УДК 621.391

ИСПОЛЬЗОВАНИЕ ПСИХОАКУСТИЧЕСКОЙ МОДЕЛИ СЛУХА ПРИ РАЗРАБОТКЕ ВОКОДЕРОВ С ЛИНЕЙНЫМ ПРЕДСКАЗАНИЕМ

Афанасьев А.А., *Академия ФСО России, 8-(4862)-41-99-47*
Илюшин М.В., *Академия ФСО России, 8-(4862)-41-99-47*

Для улучшения восприятия синтезированной речи при реализации процедуры анализа через синтез предлагается выбор наилучших в рамках заданных ограничений параметров кодека с линейным предсказанием производить на основе вычисления оценок по критерию модифицированного спектрального искажения MBSD (Modified Bark Spectral Distortion) с применением психоакустической модели слуха человека.

В настоящее время наблюдается широкое использование информационных технологий в телекоммуникационных сетях связи. Переход к цифровой обработке сигналов и пакетной передаче данных позволил предоставить пользователям более широкий спектр инфокоммуникационных услуг. При этом достаточно большую часть телетрафика в различных приложениях составляет передача речевых сигналов.

Основной проблемой цифрового представления речевого сигнала является задача качественного и компактного кодирования данных для их передачи по цифровым каналам связи. Решение этой проблемы позволит в условиях заданного критерия качества связи увеличить пропускную способность линейных трактов и каналов передачи. Часто в некоторых задачах кодирования речевого сигнала предполагается снизить скорость передачи при сохранении качественных показателей ее восприятия. В кодеках речевых сигналов с переменной скоростью передачи, ориентированных на использование в системах связи, основанных на принципе коммутации пакетов, уместно говорить о снижении средней скорости передачи при сохранении качественных показателей синтезированного речевого сигнала.

Среди многообразия методов кодирования речевых сигналов одним из наиболее эффективных является метод линейного предсказания. Метод линейного предсказания речи принадлежит к классу методов, использующих модель речевого сигнала в виде отклика линейной системы с переменными параметрами (голосового тракта) на соответствующий сигнал возбуждения (порождающий сигнал). Анализатор речепреобразующего устройства выделяет из короткого сегмента речевого сигнала параметры состояния линейной системы и сигнала возбуждения, позволяющие синтезатору восстановить исходный сигнал с требуемой степенью верности. Для повышения качества синтезированного речевого сигнала во многих алгоритмах кодирования речи на основе линейного предсказания усложняют представления сигнала возбуждения для того, чтобы с одной стороны компактно передать его на приемную сторону, а с другой – приблизить его к виду ошибки предсказания, как идеальному сигналу воздействия на фильтр синтеза. При этом вводятся разные варианты квантования различных параметров липредера (скалярное, векторное и каскадное векторное) [1].

Известны различные алгоритмы низкоскоростного кодирования речи в вокодерах с линейным предсказанием. Во многих из них одной из базовых операций является процедура анализа через синтез. Достаточно подробно ее описание представлено в [2]. Данная процедура является итерационной и направлена на вычисление наилучших в рамках заданных ограничений параметров кодека с линейным предсказанием.

Исследования в области речевого кодирования указали на необходимость использования перцептуальных особенностей слуха человека [3]. До сих пор в качестве критерия выбора параметров кодека с линейным предсказанием при реализации процедуры анализа через синтез используются либо среднеквадратическое отклонение, либо суммарное или сегментированное отношение сигнал/шум, основанные на метрике Евклида и не учитывающие перцептуальную важность параметров кодека при синтезе речевого сигнала.

При вычислении среднеквадратического отклонения допускается, что искажения, вносимые каждым элементом вектора, имеют равный вес. В общем случае для отражения вклада отдельных элементов в искажение вводятся неравные веса в виде взвешивающей матрицы. Указанный метод позволяет лишь сравнивать форму огибающих исходного и синтезированного речевого сигнала. Поэтому для количественной оценки качества звучания синтезированного речевого сигнала во временной области чаще используют критерий отношения сигнал/шум, который учитывает общие мощности сигнала и шума на всей длительности испытательного сигнала. При исследовании некоторых речевых кодеков большое значение имеют кратковременные отношения сигнал/шум, вычисленные на коротких сегментах речевого сигнала. Таким образом, учитывается сегментный характер слухового восприятия элементов речи.

