ISSN 1993-9981 Методи та прилади контролю якості, № 2 (31), 2013

УДК 004.934

МЕТОД АССОЦИАТИВНОГО СЖАТИЯ РЕЧИ ДЛЯ УСЛУГ АСИНХРОННОЙ ПЕРЕДАЧИ ГОЛОСОВЫХ СООБЩЕНИЙ

Е.В. Добровольский**, *М.В. Чута

Одесская национальная академия связи им. А.С. Попова, ул. Кузнечная 1, г. Одесса, 65029, e-mail: onat@onat.edu.ua

Запропоновано метод компресії цифрованої мови, оптимізований для асинхронної передачі голосових повідомлень із високою якістю. В основі методу лежить структурно-статистичний аналіз, асоціативне моделювання форми хвилі, лінійне передбачення та інтервальне кодування. Компресор і декомпресор використовують спільний словник асоціацій, який формується на основі аналізу еталонної сукупності зразків мови.
Ключові слова: метод стиску, форма хвилі, асоціативне моделювання, асинхронна передача, голосове повідомлення.

Предложен метод сжатия речи, оптимизированный для асинхронной передачи голосовых сообщений с высоким качеством. В основе метода лежит структурно-статистический анализ, ассоциативное моделирование формы волны, линейное предсказание и статистическое интервальное кодирование. Компрессор и декомпрессор используют общий словарь ассоциаций, который формируется на основе анализа эталонной совокупности образцов речи.
Ключевые слова: метод сжатия, форма волны, ассоциативное моделирование, асинхронная передача, голосовое сообщение.

A new method was offered for high quality speech compression for voice messaging services. The method utilizes waveform modeling technique based on associative dictionary, linear prediction and range entropy encoding for links, residuals and voice activity intervals. Compressor and decompressor use a common associative dictionary built by analyzing a representative master set of speech samples.
Keywords: near-lossless waveform compression, asynchronous voice messaging.

Вступление

Услуги асинхронной передачи голосовых сообщений (voice messaging, голосовой чат, рация) [1÷3] являются весьма востребованными в настоящее время, о чем свидетельствует большое количество приложений, реализующих доступ к этим услугам как на различных мобильных платформах (iOS, Android, WP7, WM), так и на персональных компьютерах. С помощью данных услуг осуществляется персональная или групповая рассылка цельных фрагментов речи в симплексном режиме. Сценарии использования асинхронной передачи голосовых сообщений (АПГС) ориентированы как на индивидуальных пользователей, так и на разнообразные диспетчерские службы. С помощью механизмов АПГС также могут быть реализованы интерактивные интерфейсы в системах голосового самообслуживания (СГС), которые позволяют пользователю, без участия оператора, самостоятельно получать необходимую информацию, выполнять заказ и управление различными сервисами с помощью голосовых команд.

При использовании мобильных и публичных беспроводных сетей для передачи медиаданных и, в частности, голоса остро стоит вопрос предоставляемого сетью уровня параметров качества обслуживания (QoS, quality of service), прежде всего допустимого процента потерь пакетов, величины задержки и джиттера. Так, зачастую, при доступе через «мобильный интернет» из-за больших потерь либо задержек отсутствует возможность качественной двусторонней дуплексной передачи голоса в приложениях IP телефонии. Услуги же АПГС допускают реализацию устойчивую к потерям, большим задержкам и джиттеру и, в данной ситуации, сохранят пользователям возможность голосовой коммуникации.

Для мобильных пользователей всегда остро стоит вопрос экономии трафика, в особенности при нахождении в роуминге. Большинство применяемых на практике методов сжатия речи оптимизированы для потоковой передачи и не реализуют потенциального выигрыша в степени сжатия от многопроходной обработки, возможной при сжатии голосового сообщения целиком.

В то же время выигрыш в степени сжатия может быть использован для введения управляемой избыточности, повышающей надежность доставки в условиях передачи данных по сетям без поддержки механизмов обеспечения качества облуживания [4, 5], что актуально при доступе через публичные беспроводные сети.

Механизмы распознавания речи в системах голосового самообслуживания чувствительны к искажениям, которые вносятся современными вокодерами (LPC, CELP, MELP алгоритмы [6÷8]), моделирующими голосовой тракт человека. В то же время стандартные кодеки восстанавливающие форму волны (как, например, ADPCM G.726 [9, 10]) обеспечивают очень высокое качество передачи голоса, однако не позволяют добиться высокой степени сжатия.

Цель статьи

Вышеперечисленные факторы в совокупности со стремительным ростом вычислительных ресурсов платформ (в том числе мобильных), поддерживающих услуги АПГС, создают предпосылки для создания и развития методов сжатия голоса оптимизированных под данный вид услуг, сочетающих высокую степенью сжатия характерную для вокодеров с возможностью восстановления исходной формы волны.

