

СПОСОБ СЖАТИЯ РЕЧЕВОГО СИГНАЛА В СИСТЕМАХ ИДЕНТИФИКАЦИИ ИСТОЧНИКОВ ЗВУКА

© 2018 Д. М. Жуманиязов, В. Н. Кострова, Н. А. Коростелева

*Воронежский институт высоких технологий (г. Воронеж, Россия)
Воронежский государственный технический университет (г. Воронеж, Россия)
Открытое акционерное общество «Воронежский завод
полупроводниковых приборов-сборка» (г. Воронеж, Россия)*

В данной работе представлена методика повышения эффективности алгоритма сжатия речевых сигналов, без большой потери качества. Сжатие в этом случае означает удаление из цифрового сигнала тех шагов квантования, которые можно спрогнозировать. Авторы предлагают уменьшить число этих шагов квантования, используя модифицированный алгоритм линейной предикции с переменным порядком. Это позволяет сократить время сжатия.

Ключевые слова: речевой сигнал, идентификация, информация.

Задача эффективного представления речевого сигнала является одной из важнейших задач в идентификации говорящего. Например, на сервере LAN или WAN может быть установлена система автоматического распознавания динамиков, которая разрешает терминалу доступ к сети в соответствии с голосом абонента.¹

В этом случае существует два способа обработки информации:

1) получить идентификационные признаки говорящего от речевого сигнала на абонентском терминале и передать их на сервер для принятия решения о возможности допуска;

2) сжать речевой сигнал, не теряя при этом информации о личности говорящего, в виде WAV-файла пароля, и передать его по сети на сервер, где проводится процедура идентификации.

Одним из преимуществ первого подхода является сокращение времени передачи по сети. Его основные недостатки заключаются в том, что он снижает конфиденциальность процедур идентификации говорящего, а также возникает необходимость установки на терминалах системы первичного анализа и описания характеристик сигналов говорящего.

Таким образом, второй подход более эффективен для обработки информации относительно количества вычислений, необходимых для сжатия, и использования ASP-технологий для выбора информативных признаков и принятия решений.

Анализ известных работ в соответствии с хорошо известными методами сжатия сигнала и с учетом статистических характеристик речевого сигнала, параметры аналого-цифровых преобразователей (АЦП) выбирается в соответствии с правилами, представленными в [1, 2]: дискретизации по частоте определяется верхний предел частоты сигнала, определение диапазона рассеивания, количества шагов за сигнал к шумовому отношению и требуемой точности.

Поскольку речевой сигнал не является стационарным, параметры АЦП подбираются приблизительно с использованием наиболее катастрофической ситуации, с которой редко приходится сталкиваться.

В результате, дублирование речевого сигнала завершено дублированием дискретного преобразования. При этом возникает новая проблема: устранение избыточности АЦП. В многочисленных вариантах импульсной модуляции и адаптивное кодирование, которое используется сегодня для устранения избыточности кодирования,

Частота дискретизации остается постоянной и равна частоте Найквиста, а избыточность устраняется путем анализа значений соседних образцов сигналов.

Целью исследования является повышение эффективности алгоритма сжатия речевого сигнала без потери информации, связанной с личностными особенностями гово-

Жуманиязов Дмитрий Маратович – Воронежский институт высоких технологий, студент, tyor45zhumm@yandex.ru.

Кострова Вера Николаевна – Воронежский государственный технический университет, профессор, kost_3_veranikol@yandex.ru.

Коростелева Наталья Александровна – Открытое акционерное общество «Воронежский завод полупроводниковых приборов-сборка», сотрудник, koerolostor99_34@yandex.ru.

рящего, путем удаления тех выборок, которые могут быть предсказаны.

В данной работе предлагается уменьшить число выборок сигналов с помощью модифицированного метода линейного прогнозирования переменного порядка.

Особенность предлагаемого метода заключается в двухэтапной обработке речевого сигнала, что позволяет сократить время, необходимое для сжатия wav-файла.

Процесс осуществляется в два этапа:

1. Предварительное сжатие.
2. Окончательное сжатие.