Однако, приведенные критерии объективного метода оценивания отражают степень зашумленности речевого сигнала и показывают слабую корреляцию с результатами субъективных тестов при прослушивании речевых сегментов. Следует отметить, что если качество кодеров формы речевой волны может быть оценено по степени соответствия формы огибающей восстановленного речевого сигнала исходному с помощью названных критериев, то для алгоритмов низкоскоростного параметри-

ческого сжатия на основе линейного предсказания точное восстановление формы сигнала является сложной задачей. Следовательно, методы оценивания качества звучания синтезированного речевого сигнала во временной области мало применимы. Для того чтобы оценка качества звучания речевого сигнала отражала критерии слухового восприятия, принципы ее формирования должны быть основаны на анализе спектрально-корреляционных характеристик речи.

Недостатком процедуры анализа через синтез, используемой в стандартизированных алгоритмах низкоскоростного кодирования речевого сигнала, является несоответствие слухового аппарата человека при восприятии синтезированной речи и используемых критериев близости, определяющих правила анализа пригодности выбранных параметров кодека. В классических алгоритмах для выполнения процедуры анализ через синтез в вокодере с линейным предсказанием на передающей стороне итерационно синтезируется речевой сигнал на длительности участка квазистационарности речи и при каждой итерации изменяются параметры кодека в соответствии с используемым алгоритмом линейного предсказания. На каждой итерации вычисляется среднеквадратическая ошибка между оригинальным и синтезированным речевым сигналом, находится итерация, соответствующая наименьшей среднеквадратической ошибке. При этом параметры кодека, соответствующие данной итерации, считаются наилучшими и на основе их формируется кадр передачи кодека и производится синтез речевого сигнала на длительности участка квазистационарности речи на приемной стороне.

Для улучшения восприятия синтезированной речи при реализации процедуры анализа через синтез в вокодерах с линейным предсказанием предлагается выбор наилучших в рамках заданных ограничений параметров кодека с линейным предсказанием производить на основе вычисления оценок по критерию модифицированного спектрального искажения MBSD (Modified Bark Spectral Distortion) (1):

$$\mathrm{MBSD} = \frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{K} M(n,i)\big(D(n,i)\big), \qquad (1)$$

где $M(n,i)$ и $D(n,i)$ – индикатор искажений уровня ощущения и значение разницы интенсивности ощущения сигнала *n*-го сегмента речи в *i*-й критической полосе;
N – число сегментов в речевом фрагменте;
K – общее количество критических полос.

Использование такого подхода позволяет устранить несоответствие слухового аппарата человека при восприятии синтезированной речи и используемых критериев близости, определяющих правила анализа пригодности выбранных параметров кодека при реализации процедуры анализа через синтез.

Данный критерий является наиболее предпочтительным, так как при его использовании производится анализ спектрально-корреляционных характеристик речи с учетом модели слуха человека. При этом, он показывает высокую корреляцию с оценками, полученными на основе субъективных тестов прослушивания. Экспериментальные исследования показали, что в случае применения низкоскоростных липредерных систем слуховой аппарат человека более чувствителен к возникающим при этом частотным искажениям, нежели к амплитудным и фазовым [4]. Расчет спектрального представления, учитывающего психоакустическое восприятие речи, производится согласно выражению (2):

$$b[\text{барк}] = 13\cdot arctg(0{,}00076\, f[\text{Гц}]) + 3{,}5\cdot arctg\left(\left(\frac{f[\text{Гц}]}{7500}\right)^2\right), \qquad (2)$$

где f – частота, измеренная в Герцах;
b – частота, измеренная в барках.