Изложение основного материала

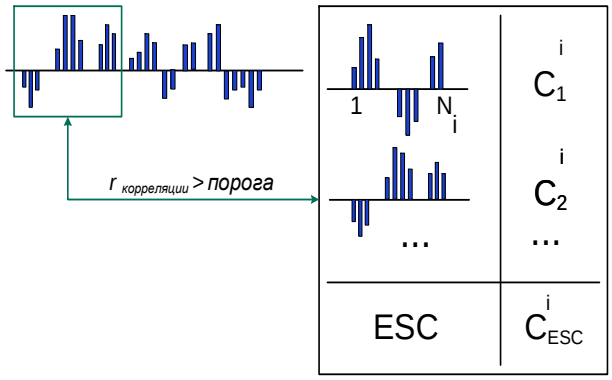
В рамках данной задачи нами предложен и совершенствуется метод сжатия речи с потерями, реализующий восстановление формы волны, близкой к исходной. Метод основан на моделировании элементов исходной последовательности отсчетов (samples) с помощью т.н. «словарей ассоциаций» (СА). В методе также применяются: структурный анализ, линейное предсказание и интервальное кодирование (range encoding [11÷13]).

СА представляет собой совокупность таблиц (рис. 1), содержащих т.н. «ассоциации» - последовательности величин имеющих высокую вероятность корреляции с фрагментами последовательностей отсчетов в сжимаемых данных.

Начальный СА формируется на основе анализа эталонной совокупности голосовых фрагментов и входит в состав кодера и декодера. Критерием добавления текущего фрагмента в СА служит низкая корреляция с уже имеющимися фрагментами в словаре. Как мера корреляции используется выборочный линейный парный коэффициент корреляции Пирсона:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}, \qquad (1)$$

где x_i – отсчеты сжимаемой последовательности, y_i – значения элементов СА, а $\overline{x}$ и $\overline{y}$ их средние выборочные, соответственно.



i – номер таблицы; N_i – длина «ассоциаций» в таблице; C^i_j – счетчик, реализующий частотный рейтинг j-ой ассоциации в i-ой таблице

Рисунок 1 – Таблица ассоциаций, составная часть словаря

Если $r < TR_{вкл.}$ (подсчитанный попарно) для текущего фрагмента и всех ассоциаций словаря, то фрагмент добавляется в таблицу, в противном случае увеличивается счетчик частотного рейтинга тех ассоциаций, для которых $r < TR_{вкл.}$. Порог $TR_{вкл.}$ является параметром алгоритма. Таким образом, в процессе последовательного перебора фрагментов эталонной совокупности происходит либо их добавление в СА, либо увеличение счетчиков частотного рейтинга соответствующих коррелированных ассоциаций. Ассоциации с низким частотным рейтингом затем удаляются из словаря для увеличения скорости поиска при сжатии/восстановлении.

Аналогично в процессе сжатия (либо восстановления) в СА могут динамически добавляться фрагменты сжимаемого (либо восстанавливаемого) файла, не коррелирующие с уже находящимися в словаре.

Процесс сжатия в предлагаемом методе

включает следующие этапы (рис.2):

1) предварительная обработка исходной последовательности отсчетов:

а) удаление всплесков (необходимо для корректного выполнения нормализации);

б) нормализация (изменение громкости фонограммы, позволяющее использовать максимум доступного динамического диапазона, не оставляющее амплитудного зазора). Нормализация необходима для эффективного поиска ассоциаций в словаре. В процессе нормализации выполняются: поиск наибольшего значения амплитуды, вычисление разницы между заданным параметром нормализации и максимальным отсчетом амплитуды, усиление до значения этого параметра;

в) очистка от импульсных шумов [14] для увеличения эффективности выделения участков голосовой активности на следующем этапе;

2) выделение участков голосовой активности (УГА) в исходной последовательности отсчетов. На дальнейших этапах обрабатываются только отсчеты принадлежащие УГА. В результирующем файле сохраняются наборы длин УГА и пауз между ними;

