

АНАЛИЗ СОВРЕМЕННОЙ ТЕХНИКИ СЖАТИЯ РЕЧЕВОГО СИГНАЛА С ИСПОЛЬЗОВАНИЕМ ВЕЙВЛЕТОВ

ANALYSIS MODERN TECHNIQUE FOR SPEECH SIGNAL COMPRESSION USING WAVELET

Hein Min Zaw

Summary. The fundamental purpose of speech is communication, i.e., the transmission of messages. Speech compression is the digital signal which is compressed by using various techniques for transmission. Compression is nothing but high input stream of data converted into smaller size. Compression is done for all, such as image, data; signals. Speech compression plays a prominent role in speech signal processing such as satellite communications, internet communications, transmission of biomedical signals and other applications. In this paper speech is compressed by discrete wavelet transform technique. Here speech compression technique is used and done using DWT.

Keywords: Speech Compression, Wavelet, discrete wavelet transform (DWT).

Hein Min Zaw

Аспирант, Курский Государственный Университет
heinminzaw13@gmail.com

Аннотация. Основной целью речи является сообщение, точнее — передача сообщений. Сжатие речи — это цифровой сигнал, который сжимается с использованием различных техник. Сжатие — это процесс преобразования потока ввода данных в меньший размер. Сжатие выполняется для всех видов информации, таких как изображение, речевые сигналы и другие данные. Сжатие речи играет важную роль в обработке речевых сигналов, например, для спутниковой связи, интернет-связи, передачи биомедицинских сигналов и др. Проблема сжатия речи в настоящее время является одной из наиболее актуальных из-за активного внедрения технологий на основе речевого управления в различные сферы человеческой деятельности. В этой статье мы представим процесс сжатия речевых сигналов методом дискретного вейвлет-преобразования.

Ключевые слова: Сжатие речи, вейвлет, дискретное вейвлет-преобразование (ДВП).

Введение

Основной целью ранних компьютеров являются вычисления. Однако с появлением разного рода достижений в области технологий компьютеры стали разрабатываться не только для выполнения вычислительных операций, но и для более широкого круга сфер их применения. Одной из них является представление и хранение нечисловой информации в формате двоичного кода. Представленные таким образом данные требуют большего пространства для хранения и более длительного времени доступа. Еще один вариант — передача данных. Время передачи будет больше, поскольку данные, которые будут отправлены, также имеют больший размер. В связи с этими недостатками представления данных в виде двоичного кода возникает необходимость в их сжатии.

Речь играет важную роль в мультимедийной системе. Речь — это очень простой способ для людей передавать информацию друг другу. С пропускной способностью только 4 кГц речь может передавать информацию с эмоциями человеческого голоса. Сжатие речи — это техника преобразования человеческой речи в эффективно кодированный формат, который впоследствии может быть декодирован для получения близкого приближения исходного сигнала. Рост потребностей в мультимедийной связи требует эффективного использования передачи

информации большого объема. Чтобы передавать речевую информацию большого объема, необходимо сжать данные. Сжатие — это процесс преобразования потока входных данных (исходного потока или необработанных данных) в другой поток данных (называемый выходные или сжатые данные), который имеет меньший размер.

Методы сжатия речи

Сжатие сигналов основано на удалении избыточного пространства между соседними образцами и / или между соседними циклами. В сжатии данных желательно представлять данные в виде как можно меньшего количества коэффициентов при незначительной потере качества сигнала. Техники сжатия речевого сигнала можно классифицировать на две основные категории: без потерь и с потерями.

Методы сжатия можно разделить на три функциональные категории:

- ◆ Прямые методы. Образцы сигнала непосредственно обрабатываются для обеспечения сжатия.
- ◆ Методы трансформации: такие как преобразование Фурье (ПФ), дискретное вейвлет-преобразование (ДВП) и дискретное косинусное преобразование (ДКП).
- ◆ Методы извлечения параметров: они применяются для извлечения некоторых особенностей

сигнала, которые позже используются для его восстановления.

Сжатие речи с использованием дискретных вейвлет-преобразований

Вейвлет-преобразование является последним мощным математическим инструментом во многих областях науки и техники, тем более в области обработки сигналов. Вейвлет-преобразование способно разложить сигнал на набор основных функций. Вейвлет является одним из недавних открытий на фоне использовавшегося ранее преобразования речевых сигналов при помощи Фурье-анализа. Качественно новый метод обработки и анализа речевых сигналов на основе вейвлет-преобразования позволяет рассматривать сигнал одновременно с позиции его временных и частотных характеристик. Использование вейвлетов стало революционным событием в методологии анализа сигналов, так как вейвлет-анализ может применяться для обработки не только линейных, но и нестационарных сигналов. Речь представляет собой нестационарный речевой сигнал из-за меняющейся во времени природы и резких изменений частот человеческой голоса. Нестационарные сигналы — это сигналы, характеризующиеся переходными изменениями их частот. Важная особенность локализации вейвлетов, наряду с их свойствами разрешения во временной частоте делают их хорошо подходящими для кодирования речевых сигналов.

