сигнал с использованием составной матрицы (СМ)

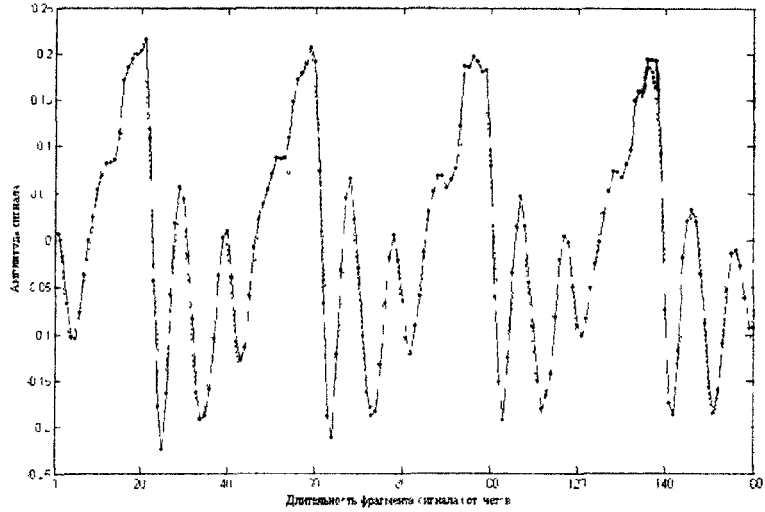

Рис. 4

Отрезок речевого сигнала, соответствующий звуку «А» при $m = 0.9$. Сплошная линия – исходный сигнал; пунктир (маркер точка) – восстановленный сигнал с использованием составной матрицы (СМ)

При m=1 форма сигнала, восстановленного с использованием суммарной матрицы, полностью повторяет форму исходного сигнала, а при уменьшении параметра m форма сигнала незначительно отличается от исходной.

Также была проведена оценка среднеквадратичной относительной погрешности [10] аппроксимации участков спектров исходного вектора в выбранных частотных диапазонах при различных значениях параметра m:

$$\delta_r = \sqrt{\frac{(\vec{x} - \hat{\vec{x}}) * A_r * (\vec{x} - \hat{\vec{x}})^T}{\vec{x} * A_r * \vec{x}^T}} \qquad (12)$$

где: $\vec{x}$ - исходный отрезок речевого сигнала;

$\hat{\vec{x}}$ - восстановленный отрезок речевого сигнала;

A_r - субполосная матрица.

В табл. 3 приведены значения среднеквадратичной относительной погрешности только для тех частотных интервалов, которые составляют заданную долю энергии m.

Таблица 3

Погрешность восстановления отрезка речевого сигнала, соответствующего звуку «А», с использованием выражения (15) при различных значениях m

Номер частотного интервала (R)	Задаваемая доля общей энергии, которая сосредоточена на минимальном количестве частотных интервалов (m)						
	1	0.98	0.96	0.94	0.92	0.9	0.84
1	$6{,}30*10^{-16}$	$1{,}9*10^{-4}$	$1{,}41*10^{-3}$	$1{,}41*10^{-3}$	$1{,}07*10^{-3}$	$1{,}07*10^{-3}$	$8{,}65*10^{-3}$
4	$1{,}07*10^{-15}$	$4{,}1*10^{-4}$	$6{,}01*10^{-3}$	$6{,}01*10^{-3}$	$8{,}58*10^{-3}$	$8{,}58*10^{-3}$	$1{,}44*10^{-2}$
2	$1{,}10*10^{-15}$	$4{,}3*10^{-4}$	$3{,}23*10^{-3}$	$3{,}23*10^{-3}$	$6{,}53*10^{-3}$	$6.53*10^{-3}$	–
3	$1{,}21*10^{-15}$	$6{,}3*10^{-4}$	$4{,}98*10^{-3}$	$4{,}98*10^{-3}$	–	–	–
5	$3{,}03*10^{-15}$	$2{,}88*10^{-3}$	–	–	–	–	–
6	$3{,}14*10^{-15}$	–	–	–	–	–	–

Результаты, приведенные в табл. 3 свидетельствуют о том, что субполосное преобразование на основе составной матрицы позволяет достигать небольших величин погрешности восстановления исходных речевых данных, а, следовательно, более точно восстанавливает исходный отрезок речевого сигнала. Этот факт позволяет говорить о целесообразности использования такого подхода в задачах обработки речевых данных и сжатия, т.к. близость формы восстановленного сигнала к исходной позволяет говорить о достаточно высокой степени разборчивости и узнаваемости диктора.

