

Е.Г. ЖИЛЯКОВ, д-р техн. наук, проф., зав. каф. БелГУ (Белгород, Россия),

А.А. ФИРСОВА, аспирант БелГУ (Белгород, Россия)

ОБ АНАЛИЗЕ И СИНТЕЗЕ РЕЧЕВЫХ СИГНАЛОВ В IP-ТЕЛЕФОНИИ

Рассмотрены методы анализа и синтеза речевых сигналов. Исследованы особенности распределения энергии звуков русской речи, соответствующих различным дикторам. Проведены исследования зависимости качества звучания синтезированного сигнала от порядка модели линейного предсказания. Рассмотрены существующие методы обнаружения и кодирования пауз в IP-телефонии. Ил.: 1. Табл.: 2. Библиогр.: 10 назв.

Ключевые слова: речевой сигнал, синтез речевого сигнала, распределение энергии звуков русской речи, линейное предсказание, обнаружение и кодирование пауз, IP-телефония.

Постановка проблемы. В настоящее время IP-телефония является одним из наиболее интересных методов передачи речевой информации в режиме реального времени. Эта технология позволяет использовать IP-каналы для осуществления телефонной связи. Быстрое развитие и распространение IP-телефонии обусловлено широким распространением IP-сетей, а также богатым спектром предоставляемых услуг. Тем не менее, вопрос снижения загруженности каналов связи не может не интересовать специалистов в области передачи информации. Одним из несомненных преимуществ IP-телефонии является возможность использования в качестве оконечного оборудования персональных компьютеров, обладающих большими возможностями для преобразования информации в форму, позволяющую снизить объем битовых представлений, передаваемых в канал связи. Уменьшения объема передаваемых данных можно добиться несколькими способами: уменьшением частоты дискретизации, уменьшением разрядности при кодировании значений сигнала, обнаружением и кодированием пауз в речевых данных. Наиболее интересным остается последний способ уменьшения объема битовых представлений передаваемых данных. Любой алгоритм обнаружения и кодирования пауз основывается на результатах анализа речевого сигнала. Существующие алгоритмы не позволяют обеспечить стопроцентное обнаружение пауз в сигнале. Таким образом, вопрос поиска методов и алгоритмов для обнаружения пауз, несомненно, является актуальным.

Анализ литературы. Вопросам, связанным с IP-телефонией посвящено большое количество публикаций [1 – 3]. Эта технология позволяет осуществить связь компьютер – компьютер, компьютер – телефон или телефон – телефон, используя при этом каналы IP-сети. Речевой сигнал по каналам

такой сети передается в цифровом виде, закодированном определенным образом. В IP-телефонии алгоритм преобразование сигнала в определенный код и обратно называется кодеком. Этот же термин применяется и к устройству, реализующему эти преобразования. Обычно выделяют 3 типа кодеков:

1. Кодеки с импульсно-кодовой модуляцией (ИКМ) и адаптивной дифференциальной импульсно-кодовой модуляцией (АДИКМ).
2. Кодеки с вокодерным преобразованием речевого сигнала.
3. Комбинированные (гибридные) кодеки.

Наилучшее качество обеспечивают кодеки первого типа, но при этом они наиболее требовательны к полосе канала, поэтому они применяются лишь в тех случаях, когда требуется обеспечить максимальное качество кодирования речевой информации при небольшом числе одновременных разговоров. Наиболее распространенными являются кодеки третьего типа, позволяющие обеспечить приемлемое качество звучания при достаточно высокой компрессии сигнала.

Большинство существующих кодеков используют дополнительно алгоритмы обнаружения и кодирования пауз. Эти алгоритмы основаны на анализе сигнала и выявлении различий между информационным фрагментом и паузой.

Анализ – одна из основных составляющих обработки сигналов, основной целью которой является сравнение сигналов друг с другом для определения сходства и различия между ними. В [4] выделяют три основные составляющие анализа сигналов:

- измерение числовых параметров сигналов: энергия, средняя мощность и среднеквадратическое значение;
- разложение сигнала на элементарные составляющие для их рассмотрения по отдельности либо для сравнения свойств различных сигналов с использованием рядов и интегральных преобразований (ряд Фурье, преобразование Фурье);
- количественное измерение степени "похожести" различных сигналов с применением аппарата корреляционного анализа.