Вейвлет-сжатие представляет собой форму прогностического сжатия, где количество шума в наборе данных может быть оценено относительно прогностической функции. Большинство современных методов сжатия используют двухэтапный процесс: во-первых, применяется функция прогнозирующего сжатия (например, вейвлет-преобразование). Если выбор функции прогностического сжатия хорош, результатом будет новый набор данных с меньшими значениями и большим количеством повторений. Во-вторых, шаг сжатия кодирования, который будет представлять набор данных в его минимальной форме (кодирование Хаффмана, длина выполнения). Сжатие речевых сигналов имеет широкий спектр практического применения. Одним из примеров является цифровая сотовая технология, в которой многие пользователи делят одну и ту же полосу частот. Сжатие позволяет большему количеству пользователей использовать эту систему в одно и то же время.

Описанные выше функции называются вейвлетами. Важными и необходимыми свойствами вейвлетов являются допустимость и условия регулярности, и это те свойства, которые дали вейвлетам их имя. Рассмотрим существенно значимые свойства функций непрерывного вейвлета $\psi(t)$;

Свойство 1: Условие допустимости 1:

$$\int_{-\infty}^{\infty} |\psi(t)|^2 \cdot d(t) < \infty.$$

Свойство 2: Условие допустимости 2:

$$\int_{-\infty}^{\infty} \psi(t) \cdot d(t) = 0.$$

Свойство 3: Условие допустимости 3:

$$|\psi|^2|_{\omega=0}.$$

Существует бесконечное множество функций, которые удовлетворяют указанным выше свойствам, среди которых вейвлет Хаара является простейшей из них.

Дискретное вейвлет-преобразование (ДВП) носит определение «малой волны», которая обладает своей энергией, сконцентрированной во времени, служащей инструментом анализа переходных, нестационарных или изменяющихся во времени сигналов. Оно обладает колебательными, волновыми свойствами, но также позволяет провести одновременный анализ времени и частоты сигнала. ДВП является важным инструментом для анализа речевого сигнала, который является нестационарным. ДВП состоит из коэффициентов разложения исходного сигнала в отношении базиса $\psi(t)$, являющегося расширенным вариантом функции ψ , называемой материнским вейвлетом.

Выражение для ДВП выглядит таким образом:

$$\psi_{j,k}(t) = \frac{1}{s_0^j} \psi\left(\frac{t - k\tau_0 s_0^j}{s_0^j}\right), j, k \in Z$$

Так выглядит разложение сигнала на аппроксимацию и подробные коэффициенты с использованием каскадных фильтров, состоящих из фильтров низких и высоких частот. Выход фильтра нижних частот обеспечивает коэффициенты аппроксимации, а фильтр верхних частот дает подробные коэффициенты.

Частота является одним из важнейших параметров характеристики речевого сигнала. Рассмотрим в качестве примера человеческий голос. Для многих сигналов низкочастотное содержимое является наиболее важной частью, которая отвечает за идентичность сигнала. Высокочастотный контент передает детали сигнала. Рассмотрим

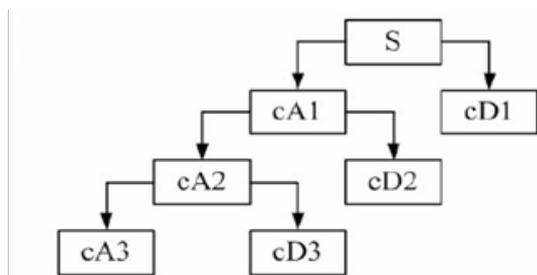


Рис. 1

речевой сигнал, итак, если удалить высокочастотные компоненты, голос звучит по-другому, но можно догадаться, что человек говорит. Однако, если удалить некоторое число низкочастотных компонентов, мы услышим тарбарщину. Исходный сигнал S проходит через два фильтра и представляется в виде двух сигналов (A и D).

Из рис. 1 мы можем определить вейвлет-разложение как разложение исходного сигнала S на разные полосы частот, и это разложение проводится путем выбора соответствующей вейвлет-функции с требуемыми моментами исчезновения и уровнем разложения.

Вейвлеты полезны для этого метода сжатия звука, потому как вейвлеты с большим количеством исчезающих моментов могут представить точную спецификацию полос пропускания каждого поддиапазона в разложении и компактности вейвлета представляемого сигнала. В выборе оптимальной вейвлет-функции для достижения наилучших результатов можно также использовать несколько разных критериев. Идея вейвлет-преобразований заключается в том, что сдвиг должен допускать только изменения во времени, но не формы. Изменение временной функции приводит к соответствующим изменениям в частотной функции.

