

СИСТЕМА СЖАТИЯ РЕЧЕВЫХ СИГНАЛОВ С ПОМОЩЬЮ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ

Изюмов В.В., группа ТКС-06м

Руководитель доц. каф. АТ Дегтяренко И.В.

В настоящее время наблюдается активное развитие и внедрение новых средств связи и телекоммуникаций, в частности, современных цифровых телефонных сетей и соответствующего абонентского оборудования, а также развитие компьютерной телефонии и спутниковых средств связи. Использование цифровых методов представления, обработки и передачи приводит к многократному увеличению занимаемой полосы частот и, как следствие этого, к многократному увеличению скорости передачи информационных сообщений. Кроме того, информация в подавляющем большинстве случаев носит частный, конфиденциальный характер, поэтому все чаще к передающей аппаратуре и каналам связи предъявляется требование обеспечения защиты передаваемой информации от несанкционированного доступа.

Решение проблемы лежит в области разработки эффективных методов цифрового преобразования — сжатия (кодирования) различных сообщений, являющихся, как правило, нестационарными случайными процессами. Поскольку речь является наиболее распространенным аналоговым информационным сообщением, вопросы сжатия (цифрового кодирования) речи являются наиболее актуальными и традиционными. На данный момент имеется множество различных решений сжатия речевого сигнала, которые можно разбить на три категории.

К первой категории относятся так называемые кодеры формы. Их принцип основан на точной передаче формы сигнала. Скорость кода для кодеров формы

лежит в пределах 64–32 Кбит/с. Дальнейшее уменьшение скорости кода приводит к уменьшению качества речи.

Ко второй категории относятся вокодеры. Суть которых кодировать не форму сигнала, а сами параметры речевого сигнала. Вокодеры позволяют очень сильно сократить скорость кода. При обеспечении разборчивой речи, скорость кода в вокодерах достигает 4,8 Кбит/с. Вокодеры используются в основном только там, где главным образом необходимы разборчивость и высокая степень сжатия: в военной связи, авиации, космической связи.

К третьей категории относятся гомоморфные вокодеры, которые используют принципы кодеров формы и вокодеров. Добиваясь компромисса между разборчивостью речи и скоростью кода.

В настоящее время продолжают работы по разработке новых алгоритмов или модификации существующих, позволяющих улучшить качество восстанавливаемого сигнала без увеличения битовой скорости, или увеличить степень сжатия информации при сохранении приемлемого качества сигнала.

На основе данного обзора можно сделать вывод о том, что несмотря на существование стандартных алгоритмов кодирования речи, у разработчиков и научных работников есть огромный простор для деятельности, направленной на дальнейшее совершенствование технологии сжатия данных.

Одним из направлений развития технологий сжатия данных может быть применение нейросетевых технологий. Нейронные сети, как метод, обладают рядом преимуществ, такими как:

Способность к обучению. Могут формировать зависимость между любыми образцами входов и выходов.

Обобщение. Выделение основной сути зависимости «вход–выход», и делать вывод о зависимостях, которых она не знает.

Нелинейность. Сеть может нелинейные, непараметрические функции от входа. Это преимущество, так как речь — очень нелинейный процесс.

Помехоустойчивость. Сеть устойчива к физическим повреждениям и зашумленным входам.

Однородность. Сети предлагают однородную вычислительную парадигму, которая может интегрировать связи от различных типов входов.

Параллелизм. Сети параллельны по природе. Так что они хорошо решаются с помощью многопроцессорных систем или нейроконтролеров, что, в конечном счете, позволит очень быстро обрабатывать речь или другие данные.

Идея системы сжатия заключается в том, чтобы из входного сигнала выделять речь и определять ее структуру — фонемы, которые в последствии передавать в канал связи. А на приемной стороне по полученной информации синтезировать речевой сигнал.

Модель кодера создавалась средствами среды MATLAB. Так как реализация кодера с помощью нейронной сети (рис. 1) для полного набора фонем представляет некоторую сложность, то было принято упрощение, заключающееся в том, чтобы построить сеть, распознающую 6 гласных фонем (а, э, и, ы, у) и 2 «фонемы», обозначающих паузу, и неопознанные фонемы. Также нейронная сеть тренируется только на одного диктора.

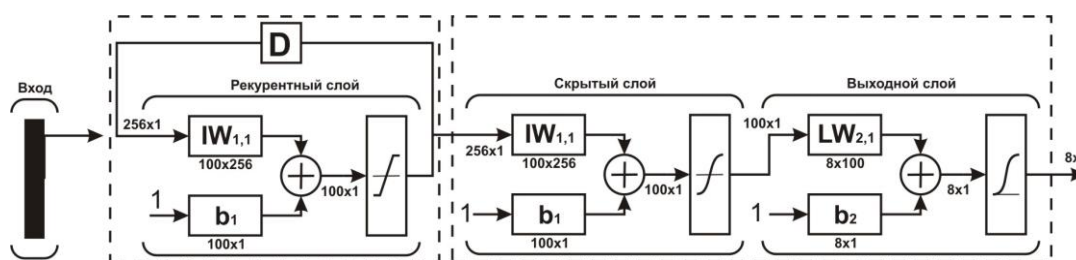


Рисунок 1 — Структур нейронной сети

Работа модели заключается в том, что поток оцифрованной речи с частотой дискретизации 22050 Гц разбивается на кадры длиной 256 отсчетов и далее подвергается вейвлет преобразованию, получая матрицу 16x16 коэффициентов, которые потом преобразуются во входной вектор сети. Сеть состоит из двух подсетей. Первая — рекуррентная сеть Хопфилда, функция

которой заключается в том, чтобы входной вектор ассоциировать с запомненными шаблонами, подсеть также обладает свойством подавления помех. Далее двухслойная подсеть классифицирует «очищенный» вектор и вырабатывает выходной вектор, каждый элемент которого соответствует определенной фонеме. То есть элемент выходного вектора ближайший к единице является предполагаемой фонемой содержащийся в кадре.

Сеть обучается стандартным методом градиентного спуска. Примеры шаблонов обучения представлены на рис. 2.

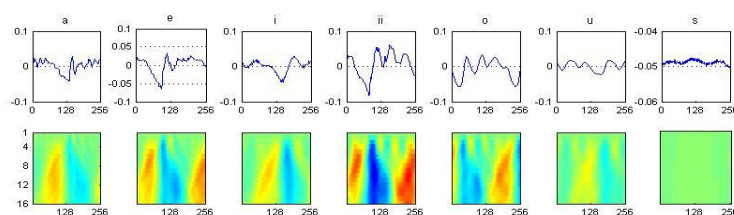


Рисунок 2 — Примеры обучающих образцов (внизу вейвлет коэффициенты подаваемые на вход)

Сеть правильно распознает 58–64 образца из 80 предъявленных, в которые входят 40 обучающих образцов. Сеть точно определяет 40 обучающих образцов и на 18–24 образца дает правильное предположение. То есть сеть дает примерно 75% верных ответов.

Данная структура нейронной сети имеет хорошую верность результата, а также проявляет свойства подавления зашумленности в предъявляемых векторах. Для повышения полученных результатов предлагается структуру и параметры системы оптимизировать с помощью метода генетических алгоритмов. С помощью ГА можно оптимизировать параметры вейвлет преобразования (количество коэффициентов и тип вейвлета), параметры рекуррентного слоя (количество шагов восстановления входного вектора), количество нейронов в скрытом слое и тип активационной функции.