На первом этапе wav-файл обрабатывается по оригинальной методике, заключающейся в аппроксимации речевого сигнала полилинией, с возможностью установить степень его отклонения от исходного сигнала.

На втором этапе области wav-файла, которые не были затронуты во время начальной процедуры сжатия, аппроксимируются с помощью полинома, порядок которого определяется в соответствии с точностью, необходимой для восстановления речевого сигнала из архивного файла.

Поскольку речевой сигнал является непрерывной функцией $S(t)$, спектр которой ограничен верхней частотой F , он определяется последовательностью его выборок, временной интервал которых рассчитывается по следующей формуле:

$$T_k = \Delta t = \frac{1}{2F}. \quad (1)$$

Таким образом, сигнал $S(t)$ можно описать по следующей формуле:

$$\begin{aligned} S(t_1) &= S_0 + k_1(t - t_0) \text{ for } t_0 \leq t \leq t_1, \\ S(t_1) &= S_0 + k_1(t - t_0) + k_2(t - t_1) \text{ for } t_1 \leq t \leq t_2, \\ S(t_1) &= S_0 + k_1(t - t_0) + k_2(t - t_1) + k_3(t - t_2), \\ &\text{for } t_2 \leq t \leq t_3, \end{aligned} \quad (5)$$

где k_i можно определить следующим образом:

$$\begin{aligned} k_1 &= \operatorname{tg} \gamma_1 = \frac{S_1 - S_0}{t_1 - t_0}, \\ k_2 &= \frac{S_2 - S_1}{t_2 - t_1} \pm \frac{S_1 - S_0}{t_1 - t_0}, \\ k_3 &= \frac{S_3 - S_2}{t_3 - t_2} \pm \frac{S_2 - S_1}{t_2 - t_1} \pm \frac{S_1 - S_0}{t_1 - t_0}, \end{aligned} \quad (6)$$

В общем:

$$S(t) = \sum_{i=-\infty}^{\infty} S(t_i) \cdot \varphi_i(t) = \sum_{i=-\infty}^{\infty} S(i \cdot \Delta t) \cdot \varphi_i(t), \quad (2)$$

$$\text{здесь } \varphi_i(t) = \frac{\sin 2\pi F(t - i \cdot \Delta t)}{2\pi F(t - i \cdot \Delta t)} - \text{ пример}$$

функции дискретного значения.

$$\varphi_i(t) = \begin{cases} 1; t = i \cdot \Delta t, \\ 0; t = k \cdot \Delta t, k \neq i. \end{cases} \quad (3)$$

В течение ограниченного периода времени τ речевого сигнала число отсчетов сигнала N определяется выражением:

$$N = \frac{\tau}{\Delta t} = 2\tau F. \quad (4)$$

С учетом стационарности квазисигнала, а также некритического состояния систем сбора данных к реальному времени обработки разработан метод снижения кодирующей избыточности речевого сигнала с помощью АЦП.

Минимизация ошибки восстановленного сигнала заключается в нахождении тех фиксированных значений аргумента $t_0, t_1, t_2, \dots, t_n$, которые обеспечивают слияние разбитого участка от вершин $S_0, S_1, S_2, \dots, S_n$ к функции $S(t)$ таким образом, чтобы для всего диапазона аргументов изменение абсолютной погрешности не превышало допустимых значений.

Функции в этих точках могут быть представлены следующим образом:

$$S(t) = S_0 + \sum_{j=0}^n (t - t_j) \cdot \psi [\text{sign}(t - t_j)] \quad (7)$$

где $\Psi \begin{cases} 1; (t - t_j) > 0, \\ 0; (t - t_j) < 0. \end{cases}$

Погрешность аппроксимации определяется оставшимся членом интерполяционной формулы.

$$S(t) = S(t_j) + \frac{(t - t_j) \cdot |S(t_{j+1}) - S(t_j)|}{(t_{j+1}) + (t_j)}, \quad (8)$$

и оставшийся член расширения функций с тем же интервалом будет выглядеть следующим образом:

$$|R(t)|_{\max} = \frac{|S^n(t)_{\max}|}{2!} \left[\frac{t_{j+1} - t_j}{2} \right]^2 \quad (9)$$

Получение $\Delta S_{\max} = |R(t)|$, данной формулы для интервала выборки:

$$\Delta t^* = t_{j+1} - t_j = \sqrt{\frac{8 \cdot \Delta S_{\max}}{|S^n(t)_{\max}|}} \quad (10)$$

Задавая верхнюю частоту полосы сигнала, мы можем определить отклонение реального значения сигнала от предсказанного.