Принцип вейвлет-преобразования заключается в разложении сигнала на различные полосы частот. Этот процесс осуществляется благодаря выбору наиболее подходящей вейвлет-функции и процессу вычисления дискретного вейвлет-преобразования (ДВП). Сжатие сигнала основано на концепции, согласно которой небольшое число коэффициентов аппроксимации (на выбранном уровне) и некоторые из подробных коэффициентов могут точно представлять регулярные компоненты сигнала. Выбор уровня разложения для ДВП обычно зависит от типа анализируемого сигнала или некоторых других критерия, например, энтропии. Для обработки речевых сигналов декомпозиция до шкалы 5 является наиболее приемлемой.

Идея сжатия речевого сигнала с использованием вейвлет-преобразования связана с относительной скуд-



Рис. 2. Структура системы сжатия

ностью вейвлет-домена представления конкретного сигнала. Вейвлеты трансформируют речевую информацию в виде нескольких соседних коэффициентов. Поэтому в результате вейвлет-коэффициенты имеют менее 5% от максимального значения.

Это значит, что основная часть энергии речевого сигнала представлена высокоценными коэффициентами, число которых невелико. Поэтому малозначные коэффициенты представляется возможным усечь или обнулить, а затем использовать в процессе восстановления сигнала. Эта схема сжатия обеспечивала сегментное отношение сигнал-шум (SEGSNR) 20 дБ с коэффициентом, равным 10%.

Для расчета пороговых значений используются два разных подхода.

- ◆ Первый, носящий название глобального порога (Global Thresholding), содержит вейвлет-расширение сигналов с последующим сохранением наибольших коэффициентов, имеющих абсолютную величину. При этом пользователь может вручную устанавливать глобальный порог или другие параметры, например, относительную квадратную норму производительности восстановления или производительность сжатия. Следует отметить, что необходимо выбрать только один параметр.
- ◆ Второй подход, названный пороговым значением уровня (By Level Thresholding), представляет собой применение визуально определенных уровней порогов по отношению к каждому из уровней разложения в процессе вейвлет-преобразовании.

Сжатие сигнала осуществляется благодаря первому усечению малозначных коэффициентов. Данный подход к сжатию представляет собой процесс кодирования последовательных нулевых коэффициентов, представленных двумя байтами. Один байт показывает последовательность нулей в векторе вейвлет-преобразования, а второй байт отражает количество нулей, расположенных последовательно.

Заключение

Проблема сжатия речевого сигнала является крайне актуальной в настоящее время, так как наблюдается увеличение роли мультимедийной связи в различных областях современной науки и техники. Дискретное вейвлет-преобразование, как было отмечено выше, довольно успешно может применяться для сжатия речевых сигналов. Сжатие речи осуществляется с помощью

пренебрежения малыми и меньшими значимыми коэффициентами и данными и отбрасыванием их. Затем речевой сигнал подвергается процессу квантования и кодирования. Именно благодаря этому информация передается практически без потерь и искажений, что особенно важно в тех областях, где требуется максимальная точность: медицина, судебная экспертиза, криминалистика и другие сферы применения технологии анализа речевых сигналов.

ЛИТЕРАТУРА

1. W. Kinsner and A. Langi»speech and image signal compression with wavelets» W. Kinsner and A. Langi.
2. Amandeep Kour, Vimal kishore Yadav, Vikas maheshwari, Deepak Prashar, «A Review on Image Processing» International Journal of Electronics Communication and Computer Engineering, vol. 4, Issue 1 ISN(online): 2249–071X, ISSN(print):2278–4209, 2013.
3. Vimal Kishore Yadav, Pratiksha Singh Gaur, Sucheta Yadav,»Impending Form Interpretations for Delay to Ramp and Step Input On-Chip VLSI RLC Annex «, «International Journal of Advanced Research in Computer and Communication Engineering(IJARCC)», vol. 3, Issue 11, ISSN (Online): 2278–1021, ISSN (Print): 2319–5940,2014.
4. Fgee, E.B., Phillips, W.J. and Robertson, W. «Comparing Audio Compression using Wavelets with other Audio Compression Schemes», IEEE Canadian Conference on Electrical and Computer Engineering, IEEE, Edmonton, Canada, 1999, pp. 698–701.
5. Agbinya, J.I. «Discrete Wavelet Transform Techniques in Speech Processing», IEEE Tencon Digital Signal Processing Applications Proceedings, IEEE, New York, NY, 1996, pp 514–519.
6. Kinsner, W. and Langi, A. « Speech and Image Signal Compression with Wavelets», IEEE Wescanex Conference Proceedings, IEEE, New York, NY, 1993, pp. 368–375.

© Хеин Мин Зо (heinminnzaw13@gmail.com).

Журнал «Современная наука: актуальные проблемы теории и практики»



Курский Государственный Университет