Более подробно данный вопрос изложен в [5]. Подробное описание критерия MBSD можно найти в [6, с.63-75].

Согласно данному критерию, синтезированный и оригинальный речевые сигналы на сегменте квазистационарности подвергаются делению на критические полосы в каждой из которых вычисляется интенсивность ощущения сигнала и порог шумового маскирования, далее в каждой полосе определяется разность между оригинальным и искаженным значением интенсивности ощущения. Если полученное значение $D(n,i)$ превышает вычисленный порог шумового маскирования $NMT(n,i)$, то индикатору искажения уровня ощущения $M(n,i)$ присваивается значение 1, в противном случае значение 0.

Известны работы [7, 8], в которых авторы предлагают использовать кепстральное расстояние (CD – Cepstral Distance) [7] и разность значений громкости [8] между сегментами исходного и синтезированного речевого сигнала в целях выбора оптимальных параметров системы преобразования речевого сигнала. Указанные меры искажений имеют определенные недостатки.

При вычислении CD не учитываются особенности восприятия речи аудиторной системой человека, а приведенные автором сведения о коэффициенте корреляции с результатами средней оценки мнений MOS (R=0,93) [7] имеют расхождения с данными, представленными в [9] (R=0,63).

Необходимо отметить, что идея, опубликованная в [8], не отражает в полной мере применение психоакустической модели восприятия речи как критерия выбора наилучших в рамках заданных ограничений параметров кодека с линейным предсказанием при реализации процедуры анализа через синтез. В [8] авторы предлагают использовать меру искажений громкостей исходного и синтезированного речевого сигнала. В процессе расчета указанной меры в рамках процедуры анализа через синтез не определяются пороги шумового маскирования в критических частотных полосах, что отрицательно влияет на точность определения перцептуально значимых искажений речевого сигнала. Этот факт подтверждается результатами исследований, представленными в [6, с. 72-74].

Алгоритм функционирования предложенной системы для улучшения восприятия синтезированной речи при реализации процедуры анализа через синтез в вокодерах с линейным предсказанием включает процедуры вычисления порогов шумового маскирования и громкости сигналов в критических частотных полосах. Необходимо отметить, что в предложенной системе разделение частотной шкалы происходит таким образом, чтобы отразить частотную избирательность аудиторной системы человека и формантную структуру речевого сигнала. Параметры модифицированной схемы разделения частотной шкалы приведены в [10].

Алгоритм функционирования предложенного метода представлен на рис. 1.