Одним из основных способов анализа речевых сигналов является использование преобразования Фурье для дискретных значений [4 – 6].

В кодеках IP-телефонии часто используется анализ особенностей речевых сигналов, основанный на модели авторегрессии. Распределение спектральной плотности модели авторегрессии дает наглядное представление о местоположении максимумов при различных значениях частот для разных звуков. Модель авторегрессии широко используется различными кодеками в алгоритмах VAD для обнаружения активности речи, а также в алгоритмах синтеза речи. Это позволяет значительно снизить объемы передаваемой информации.

В современных системах IP-телефонии применяются алгоритмы синтеза речи, основанные на линейном предсказании [7, 8].

Цель статьи: оценка методов анализа и синтеза речевых сигналов, сравнение методов обнаружения пауз в IP-телефонии.

Алгоритмы анализа речевых сигналов и методы обнаружения пауз.

Согласно акустической теории речеобразования восприятие звуков определяется областями максимальной концентрации энергии – формантами.

Каждому звуку речи соответствует своя форма акустического речевого тракта с присущей ему структурой формант. Форманта характеризуется амплитудой и частотой для гласных – добротностью или шириной полосы пропускания B_i . Частоты формант изменяются в достаточно широких пределах как для звуков, произносимых различными дикторами, так и для звуков, произносимых одним и тем же диктором, но в различных звукосочетаниях. Однако, области, в которых располагаются соответствующие формантные частоты, более или менее одинаковы даже для различных языков.

Другой важной особенностью речевых сигналов является коррелированность его значений. На основе этой особенности возможно осуществление анализа сигнала на обнаружение пауз, а также синтез речевого сигнала по некоторому количеству отсчетов. Было проведено исследование синтеза речевого сигнала для различных длин окна анализа при изменении порядка модели от 1 до 30 с использованием метода линейного предсказания. Синтез проводился по p первым отсчетам и p коэффициентам предсказания. Суть метода заключается в том, что сигнал разбивается на окна одинаковой длины, затем для каждого окна рассчитывается p коэффициентов авторегрессии:

$$r(i) = \sum_{k=0}^{N-i} x(k)x(k+i), \quad (1)$$

где $i = 0, \dots, p$; p – порядок модели; N – длина окна анализа; x – анализируемый сигнал.

Для восстановления сигнала используются рассчитанные коэффициенты авторегрессии и первые p отсчетов:

$$x_c(i) = \sum_{k=1}^p a(k)x(i-k), \quad (2)$$

где $i = 1, 2, \dots, N$; p – порядок модели; N – длина окна анализа; x – анализируемый сигнал; a – коэффициенты предсказания, рассчитанные с использованием алгоритма Дурбина и коэффициентов авторегрессии r .

Использование этого метода позволяет уменьшить объем передаваемых данных, но при этом значительно влияет на качество звучания речи, добавляя "металлический оттенок". Качество звучания улучшается с увеличением порядка модели предсказания. При этом разборчивость хорошая даже при малых порядках модели.

Другим методом, используемым в кодеках IP-телефонии и основанном на автокорреляционном анализе, является метод определения активности речи VAD. Он основан на оценке значения усредненного автокорреляционного LPC-параметра A и сравнении со средними значениями автокорреляции сигнала r , вычисленными в текущем фрейме [7]:

$$Df_h = A_h(0)r_h(0) + 2\sum_{i=1}^p A_h(i) \frac{r_h(i)}{r_h(0)}, \quad (3)$$

где Df_h – средний спектр, вычисленный для h -го отрезка; $r_h(i)$ – коэффициенты автокорреляции h -го отрезка входного сигнала; $A_h(i)$ – коэффициенты автокорреляции средних LPC-параметров h -го отрезка; p – порядок модели.

На рис. 1 представлена структурная схема VAD с обработкой в частотной области.

