ко А.В. Новый метод оптимального субполосного преобразования в задаче сжатия речевых данных. – «Вопросы радиоэлектроники», сер. ЭВТ, 2010, вып. 1, с. 49-55.

6. Жиляков Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным. Белгород, БелГУ, 2007. 160 с.

7. Жиляков Е.Г., Белов С.П. и Прохоренко Е.И. Вариационные методы частотного анализа звуковых сигналов. – В сб.: Труды учебных заведений связи, 2006, № 174 с. 163-170.

8. Жиляков Е.Г., Белов С.П. и Прохоренко Е.И. О субполосном преобразовании звуковых сигналов. "Труды РНТО РЭС им. А.С. Попова", .сер. Цифровая обработка сигналов и ее применение, 2006, вып. VIII-1, с. 167-169.

9. Гантмахер Ф.Р. Теория матриц. М., Физматлит, 2004. 560 с.

10. Сизиков В.С. Математические методы обработки результатов измерений. Учеб. для вузов. СПб., Политехника, 2001.

Статья поступила 12.10 2010

А.А. Фирсова, д. ф.-м. н. А.Н. Чеканов (БелГУ)

A.A. Firsova, A. N. Chekanov

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ АЛГОРИТМОВ ОБНАРУЖЕНИЯ ПАУЗ В IP-ТЕЛЕФОНИИ

COMPUTER MODELLING OF ALGORITHMS OF DETECTION OF PAUSES IN IP — TELEPHONY

В статье рассмотрены различные алгоритмы обнаружения пауз Проведен анализ возможности использования различных алгоритмов обнаружения пауз в режиме реального времени Проведено сравнение эффективности использования различных алгоритмов

Keywords. speech signal, speech signal analysis, a model of VAD, pause detection algorithm, the frequency representation, IP-telephony

Развитие информационно-телекоммуникационных систем направлено на обеспечение возможности использования естественных форм общения (речь, изображение) с помощью современных средств обработки сигналов. В настоящее время обработка сигналов

в информационно-телекоммуникационных системах осуществляется преимущественно с помощью компьютерных технологий. К наиболее широко используемым способам обработки речевых данных относятся запись речевых сообщений для последующей обработки и хранения, IP-телефония, управление голосом и т.д. При IP-телефонии происходит обработка речевых сигналов в режиме реального времени. При выборе алгоритмов обработки речевых данных в таких системах особое внимание уделяется вопросам уменьшения объемов битовых представлений при передаче речевого сообщения, а также скорости обработки отрезков речевых данных.

Речевые сигналы на 16% состоят из пауз между звуками речи. Во время диалога это значение возрастает до 60%. Одним из эффективных способов уменьшения объемов битовых представлений является обнаружение участков сигнала, относящихся к паузам и передаче только начала и длительности паузы [1].

При работе с речевыми данными в режиме реального времени важно, чтобы время обработки сигнала не превышало заданного порога. Задержка на обработку речевых сигналов на передающей стороне не должна превышать 20 мс, а на декодирование – 10 мс [2].

В современных информационно-телекоммуникационных системах широкое применение нашли алгоритмы VAD (Voice Activity Detector). Реализация алгоритмов VAD основана на различиях речевого сигнала и шума. При этом проверяются два условия: превышение сигналом порога по энергии и стационарность сигнала.

Для определения является сигнал стационарным или нет, проверяется выполнение условия [1]:

$$|Df_n - Df_{n-1}| < \Delta D_{nop}, \qquad (1)$$

где: ΔD_{nop} – пороговое значение;

Df_n – средние значения автокорреляции сигнала, вычисленные для n-го отрезка;

Df_{n-1} – средние значения автокорреляции сигнала, вычисленные для $(n-1)$-го отрезка.

Если условие (1) выполняется, то текущий фрейм считается стационарным, иначе – нестационарным [1,3,4].

В неравенстве (1) Df_n определяется следующим образом:

$$Df_n = A_n(0)r_n(0) + 2\sum_{i=1}^{p} (A_n(i)\cdot r_n(i)/r_n(0)), \qquad (2)$$

$$r(i) = \sum_{k=0}^{N-1} x(k)x(k+i), \quad (3)$$

$$A(i) = \sum_{k=0}^{p-1} a(k)a(k+i), \quad (4)$$

где: $r_n(i)$ – коэффициенты автокорреляции n-го отрезка входного сигнала;

$A_n(i)$ – коэффициенты автокорреляции средних LPC-параметров n-го отрезка;

p – порядок модели линейного предсказания;

i – изменяется от 0 до p;

N – длина окна анализа;

x – анализируемый сигнал;

a – средние LPC-параметры, рассчитываемые на основе средних коэффициентов автокорреляции с использованием алгоритма Дурбина.

