

ТЕОРЕТИЧЕСКИЕ ПОДХОДЫ К ФОРМИРОВАНИЮ СЕГМЕНТОВ АНАЛИЗА РЕЧИ

Афанасьев А. А.

Академия Федеральной службы охраны Российской Федерации

При обработке речевых сигналов основополагающей тенденцией является широкое использование информационных технологий в телекоммуникационных сетях связи. Переход к цифровым сетям связи и пакетной передаче данных позволил предоставить пользователям более широкий спектр защищенных 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. Широкое распространение в инфокоммуникациях в настоящее время получили методы кодирования речевых данных с переменной скоростью передачи и асинхронным вводом в канал связи.

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.

Выбор длины данного сегмента является весьма важной задачей. Его увеличение приводит к уменьшению скорости передачи в канале связи, а сокращение к повышению качественных характеристик синтезируемого сигнала, так как сигнал становится близким к стационарному, и, уменьшению времени задержки сигнала на обработку. Анализ речевого сигнала на фиксированном выбранном сегменте квазистационарности является достаточно грубым допущением, так как за пределами сегмента сигнал представляется равным нулю, что не соответствует действительности и приводит к появлению искажений на стыках сегментов при их анализе и кодировании, а также искажений в восприятии синтезированного речевого сигнала на приеме, при этом на вокализованных участках речи длина сегмента стационарности может быть увеличена, что связано с линейным характером образования речевого сигнала на этом участке, а на шумоподобных участках желательно ее уменьшать, так как речевой сигнал в данном случае имеет нестационарные свойства. Деление речевого сигнала на вокализованные и шумоподобные (невокализованные) сегменты достаточно широко используется в технике эффективного речевого кодирования. В результате применения предлагаемого способа налицо снижение требуемой скорости передачи при сохранении качественных показателей синтезированного речевого сигнала.

Сущность способа, предлагаемого в докладе, заключается в следующем. Выполняется анализ на основе линейного предсказания поступающего на вход вокодера речевого сигнала, деля его на сегменты по 20 миллисекунд, если принимается

решение о том, что сигнал является активной речью, то выделяется переход огибающей сигнала через нулевое значение и от положения отсчета со значением наиболее близким нулю выбирается длина сегмента соответствующая 20 миллисекундам и рассчитывается значение частоты основного тона и сигнала тон-шум. Если принимается решение о вокализованности анализируемого сигнала, то увеличивается длительность сегмента квазистационарности на количество отсчетов кратное периоду основного тона, но не более чем на 60 миллисекунд с обязательной проверкой на вокализованность следующих сегментов по 20 миллисекунд. Если принимается решение о шумоподобности следующего сегмента, то граница сегмента анализа выбирается кратной количеству отсчетов на периоде основного тона, но не более половины следующего сегмента длительностью 20 миллисекунд, если принимается решение о шумоподобности анализируемого сегмента, то длина сегмента анализа уменьшается, при этом граница сегмента формируется на значении близком нулю и кратном вычисленному периоду основного тона.

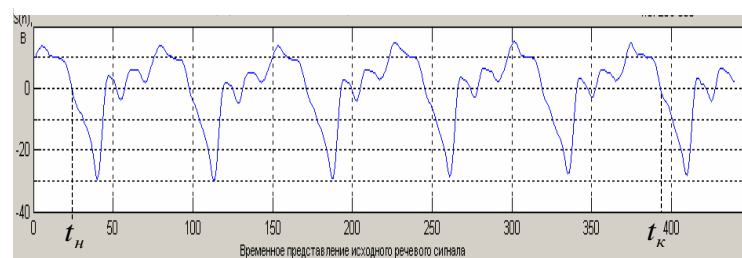


Рис.1. Временное представление установившегося режима следования сегментов квазистационарности речевого сигнала

При таком подходе с высокой вероятностью можно утверждать, что начальный и конечный отсчеты во вновь сформированном сегменте будут иметь значения близкие нулю, что значительно уменьшит возможные искажения на стыках сегментов.

В большинстве вокодеров с линейным предсказанием в качестве одного из параметров входящих в кадр передачи является частота основного тона говорящего, таким образом использование данного параметра в качестве одного из критериев формирования границ сегмента квазистационарности речевого сигнала не приведет к повышению скорости передачи в таких вокодерах, а существенно сократит ее, при этом повышается качество синтезированной речи на выходе вокодера с линейным предсказанием, так как отсчеты на стыках сегментов квазистационарности имеют практически нулевую энергию. В некоторых методах линейного предсказания частота основного тона говорящего не передается в кадре передачи, для таких методов данный параметр будет дополнительным, что снизит эффект от внедрения предлагаемого способа, однако общий выигрыш от ее использования будет достаточно велик. Методы выделения частоты основного тона речевого сигнала хорошо известны и нашли широкое распространение в речевом кодировании.

Использование данного способа для выделения сегментов квазистационарности при анализе речевого сигнала в вокодерах с линейным предсказанием будет рационально для класса систем кодирования речевого сигнала с переменной скоростью передачи.

Такой подход позволяет значительно снизить требуемую пропускную способность каналов связи для функционирования вокодеров с линейным предсказанием, при этом качественные показатели синтезированного сигнала значительно повышаются.