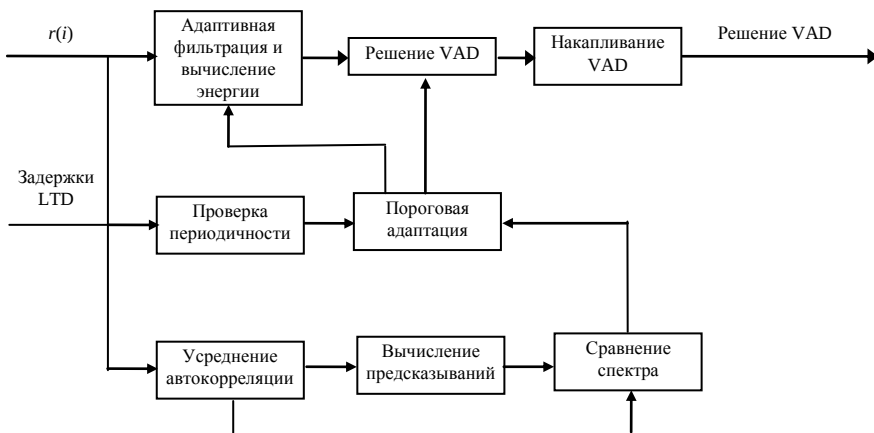


Рис. Структурная схема VAD с обработкой в частотной области, применяемая в системах GSM и IP-телефонии

Для оценки качества работы этого алгоритма были проведены исследования работы алгоритма VAD для различных значений порядка модели p . Вероятность ошибки "ложная тревога" рассчитывалась по формуле:

$$P_{\text{л.т.}} = \frac{N_{\text{пр_пауз}}}{N_{\text{сигн}}}, \quad (4)$$

где $N_{\text{пр_пауз}}$ – количество отсчетов, принятых за паузу в заведомо известном полезном сигнале; $N_{\text{сигн}}$ – количество отсчетов в заведомо известном полезном сигнале.

Вероятность ошибки "пропуск цели" можно рассчитать как:

$$P_{\text{п.ц.}} = \frac{N_{\text{пр_сигн}}}{N_{\text{пауз}}}, \quad (5)$$

где $N_{\text{пр_сигн}}$ – количество отсчетов, принятых за полезный сигнал в заведомо известном фрагменте пауз; $N_{\text{пауз}}$ – количество отсчетов в заведомо известном фрагменте пауз.

В табл. 1 представлены результаты оценки вероятности принятия ошибочных решений при использовании алгоритма VAD для различных значений порядка модели. При этом за нулевую принималась гипотеза о наличии паузы. В этом случае $P_{\text{л.т.}}$ – вероятность ошибки "ложная тревога" (когда речевой сигнал принимается за паузу), а $P_{\text{п.ц.}}$ – вероятность ошибки "пропуск цели" (когда пауза принимается за речевой сигнал). Как видно, линейной зависимости значения вероятности принятия ошибочного решения от значения порядка модели нет. Обычно этот алгоритм используется совместно с алгоритмом линейного предсказания и значение порядка выбирается, основываясь на критерии Акаике для значений ошибки синтеза сигнала.

Таблица 1

Оценка вероятности принятия ошибочного решения при различных параметрах модели VAD

Параметры	$P_{\text{л.т.}}$	$P_{\text{п.ц.}}$
$p = 2$	0,006	0,014
$p = 3$	0,006	0,129
$p = 4$	0	0,257
$p = 5$	0	0,257
$p = 10$	0	0,357
$p = 15$	0	0,257
$p = 20$	0	0,171
$p = 25$	0	0,186
$p = 30$	0	0,229

Использование этого алгоритма позволяет значительно снизить объем передаваемых данных. Это связано с тем, что в диалоге до 60% пауз. Фрагменты сигнала, принятые за паузу, не передаются, передается только информация о начале и длительности (или конце) паузы.

Как видно, описанные методы не являются идеальными. Для обнаружения пауз в речевом сигнале можно использовать метод оценки, основанный на вариационном методе анализа речевых сигналов [9, 10]. Этот метод позволяет осуществить частотный анализ энергетических характеристик исследуемой функции. Полный набор долей энергии отрезка сигнала можно определить в этом случае следующим образом

$$P_r \approx P_{1r} = \|\vec{y}_{1r}\|^2 = \sum_{k=1}^{J_{r_r}} (y_{kr}^2)_1, \quad (6)$$

причем,

$$\vec{y}\vec{y} = (\vec{y}_{1r}^T, \dots, \vec{y}_{1R}^T)^T = A\vec{A}\vec{x}, \quad (7)$$

где $A\vec{A}$ – блочная матрица, состоящая из собственных значений собственных векторов субполосной матрицы A_r [9, 10]; $\vec{x}$ – анализируемый отрезок;

$$A_r = \{a_{ik}^r\},$$

$$a_{ik}^r = (\sin(v_{r+1}(i-r)) - \sin(v_r(i-k))) / (\pi(i-k)), \quad i, k = 1, 2, \dots, N. \quad (8)$$

Субполосная матрица является симметричной и неотрицательно определённой. Поэтому она обладает полной системой ортонормальных собственных векторов.