Второе условие принятия решения о наличии или отсутствии шума:

$$E < E_{пор}, \quad (5)$$

где: E – остаточная энергия;

$E_{пор}$ – значение порога.

Остаточную энергию определяется по формуле:

$$E = A(0)r(0) + 2\sum_{i=1}^{p} A(i)r(i), \quad (6)$$

где: $r(i)$ – коэффициенты автокорреляции входного сигнала;

$A(i)$ – коэффициенты автокорреляции средних LPC-параметров;

p – порядок модели линейного предсказания.

Пороговые значения $E_{пор}$ и $\Delta D_{пор}$ могут быть определены экспериментально на основе анализа обучающей выборке сигнала, относящегося к паузам. Здесь для определения порога анализировалось 100 фрагментов пауз одинаковой длительности N (N=64, N=128 отсчетов). Для каждого отрезка вычислялись значения E (6) и Df_n (2). В качестве $E_{пор}$ выбиралось максимальное значение E среди всех проанализированных отрезков. В качестве $\Delta D_{пор}$ выбиралось максимальное значение, из полученных на этапе обучения, абсолютных величин разностей между Df соседних фрагментов

Полученные на этапе обучения значения $E_{пор}$ и $\Delta D_{пор}$ используются в дальнейшем для принятия решения об отсутствии или на-

личии паузы. Принятие решения о наличии пауз осуществляется в том случае, если выполняются условия (1) и (5).

Исследование эффективности работы метода проводилось для различных значений порядка модели предсказания p. Сигнал разбивался на окна одинаковой длины N, для каждого окна анализа определялись значения E (6) и Df_n (2). Затем проверялись условия (1) и (5). Если выполняются оба условия, то принимается решений о наличии паузы, иначе – принимается решение о наличии полезного сигнала.

При выборе параметров для такой модели необходимо учитывать различные критерии. Одним из основных критериев является оценка вероятности принятия ошибочного решения о наличии или отсутствии пауз. Для систем реального времени другим важным критерием является время обработки речевого сигнала.

Оценка вероятности принятия ошибочного решения осуществлялась на основе определения вероятностей ошибок первого и второго рода. При этом за основную принималась гипотеза о наличии паузы. В этом случае $P_{л.т.}$ – вероятность ошибки «ложная тревога» (когда основная гипотеза о наличии паузы ошибочно отвергается), а $P_{п.ц.}$ – вероятность ошибки «пропуск цели» (когда основная гипотеза о наличии паузы ошибочно принимается).

Вероятность принятия ошибочного решения определялась в два этапа. На первом этапе анализировался фрагмент сигнала, относящийся к паузе, отличающийся от обучающей выборки. Вероятность ошибки «ложная тревога» определялась как:

$$P_{л.т.} = 1 - N_o/N_n, \tag{7}$$

где: N_o – количество отрезков, отнесенных к паузе,

N_n – количество отрезков паузы.

На втором этапе анализировался фрагмент сигнала, относящийся к речи. Вероятность ошибки «пропуск цели» определялась как:

$$P_{п.ц.} = 1 - N_o/N_p, \tag{8}$$

где: N_o – количество отрезков, отнесенных к паузе,

N_p – количество отрезков речевого сигнала.

Для определения значения вероятности $P_{л.т.}$ анализировалось 3992 отрезка. Для определения значения вероятности $P_{п.ц.}$ анализировалось 3843 отрезка. В табл. 1 представлены результаты исследования работы алгоритма VAD при различных значениях длины окна

анализа для значения порядка фильтра равного 8, которое наиболее часто используется в фильтрах линейного предсказания [1].

Таблица 1

Оценка вероятности принятия ошибочного решения алгоритма VAD

Параметры	P_{1m}		$P_{n\text{ч}}$	
	N=64	N=128	N=64	N=128
p=8	0,16	0,15	0,00	0,00

Основную опасность при обработке сигнала представляют ошибки «пропуск цели», поэтому при разработке алгоритма VAD главное, чтобы вероятность $P_{n\text{ч}}$ была минимальна, при этом вероятность P_{1m}, чаще всего, выбирается достаточно большой.

Таким образом, рассмотренный метод имеет достаточно большое значение P_{1m}, что не позволяет минимизировать объем передаваемых данных и приводит к тому, что сегментация не является достоверной.

Исследования тонкой структуры энергетического спектра речевого сигнала в частотной области позволили установить, что энергия звуков речи распределена неравномерно и сосредоточена в достаточно узких частотных интервалах, в то время как энергия отрезка сигнала, принадлежащего паузе, распределена равномерно во всем анализируемом частотном диапазоне. В связи с этим предлагается в качестве процедуры обнаружения пауз использовать метод, основанный на принципе учета отличий распределения энергии речевого сигнала по частотному диапазону, соответствующему звуку, по сравнению с распределением энергии сигнала в паузе.