На основе вышеизложенного создан алгоритм реализации процедуры предварительного сжатия голосовой информации.

Она включает в себя следующие шаги:

1. Установление уровня допустимой абсолютной погрешности сигнала восстановления;

2. Установление минимального размера М при сжатии буфера;

3. Определение коэффициента прогнозирования;

4. В случае, если будет происходить отклонение коэффициента $\Delta k \leq \Delta$, то включение данного образца при сжатии буфера, необходимо будет увеличить производительность m буферного счетчика на 1 и перейти к пункту 3.

Если неравенство не выполняется, то проверяем счетчик буфера m : Если $m < M$ – установите $0 = m$ и перейдите к пункту 3; Если $m \geq M$ после сжатия все поле;

5. Если конец wav-файла не найден, то переходим к пункту 3. Линейное определе-

В этом случае отрезок линии в пределах интервала времени $[t_j + t_{j+1}]$ определяется выражением:

ние используется для реализации процесса второй степени сжатия.

Сигнал $S(t)$ представлен в цифровой форме S_n , $n = 1, 2, \dots, N$, где N -число образцов сигнала, который получен путем выборки на определенной частоте F . Этот сигнал S_n , $n = 1, 2, \dots, N$, может быть представлен как линейная комбинация предыдущих значений сигнала и влияние u_n .

$$S_n = - \sum_{k=1}^p a_k \cdot S_{n-k} + G \cdot u_n, \quad (11)$$

где G -коэффициент усиления, а p -порядок предсказания. Тогда, зная значения сигнала S_n , задача сводится к поиску коэффициентов a_k .

Что касается оценки, мы будем использовать метод наименьших квадратов, предполагающий сигнал S_n как детерминированный.

Значения сигнала S_n будут выражены через его оценивающие значения \tilde{S}_n по следующей формуле:

$$\tilde{S}_n = - \sum_{k=1}^p a_k \cdot S_{n-k} \quad (12)$$

Тогда ошибку прогноза можно описать следующим образом:

$$e_n = S_n - \tilde{S}_n = S_n - \sum_{k=1}^p a_k \cdot S_{n-k} \quad (13)$$

Используя метод наименьших квадратов, параметры k а выбираются таким образом, чтобы минимизировать среднее или сумму квадратов ошибки прогноза. Чтобы найти коэффициенты k а, воспользуемся

матричным методом, называемым методом Дарбина.

Расчет коэффициентов линейного прогноза и погрешности прогноза осуществляется по следующему алгоритму: из коэффициентов линейного прогноза погрешность прогноза составляет:

1. Сегментация речевого сигнала на стационарных интервалах;

2. Для разделенных интервалов формируется система линейных уравнений, которая решается матричным методом или методом Дарбина с использованием автокор-

реляционной функции (метод выбирается пользователем);

3. Вычисляется ошибка прогноза.

Разработанные алгоритмы реализованы в среде MATLAB. Разработано программное обеспечение, позволяющее реализовать процесс сжатия речевого сигнала с помощью предложенного метода.

В качестве примера было проведено сжатие wav-файлов (Формат PCM, F = 8 кГц), содержащих входное слово «приводной» размером 6,54 Кбит.