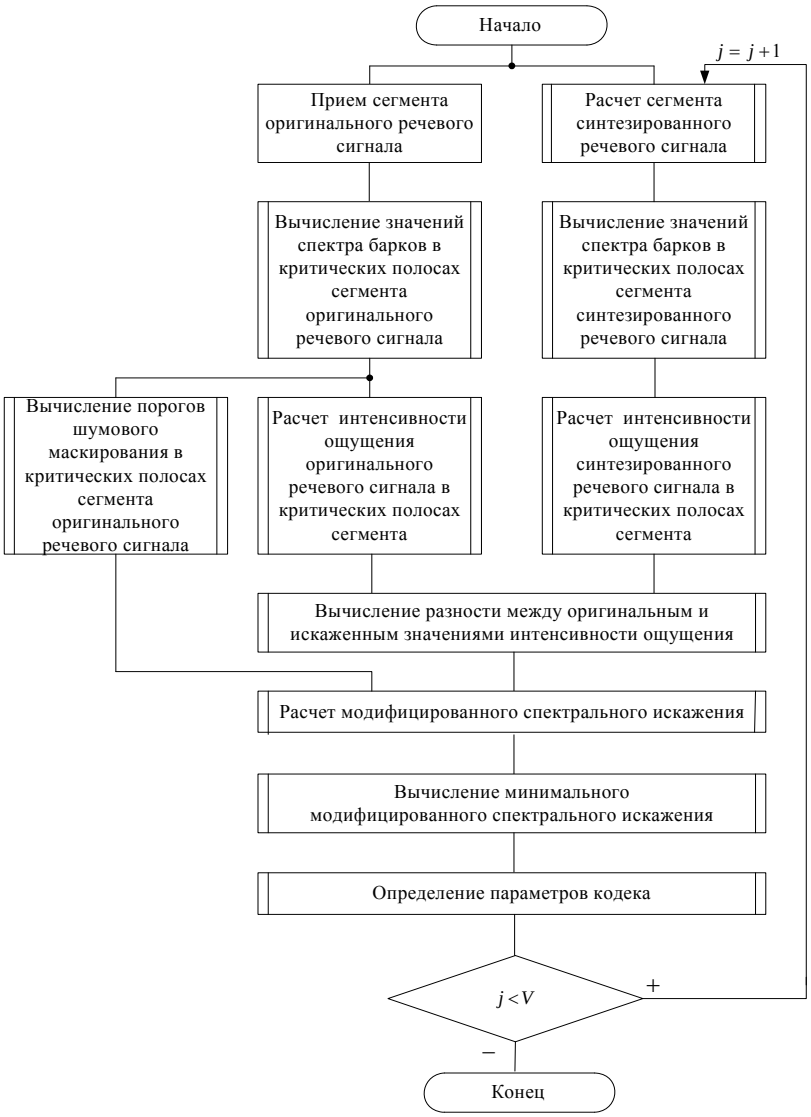

Рис.1 Алгоритм реализации процедуры анализа через синтез, учитывающий психоакустическую модель слуха человека

Таким образом, при реализации процедуры анализа через синтез в вокодерах с линейным предсказанием на передающей стороне итерационно синтезируется речевой сигнал на длительности участка квазистационарности речи, при этом на каждой итерации изменяются параметры кодека в соответствии с используемым алгоритмом линейного предсказания. Для вычисления наилучших в рамках заданных ограничений параметров кодека с линейным предсказанием предлагается ввести итерационный расчет критерия модифицированного искажения спектра барков. При каждой итерации будут изменяться параметры кодека в соответствии с используемым алгоритмом линейного предсказания, для каждой итерации производится расчет минимального перцептуального искажения. Параметры кодека, соответствующие итерации с минимальным искажением, считаются наилучшими и используются для синтеза речевого сигнала на длительности участка квазистационарности речи на приемной стороне.

К достоинствам предлагаемого подхода следует отнести тот факт, что в вокодерах с линейным предсказанием устраняется несоответствие слухового аппарата человека при восприятии синтезированной речи и используемых критериев близости, определяющих правила анализа пригодности выбранных параметров кодека при реализации процедуры анализа через синтез. Были проведены экспериментальные испытания согласно «ГОСТ Р 51061-97. Системы низкоскоростной передачи речи по цифровым каналам. Параметры качества речи и методы измерений» (М.: Госстандарт России, 1997. – 230 с.), которые показали, что применение данного способа позволяет повысить субъективное качество восприятия синтезированной речи в среднем на 0,11 балла.

Заключение

Поставленная задача кодирования с линейным предсказанием при реализации процедуры анализа через синтез достигается путем исключения итерационного расчета и минимизации среднеквадратической ошибки, основанной на метрике Евклида. При этом анализ и выбор наилучших в рамках заданных ограничений параметров кодека с линейным предсказанием производят на основе вычисления оценок по критерию модифицированного искажения спектра барков MBSD (Modified Bark Spectral Dis-

tortion), который рассчитывают на каждом квазистационарном сегменте анализа речевого сигнала.

Полученные решения указывают на возможность обеспечить более качественное восприятие синтезированной речи в вокодерах с линейным предсказанием за счет учета психоакустических особенностей слуха человека, реализация которых основана на выполнении процедуры анализа через синтез.