Для анализа особенностей речевых сигналов можно использовать метод вычисления точных значений долей энергии, попадающих в заданный частотный интервал [5].

Полный набор долей энергии отрезка сигнала можно определить следующим образом:

$$P_r = \vec{x}^{\,T} A_r \vec{x}, \qquad (9)$$

где: $\vec{x}$ – анализируемый отрезок сигнала;

r – номер частотного интервала, изменяющийся от 1 до R;

A_r – субполосная матрица, рассчитанная для r-го частотного ин тервала:

$$A_r=\{a^r_{ik}\}$$

$$a^r_{ik}=(\sin(v_{r+1}(i-k))-\sin(v_r(i-k)))/(\pi(i-k)),\ \ i,k=1,\ldots,N, \qquad (10)$$

где v_r, v_{r+1} – границы r-ого частотного интервала, причем:

$$0 \le v_r < v_{r+1} \le \pi, \quad r=1,\ldots,R, \qquad (11)$$

$$v_{r+1}-v_r=\pi/R, \qquad (12)$$

где R – количество частотных интервалов, на которые разбивается частотная ось.

Для принятия решения о наличии или отсутствии паузы вычисляется решающая функция для проверки гипотезы о том, что анализируемый отрезок сигнала соответствует паузе между звуками речи (основная гипотеза) [6]:

$$W_{NR}=f^m_{NR}/R, \qquad (13)$$

где f^m_{NR} – минимальное количество частотных интервалов (частотная концентрация), в которых сосредоточена заданная доля энергии m звукового отрезка, т.е.:

$$f^m_{NR}=\min d^m_{NR}. \qquad \textbf{(14)}$$

Здесь выполняется неравенство:

$$\sum_{k=1}^{d^m_{NR}} P_{(k),N} \ge m\|\vec{x}_N\|^2 = m\sum_{i=1}^{N} x_i^2, \qquad (15)$$

где: $\vec{x}_N$ – анализируемый отрезок сигнала,

m – заданное значение доли энергии сигнала,

$P_{(k),N}$ – упорядоченные по убыванию доли энергий сигнала, попадающих в заданные частотные интервалы, т.е.:

$$P_{(k),N} \in \{P_{rN},\ r=1,\ldots,R\}\ P_{(k+1),N} \le P_{(k),N},\ k=1,\ldots,R \qquad (16)$$

где P_{rN} – доли энергий сигнала, попадающих в заданные частотные интервалы, определяемые с помощью (9).

Если выполняется неравенство:

$$W_{NR} < w_{пор}, \qquad (17)$$

то основная гипотеза отвергается, в противном случае принимается решение о наличии паузы.

$w_{пор}$ в (17) – пороговое значение, которое выбирается на основе анализа особенностей распределения долей энергии звуков речи и шума [6]. Анализ особенностей распределения энергии по частотным интервалам звуков русской речи показал, что все звуки речи имеют различное распределение долей энергии по частотным интервалам, при этом основная энергия сигнала сосредоточена в узком частотном диапазоне. Здесь представлены результаты эксперимен-

тов для пороговых значений $w_{пор}$=0,4 и $w_{пор}$=0,5.

Для оценки эффективности метода анализировались отрезки одинаковой длины N (64, 128 отсчетов). Проводились эксперименты при различных значениях количества частотных интервалов, на которые разбивается частотная ось R: 16, 32, 64; и значения заданной доли энергии m=0,80÷0,99.

Оценка вероятностей $P_{пт}$ (когда основная гипотеза о наличии паузы ошибочно отвергается) и $P_{пц}$ (когда основная гипотеза о наличии паузы ошибочно принимается) осуществлялась так же как и при исследовании эффективности алгоритма VAD (7), (8).

Сравнение результатов работы алгоритма показывает, что при наименьшей вероятности $P_{пц}$ меньшее значение вероятности $P_{пт}$ достигается при N=128, R=32, $w_{пор}$=0,5, m=0,96. В табл. 2 представлены результаты экспериментов при некоторых параметрах модели.

Таблица 2

Оценка вероятности принятия
ошибочного решения алгоритма без обучения при N=128 R=32

Параметры	$P_{пт}$		$P_{пц}$	
	$w_{пор}$ =0,4	$w_{пор}$ =0,5	$w_{пор}$ =0,4	$w_{пор}$ =0,5
m=0.96	0,02	0,15	0,06	0,00

Сравнение рассмотренного метода с работой алгоритма VAD показывает, что на различных участках сигнала рассмотренный алгоритм может работать с меньшим значением вероятности ошибки. Но этот метод существенно зависит от типа шума и особенностей речевого аппарата диктора и на некоторых участках работает хуже алгоритма VAD. Для анализируемого фрагмента вероятность $P_{пт}$ для $w_{пор}$ =0,5, m=0,96 ($P_{пц}$ ≈0, $P_{лт}$ ≈0,15) такая же, как и вероятность $P_{лт}$ алгоритма VAD ($P_{пц}$ ≈0, $P_{лт}$ ≈0,15).