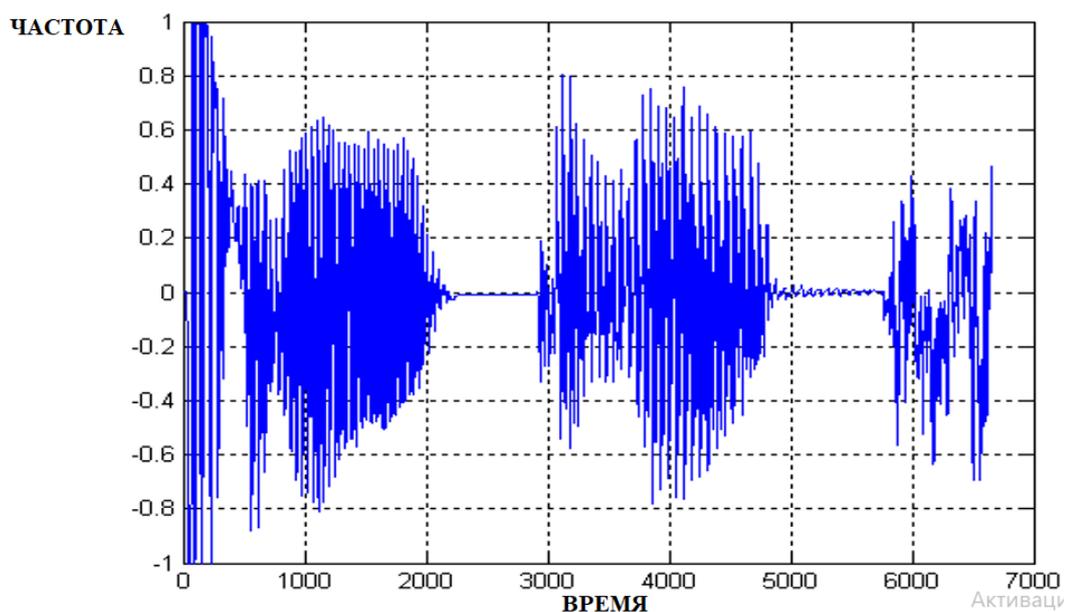


Рисунок 1. Wav-файл для украинского слова «приводной».

Размер полученного архивного файла в формате mat имеет Габариты 3,51 Кбит, коэффициент сжатия 54 %.

Для иллюстрации эффективности разработанного метода проведено исследование, в котором проведено сравнение размеров архивных файлов из различных программных продуктов.

Исходные файлы речевого сигнала хранились в wavformat. Сжатие исходного файла осуществлялось архиватором WinZip 8.0, конвергенцией в форматы mp3, wma и разработанным программным обеспечением. Результаты экспериментов приведены в таблице 1.

Таблица 1

Сравнение эффективности методов сжатия сигнала

Лингвистический материал	Размер файлов, сжатых разными способами				
	Размер файла в формате .wav, Kbit	Размер файла в формате .zip, Kbit	Размер файла в формате .mp3, Kbit	Размер файла в формате .wma, Kbit	Размер файла в формате .mt, Kbit
Числа 1-10	78,20	34,10	27,40	28,10	30,54
	5,72	3,78	3,12	3,31	3,19
	12,00	7,42	5,28	5,40	5,52
	6,54	4,77	3,51	3,44	3,51
	3,90	21,00	17,20	17,30	17,39

Приведенные в таблице 1 результаты показывают, что разработанное программное обеспечение выполнило сжатие более эффективно, чем WinZip 8.0 archiver и WMA codec.

Эксперимент реализован с целью выявления преимуществ и недостатков использования предыдущей процедуры сжатия. В качестве критериев оценки были указаны Размер оригинального мат-файла и время, затраченное на обработку лингвистического материала.

Результаты эксперимента приведены в таблице 2.

Из таблицы 2 видно, что использование предыдущей процедуры сжатия позволяет увеличить скорость обработки речевого сигнала на 25-35 % при увеличении размера выходного файла примерно на 10 %.

Таким образом, целесообразно поставить в программное обеспечение выбор режима сжатия.