3) последовательное сканирование множества отсчетов на наличие участков «схожих» с элементами словаря. Под схожестью понимается наличие корреляции с коэффициентом выше порогового значения TRпоиска, являющегося параметром алгоритма. Дополнительно, порогом $TR_{пика}$, ограничивается максимально допустимое различие отсчетов исследуемого на предмет корреляции участка и элемента словаря:

$$r > TR_{поиска}, \qquad (2)$$

$$\max\left|y_j - x_j\right| < TR_{пика}. \qquad (3)$$

Сравнение начинается с таблицы словаря, содержащей ассоциации максимальной длины (которая задается как параметр алгоритма). Если по условиям (2) и (3) была найдена подходящая ассоциация, то на блок статистического кодирования отправляется ссылка на элемент таблицы, ее частотный рейтинг C_j^i (определяющий кодовое пространство, выделяемое на этапе кодирования для данной ссылки) и квантованный масштабирующий множитель k, обеспечивающий минимизацию расстояния, т.е.:

$$\sum_{j=1}^{Ni}\left|y_j - kx_j\right| \Rightarrow \min. \qquad (3)$$

Если подходящей ассоциации для данной длины не было найдено, поиск начинается в таблице с меньшей (на один шаг) длиной ассоциаций. В то же время на блок кодирования подается символ ухода (ESC) информирующий декодер о переходе к поиску ассоциаций меньшей длины. Кодовое пространство для символа ухода выделятся исходя из его динамически формируемого частотного рейтинга.

Далее механизм поиска работает аналогично, в случае неудач постепенно спускаясь к таблице с минимальной длинной ассоциаций (которая задается как параметр алгоритма). Если поиск был неудачен и для этой таблицы, то текущий отчет моделируется с помощью линейного предиктора первого порядка. Разница между истинным и модельным значением отсчета в этом случае квантуется и передается на вход блока интервального кодирования. Кодовое пространство при этом распределяется согласно заранее просчитанной статической таблице частот, которая является параметром алгоритма.

Таким образом, результирующий файл для предложенного метода состоит из следующих потоков структурно□статистической информации:

– длины участков голосовой активности и пауз;

– вероятности уходов, рассчитанные для всех словарей;

– ссылки на найденные ассоциации;

– коэффициенты масштабирования;

– квантованные дельты (для отсчетов моделируемых линейным предсказанием).

Предложенный метод был реализован в среде Matlab. Достигнутая степень сжатия (исходный файл PCM 16 бит 8кГц) составила порядка 6:1 (около 2,7 бит/отсчет) при качестве восстановленной речи сопоставимом с ADPCM 32 (4 бита/отсчет).