Выводы

Проведенные вычислительные эксперименты показали, что метод субполосного преобразования на основе составной матрицы можно использовать для сжатия речевых данных, при этом возможно получить коэффициент сжатия порядка 3-4 раз, без учета квантования результатов субполосного преобразования.

Экспериментально было установлено, что предлагаемый метод субполосного преобразования на основе использования составной матрицы обладает малой погрешностью восстановления. Таким образом, использование нового метода позволяет со значительной степенью точности восстановить исходный отрезок речевых данных. Этот факт подтверждает сохранение высокой степени разборчивости и узнаваемости в восстановленных речевых сообщениях.

Литература

1. Ковалгин Ю.А. и Вологдин Э.И. Цифровое кодирование звуковых сигналов: Учеб. пособ. СПб., КОРОНА-принт, 2004. 240 с.

2. Жиляков Е.Г., Белов С.П. и Прохоренко Е.И. Методы обработки речевых данных в информационно-телекоммуникационных системах на основе частотных представлений: Белгород, 2007. 136 с.

3. Шелухин О.И. и Лукьянцев Н.Ф. Цифровая обработка и передача речи. Под ред. О.И. Шелухина. М., Радио и связь, 2000. 456 с. с илл.

4. Болдышев А.В. и Фирсова А.А. О различиях распределения энергии звуков русской речи и шума. – В сб.: Цифровая обработка сигналов и её применение - DSPA'2010. Материалы 12-ой Междунар. конф. М., 2010.

5. Прохоренко Е.И., Болдышев А.В., Фирсова А.А. и Эсаулен-

ко А.В. Новый метод оптимального субполосного преобразования в задаче сжатия речевых данных. – «Вопросы радиоэлектроники», сер. ЭВТ, 2010, вып. 1, с. 49-55.

6. Жиляков Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным. Белгород, БелГУ, 2007. 160 с.

7. Жиляков Е.Г., Белов С.П. и Прохоренко Е.И. Вариационные методы частотного анализа звуковых сигналов. – В сб.: Труды учебных заведений связи, 2006, № 174 с. 163-170.

8. Жиляков Е.Г., Белов С.П. и Прохоренко Е.И. О субполосном преобразовании звуковых сигналов. "Труды РНТО РЭС им. А.С. Попова", .сер. Цифровая обработка сигналов и ее применение, 2006, вып. VIII-1, с. 167-169.

9. Гантмахер Ф.Р. Теория матриц. М., Физматлит, 2004. 560 с.

10. Сизиков В.С. Математические методы обработки результатов измерений. Учеб. для вузов. СПб., Политехника, 2001.

Статья поступила 12.10 2010

А.А. Фирсова, д. ф.-м. н. А.Н. Чеканов (БелГУ)

A.A. Firsova, A. N. Chekanov

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ АЛГОРИТМОВ ОБНАРУЖЕНИЯ ПАУЗ В IP-ТЕЛЕФОНИИ

COMPUTER MODELLING OF ALGORITHMS OF DETECTION OF PAUSES IN IP — TELEPHONY

В статье рассмотрены различные алгоритмы обнаружения пауз Проведен анализ возможности использования различных алгоритмов обнаружения пауз в режиме реального времени Проведено сравнение эффективности использования различных алгоритмов

Keywords. speech signal, speech signal analysis, a model of VAD, pause detection algorithm, the frequency representation, IP-telephony

Развитие информационно-телекоммуникационных систем направлено на обеспечение возможности использования естественных форм общения (речь, изображение) с помощью современных средств обработки сигналов. В настоящее время обработка сигналов