Таблица 2

Изучение эффективности предыдущей процедуры сжатия

Лингвистический материал	Длина wav-файла	Размер .Mat файл, со сжатием до, Кбит	Размер .mat файл, без использования предварительного сжатия Kbit	Разница в размерах, %	Время, за которое файл был заархивирован	Время, за которое файл был заархивирован, без использования процедур предварительного сжатия, сек	Разница во времени, %
Числа 1-10	80128	30,54	28,53	7	2,7853	45,2337	34
доклад	5816	3,19	2,97	7	0,5542	0,8479	35
Эксперимент	12320	5,52	5,14	7	0,8013	1,0737	25
Приклад	6656	3,51	3,07	13	0,7109	0,9313	24
VNT U	34752	17,39	17,22	1	1,8130	1,5497	24

Вывод. Полученные теоретические и экспериментальные результаты позволили разработать методику сжатия речевого сигнала с использованием линейного прогнозирования. Разработан алгоритм и программное обеспечение для реализации этого метода с использованием MATLAB. В исследовании мы получили:

1. Размер сжатых по разработанному методу файлов соответствует размеру файлов в форматах mp3, wma, полученных при конвертации исходного wav-файла, и улучшает их качество;

2. Использование предыдущей процедуры сжатия позволяет увеличить скорость обработки речевого сигнала на 25-35 % при увеличении размера выходного файла примерно на 10 %.

3. Изучение звуков, соответствующих русскому алфавиту, показало, что наиболее чувствительными к сжатию предыдущих процедур были звуки, соответствующие гласным и озвученным согласным (коэффициент сжатия 50-80 %), наименее чувстви-

тельными – согласные (коэффициент сжатия 10-40 %).

ЛИТЕРАТУРА

1. Головинов С. О. Цифровая обработка сигналов / С. О. Головинов [и др.] // Вестник Воронежского института высоких технологий. – 2009. – № 4. – С. 064-065.

2. Зазулин А. В. Особенности построения семантических моделей предметной области / А. В. Зазулин, Ю. П. Преображенский // Вестник Воронежского института высоких технологий. – 2008. – № 3. – С. 026-028.

3. Зяблов Е. Л. Разработка лингвистических средств интеллектуальной поддержки на основе имитационно-семантического моделирования / Е. Л. Зяблов, Ю. П. Преображенский // Вестник Воронежского института высоких технологий. – 2009. – № 5. – С. 024-026.

4. Кленяева Г. В. Современные проблемы речевой акустики и построения систем автоматического распознавания речи /

Г. В. Кленяева, А. П. Преображенский // Вестник Воронежского института высоких технологий. – 2007. – Т. 1. – № 2-1. – С. 071-074.

5. Lvovich I. Ya. Analysis of potential of error-correcting capabilities of codes / I. Ya. Lvovich, A. P. Preobrazhensky, O. N. Choporov // Life Science Journal. – 2013. – Т. 10. – № 4. – С. 830-834.

6. Методы и стандарты кодирования и сжатия речи в цифровой телефонии: Интернет-публикации. – Режим доступа: <http://dox.sbnnet.ru:8082/~serge/speech.ru/coder/>.

7. Преображенский А. П. Анализ методов кодирования разных видов информации

/ А. П. Преображенский // Вестник Воронежского института высоких технологий. – 2013. – № 10. – С. 74-77.

8. Рабинер Л. Р. Цифровая обработка речевых сигналов: пер. с англ. / Л. Р. Рабинер, Р. В. Шафер; под ред. М. В. Назарова и Ю. Н. Прохорова. – М.: Радио и связь, 1981. – 496 с.

9. Смит С. В. Научно-техническое руководство по цифровой обработке сигналов / С. В. Смит; пер. с англ. В. Н. Покровского, В. И. Силантьева. – СПб.: АВТЭКС, 2001. – 630 с.

METHOD OF COMPRESSION OF SPEECH SIGNAL IN SYSTEMS IDENTIFICATION OF SOUND SOURCES

© 2018 D. M. Zhumaniyazov, V. N. Kostrova, N. A. Korosteleva

*Voronezh Institute of high technologies (Voronezh, Russia)
Voronezh state technical University (Voronezh, Russia)
Open Joint-Stock Company «Voronezh Plant of Semiconductor
Devices-Assembly» (Voronezh, Russia)*

This paper presents a technique to improve the efficiency of the algorithm of compression of speech signals, without much loss of speech signals. Compression in this case means removing from the digital signal those quantization steps that can be predicted. We propose to reduce the number of these quantization steps using a modified variable order linear prediction algorithm. This allows to reduce the time compression.

Key words: speech signal, identification, information.