С использованием формулы (6) были проведены исследования всех звуков русского алфавита для 10 различных дикторов (5 мужчин, 5 женщин). Результаты исследований показали, что распределение долей энергии различно и зависит от самого звука, положения звука в слове, интонации и тембра голоса диктора. Выбор длины окна анализа порядка основан на стремлении минимизировать ошибку захвата других звуков. Изменение распределения энергии в зависимости от начала, середины или конца звука вызвано тем, что речевой аппарата человека не может перестраиваться мгновенно.

Распределение долей энергии по частотным интервалам может быть использовано для обнаружения пауз в речевом сигнале [9, 10]. Решающая функция для проверки гипотезы о том, что анализируемый отрезок сигнала соответствует паузе между звуковыми данными (нулевая гипотеза) имеет вид

$$S = \max(P_r / P_r^\Pi), \quad \forall r = 1, 2, \dots, R, \quad (9)$$

где в знаменателе стоят результаты предварительного усреднения по достаточно большому количеству отрезков сигнала, заведомо относящихся к паузам, долей энергий, попадающих в заданный частотный интервал [9, 10]:

$$P_r^\Pi = \sum_{k=1}^{N_y} (P_r)_k^\Pi / N_y. \quad (10)$$

Здесь N_y – количество отрезков сигнала в паузе, которые используются для усреднения, что соответствует оцениванию математических ожиданий вычисляемых долей энергий в соответствующих частотных интервалах.

Если выполняется неравенство

$$S > h_\alpha, \quad (11)$$

то нулевая гипотеза отвергается, иначе принимается решение о паузе [9, 10].

Символ h_α в правой части неравенства (11) означает порог, обеспечивающий заданный уровень вероятности ложной тревоги, то есть выполнение неравенства

$$P\{S > h_\alpha\} \leq \alpha \ll 1. \quad (12)$$

Значение порога можно получить из основанного на неравенстве Чебышева выражения

$$h_\alpha \leq \bar{S}_\Pi + D_\Pi / d\sqrt{\alpha}, \quad (13)$$

где $\bar{S}_\Pi$ – результат оценки математического ожидания решающей функции, D_Π – результат оценки дисперсии решающей функции [9, 10].

В табл. 2 представлены результаты оценки вероятности принятия ошибочного решения при использовании различных параметров модели, основанной на предложенном методе. При этом за нулевую принималась гипотеза о наличии паузы. В этом случае $P_{\text{л.т.}}$ – вероятность ошибки "ложная тревога" (когда речевой сигнал принимается за паузу), а $P_{\text{п.ц.}}$ – вероятность ошибки "пропуск цели" (когда пауза принимается за речевой сигнал). Как видно, значение вероятности принятия ошибочного решения невелико и значительно меньше, чем при использовании алгоритма VAD.

Таблица 2

Оценка вероятности принятия ошибочного решения при различных параметрах модели, основанной на новом методе

Параметры	$P_{\text{л.т.}}$	$P_{\text{п.ц.}}$
$N = 64, R = 16$	0	0
$N = 64, R = 32$	0	0
$N = 128, R = 16$	0	0
$N = 128, R = 32$	0	0,028
$N = 128, R = 64$	0	0,057

Выводы. Результаты исследования влияния порядка модели на величину вероятности ошибочного принятия решения в алгоритме VAD не выявило линейной зависимости между этими параметрами. Величина погрешности между синтезированным и исходным сигналом при реализации алгоритма линейного предсказания обратно пропорциональна значению порядка модели авторегрессии. Важно отметить, что чем выше порядок модели, тем выше качество звучания синтезированного сигнала, но разборчивость сохраняется даже при малых значениях порядка модели. Исследования на базе вариационного анализа позволили сравнить распределение долей энергии по частотным интервалам для всех звуков русской речи для различных дикторов. Исследования показали, что все звуки имеют свое особенное распределение, но при этом можно выделить некоторую схожесть для одного и того же звука разных дикторов. Выбор малой длительности окна анализа позволил выявить различие распределения энергии одного звука для начального, среднего и

последнего фрагментов. Это вызвано тем, что речевой аппарат человека не может мгновенное перестроиться и захватывает соседние звуки. Сравнение результатов оценки вероятностей принятия ошибочного решения при различных параметрах алгоритма, использующего вариационный метод анализа, с алгоритмом VAD показало, что новый алгоритм имеет меньшую вероятность принятия ошибочного решения. Таким образом, можно сказать, что использование нового метода обнаружения пауз позволит сократить вероятность принятия ошибочного решения, а значит, точнее определять паузы в речевых данных.