Литература

1. Быков С. В. Цифровая телефония: Учеб. пособие для вузов/ В. И. Журавлев, И. А. Шалимов – М.: Радио и связь, 2003. – С. 66-72.
2. Шелухин О. И. Цифровая обработка и передача речи / Н. Ф. Лукьянцев, М.: Радио и Связь, 2000г. – С. 102-166;
3. Попов, О. Б. Цифровая обработка сигналов в трактах звукового вещания / С.Г. Рихтер, Учебное пособие для вузов. – М.: Горячая линия. Телеком, 2007. – с. 341.
4. Павловец А. Н., Использование закономерностей психоакустики в процедуре квантования параметров гармонической модели речевого сигнала / Петровский А.А.. // Речевые технологии. 4, 2008, С. 55-60.
5. Радзишевский А. Ю. Основы аналогового и цифрового звука - М.: Изд.дом "Вильямс", 2006 – С. 105-109.
6. Yang, W. Enhanced Modified Bark Spectral Distortion (EMBSD): An Objective Speech Quality Measure Based On Audible Distortion And Cognition Model / A Dissertation of the Requirement for the Degree Doctor of Philosophy – May, 1999.
7. Поляков А. Н. Об одном из способов решения задачи определения оптимальных управляющих параметров системы низкоскоростной компрессии речевой информации // Телекоммуникации. 2008. №3. – С. 15–18.
8. Hauenstein, M. “On the application of a psychoacoustically motivated speech-quality measure in CELP speech-coding” / N. Goertz. in the proc. Of the 9th European Signal processing conference (EUSIPCO’98), vol. III, pp. 1421-1424, Rhodes, Greece, 1998.
9. Шалимов И. А. Практикум по цифровой телефонии: учеб. пособие / Академия ФСБ России, 2008. – 344 с. 111 ил., 139 табл.
10. Лившиц М. З. Широкополосный CELP – кодер с мультиполосным возбуждением и многоуровневым векторным квантованием по кодовой книге с реконфигурируемой структурой / М. Парфенюк, А. А. Петровский. Цифровая обработка сигналов, № 2, 2005. – С. 20–35.

USE OF PSYCHOACOUSTIC MODEL OF HEARING BY WORKING OUT VOCODERS WITH THE LINEAR PREDICTION

Afanasjev A. A., Ilushin M. V.

Materials of given article can be used in systems of teleinformation communications for effective coding of speech signals. A main objective of the presented work is improvement of perception of the synthesised speech at realisation of the analysis procedure through synthesis in vocoders with a linear prediction. The task in view in vocoder with a linear prediction at realisation of procedure of the analysis through synthesis is reached by an exception of iterative calculation and minimisation of the mean square error based on metrics Evclid. Thus the analysis and a choice of the best within the limits of the set restrictions of parametres of the codec with a linear prediction make on the basis of calculation of estimations by criterion of the modified bark spectral distortion criterion (MBSD) which count on everyone near stationary segment of the analysis of a speech signal.

У в а ж а е м ы е а в т о р ы !

Редакция научно-технического журнала "Цифровая обработка сигналов" просит Вас соблюдать следующие требования к материалам, направляемым на публикацию:

1) Требования к текстовым материалам и сопроводительным документам:

- ***Текст - текстовый редактор Microsoft Word.***
- ***Таблицы и рисунки должны быть пронумерованы. На все рисунки, таблицы и библиографические данные указываются ссылки в тексте статьи.***
- ***Объем статьи до 12 стр. (шрифт 12). Для заказных обзорных работ объем может быть увеличен до 20 стр.***
- ***Название статьи на русском и английском языках.***
- ***Рукопись статьи сопровождается:***
  - ***краткой аннотацией на русском и английском языках;***
  - ***номером УДК;***
  - ***сведениями об авторах (Ф.И.О., организация, должность, ученая степень, телефоны, электронная почта);***
  - ***ключевыми словами;***
  - ***актом экспертизы (при наличии в вашей организации экспертной комиссии).***

2) Требования к иллюстрациям:

Векторные (схемы, графики) - желательно использование графических редакторов Adobe Illustrator или Corel DRAW.

- ***Растровые (фотографии, рисунки) - М 1:1, разрешение не менее 300dpi, формат tiff.***