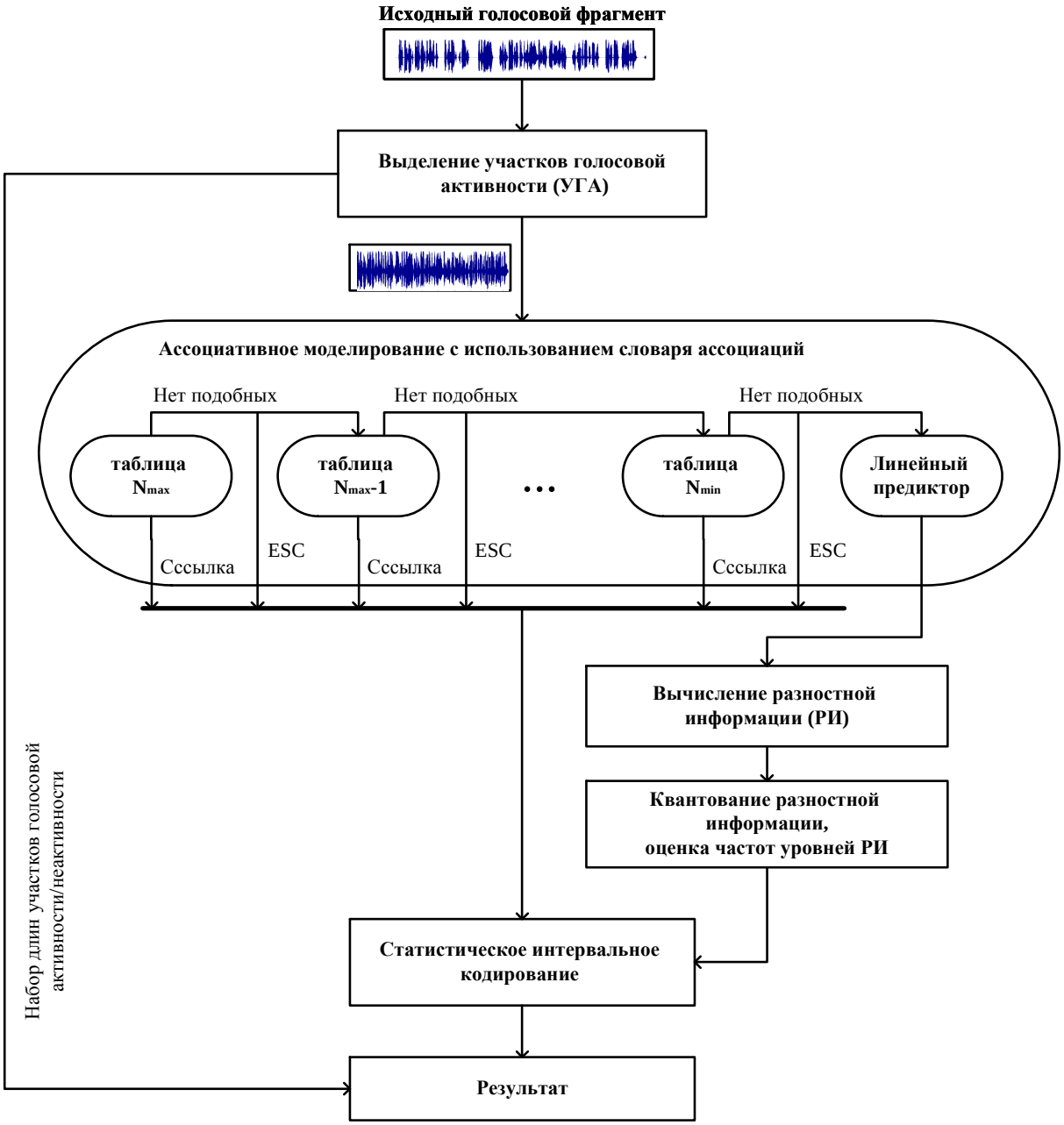


Рисунок 2 – Схематическое изображение основных этапов ассоциативного сжатия

ВЫВОДЫ

Несмотря на большую популярность вокодерных методов сжатия речи, для некоторых задач применение алгоритмов восстанавливающих форму волны не потеряло свою актуальность, в частности, они могут быть с успехом применены в приложениях передачи голосовых сообщений, обеспечивая высокое качество восстановления речи.

Предложенный метод показывает наличие у алгоритмов сжатия голоса, относящихся к классу восстанавливающих форму волны, далеко не исчерпанного потенциала по улучшению эффективности сжатия, например, путем создания статистических моделей связанных со структурными элементами сжимаемых данных.

1. Voicemail / Wikipedia, the free encyclopedia [Электронный ресурс]. – Режим доступа: http://en.wikipedia.org/wiki/Voicemail. 2. Voice Messaging Comes To Whatsapp / TechCrunch

[Электронный ресурс]. – Режим доступа: http://techcrunch.com/2013/08/07/voice-messaging-comes-to-whatsapp/. 3. Voice Messaging Services/Alcatel-Lucent Enterprise [Электронный ресурс]. – Режим доступа: http://enterprise.alcatel-lucent.com/?product=VoiceMessagingServices&page=overview. 4. Вегешна Ш. Качество обслуживания в сетях IP./ Шринивас Вегешна — Cisco Press, 2003. – 356 с. 5. RFC 2212 - Specification of Guaranteed Quality of Service. [Электронный ресурс]. – Режим доступа: http://tools.ietf.org/rfc/rfc2212.txt. 6. ITU-T Recommendation G.729 — Coding of speech at 8 kbit/s using conjugate-structurealgebraic-code-excited linear prediction (CS-ACELP). – 1996. 7. LPC10 presentation, Soo Hyun Bae, ECE 8873 Data Compression & Modeling, Georgia Institute of Technology, 2004 [Электронный ресурс]. – Режим доступа: http://users.ece.gatech.edu/~juang/8873/Bae-LPC10.ppt. 8. MELP Vocoder Algorithm [Электронный ресурс]. – Режим доступа: http://read.pudn.com/downloads153/ebook/668834/melp.pdf. 9. ITU-T Recommendation G.726 —.40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM). – Geneva, 1990. 10. Сэломон Д. Сжатие данных, изображения и звука / Сэломон Д. [перевод с английского В.В. Чепыжова]. – М.: Техносфера, 2004. – 367с. – (Мир программирования цифровая обработка сигналов). 11. G. Nigel N. Martin, Range encoding: An algorithm for removing redundancy from a digitized message, Video & Data Recording Conference, Southampton, UK, July 24–27, 1979. 12. "On the Overhead of Range Coders", Timothy B. Terriberry, Technical Note 2008. 13. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. / Ватолин Д., Ратушняк А., Смирнов М., Юкин В. – М.: ДИАЛОГ-МИФИ, 2003. – 384с. 14. Сергиенко А. Б. Цифровая обработка сигналов / А. Б. Сергиенко — СПб.: Питер, 2002. — 608 с: ил.

Поступила в редакцію 08.12.2013р.

Рекомендували до друку Оргкомітет 4-ої н/п конференції студентів і молодих учених «Методи та засоби неруйнівного контролю промислового обладнання» (26-27.11.2013р., ІФНТУНГ) та докт. техн. наук, доц. Лютак І. З.