Список литературы: 1. Росляков А.В. IP-телефония / А.В. Росляков, М.Ю. Самсонов, И.В. Шибаева. – М.: Радио и связь, 2003. – 252 с. 2. Гольдштейн Б.С. IP-телефония / Б.С. Гольдштейн, А.В. Пинчук, А.Л. Суховицкий. – М.: Эко-Трендз, 2001. – 336 с. 3. Варламова Е. IP-телефония в России / Е. Варламова // Connect! Мир связи, 1999. – № 9. 4. Сергиенко А.Б. Цифровая обработка сигналов / А.Б. Сергиенко. – СПб.: Питер, 2003. – 604 с. 5. Назаров М.В. Методы цифровой обработки и передачи речевых сигналов / М.В. Назаров, Ю.Н. Прохоров. – М.: Радио и связь, 1985. – 176 с. 6. Гольденберг Л.М. Цифровая обработка сигналов. / Л.М. Гольденберг. – М.: Радио и связь, 1985. – 312 с. 7. Шелухин О.И. Цифровая обработка и передача речи / О.И. Шелухин, Н.Ф. Лукьянцев; под ред. О.И. Шелухина. – М.: Радио и связь, 2000. – 456 с. 8. Рабинер Л.Р. Цифровая обработка речевых сигналов / Л.Р. Рабинер, Р.В. Шафер; под ред. М.В. Назарова, Ю.Н. Прохорова. – М.: Радио и связь, 1981. – 495 с. 9. Жилияков Е.Г. Методы обработки речевых данных в информационно-телекоммуникационных системах на основе частотных представлений / Е.Г. Жилияков, С.П. Белов, Е.И. Прохоренко. – Белгород: БелГУ, 2007. – 136 с. 10. Белов С.П. Об уменьшении объема трафика при пакетной передаче речевых сообщений за счет кодирования пауз / С.П. Белов, Е.И. Прохоренко // Научные ведомости Белгородского государственного университета. Серия: Информатика и прикладная математика. – Белгород, 2006. – № 1 (21). – Вып. 2. – С. 141–148.

УДК 621.391

Про аналіз і синтез мовних сигналів в IP-телефонії / Жилияков Є.Г., Фірсова А.А. // Вісник НТУ "ХПІ". Тематичний випуск: Інформатика і моделювання. – Харків: НТУ "ХПІ". – 2009. – № 43. – С. 84 – 91.

Розглянуті методи аналізу і синтезу мовних сигналів. Досліджені особливості розподілу енергії звуків російської мови, відповідних різним дикторам. Проведено дослідження залежності якості звучання синтезованого сигналу від порядку моделі лінійного прогнозу. Розглянуті існуючі методи виявлення і кодування пауз в IP-телефонії. Іл.: 1. Табл. 2. Бібліогр.: 10 назв.

Ключові слова: мовний сигнал, синтез мовного сигналу, розподіл енергії звуків російської мови, лінійний прогноз, виявлення і кодування пауз, IP-телефонія.

UDK 621.391

About an analysis and synthesis of vocal signals in IP-telephony / Zhilyakov E.G., Firsova A.A. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2009. – №. 43. – P. 84 – 91.

The methods of analysis and synthesis of vocal signals are considered. The features of distributing of energy of sounds of Russian speech, proper different announcers are probed. Researches of dependence of quality of sounding of the synthesized signal are conducted from the order of model of linear prediction. The existent methods of discovery and encoding of pauses are considered in IP-telephony. Figs: 1. Tabl.: 2. Refs: 10 titles.

Keywords: vocal signal, synthesis of vocal signal, distributing of energy of sounds of Russian speech, linear prediction, discovery and encoding of pauses, IP-telephony.

Поступила в редакцию 20.10.2009.