Другой способ обнаружения пауз заключается в использовании процедуры обучения на основе анализа особенностей распределения долей энергии по частотным интервалам в паузе.

На этапе обучения для отрезков сигнала, заведомо относящихся к шуму, оцениваются характеристики вида [5,7]:

$$P_r^{П} = \sum_{k=1}^{N_y} (P_r)_k^{П}/N_y, \qquad (18)$$

где: N_y – количество отрезков сигнала в паузе, которые используются для усреднения (обучения), что соответствует оцениванию математических ожиданий вычисляемых долей энергий в соответствующих частотных интервалах;

$(P_r)_k^{П}$ – доли энергий в соответствующих частотных интервалах для N_y отрезков обучающей выборки.

В данном случае решающая функция имеет вид:

$$S = \max(P_r/P_r^{П}),\ \forall\ r = 1,\ldots,R, \tag{19}$$

где: P_r – доли энергий, попадающих в заданные частотные интервалы (9);

$P_r^{П}$ – результаты предварительного усреднения по достаточно большому количеству отрезков сигнала, заведомо относящихся к паузам, долей энергий, попадающих в заданный частотный интервал (18):

Если выполняется неравенство:

$$S > h_\alpha, \tag{20}$$

где h_α – порог, обеспечивающий заданный уровень вероятности ложной тревоги α на обучающей выборке,

то основная гипотеза о наличии паузы отвергается, в противном случае принимается решение о наличии паузы.

Для определения значения порога используется обучающая выборка относящихся к паузе данных. При этом после вычислений оценок математических ожиданий вида (19) вычисляются оценки математического ожидания и дисперсии решающей функции [5,7]:

$$\bar{S}_{П} = \sum_{k=1}^{N_y} (S_k^{П})/N_y, \tag{21}$$

$$D_{П}^2 = \sum_{k=1}^{N_y} (S_k^{П})^2 / N_y - \bar{S}_{П}^2, \tag{22}$$

где: $S_k^{П}$ – значение решающей функции на k-ом анализируемом отрезке заведомо относящихся к паузе данных;

N_y – количество отрезков сигнала обучающей выборки заведомо относящихся к паузе.

Пороговое значение, обеспечивающее заданный уровень вероятности ложной тревоги α на обучающей выборке, определяется на основе неравенства:

$$h_\alpha \le \bar{S}_{П} + D_{П}/a_m\sqrt{\alpha}\,, \tag{23}$$

где: α – вероятность ложной тревоги, задаваемая на этапе обучения;

$\bar{S}_П$ – математическое ожидание решающей функции;

$D_П$ – дисперсия решающей функции;

a_m – коэффициент, превышающий значение 2 и определяемый в процессе обучения [6].

В качестве обучающей выборки использовалось 400 отрезков сигнала, соответствующего паузе. Отрезки были получены в результате разбиения сигнала на окна одинаковой длины N (64, 128 отсчетов) с шагом 5 отсчетов.

Для оценки эффективности метода анализировались отрезки одинаковой длины N (64, 128 отсчетов). Проводились эксперименты при различных значениях количества частотных интервалов, на которые разбивается частотная ось R: 16, 32, 64.

Оценка вероятностей $P_{пт}$ (когда основная гипотеза о наличии паузы ошибочно отвергается) и $P_{пц}$ (когда основная гипотеза о наличии паузы ошибочно принимается) осуществлялась так же как и при исследовании эффективности алгоритма VAD (7), (8).

В табл. 3 представлены результаты экспериментальной оценки вероятностей ошибок «ложная тревога» и «пропуск цели».

Таблица 3

Оценка вероятности принятия ошибочного решения алгоритма с обучением на основе субполосной обработки N=128 R=32

Параметры	$P_{пт}$	$P_{пц}$
α=0,00002	0,02	0,00

Сравнение результатов работы алгоритма VAD, алгоритма без обучения и алгоритма с обучением на основе субполосной обработки показало, что алгоритм обнаружения пауз с обучением на основе субполосной обработки дает наименьшее значение вероятности $P_{пт}$ при условии, что вероятность $P_{пц}$ для всех исследованных алгоритмов одинакова. Так, для алгоритма с обучением $P_{пт}\approx 0{,}02$, а для алгоритма без обучения и алгоритма VAD $P_{пт}\approx 0{,}15$. Легко видеть, что применение алгоритма обнаружения пауз с обучением позволяет точнее определять участки отсутствия звука во фрагменте сигнала.

В случае использования описанных алгоритмов в режиме реального времени важно учитывать время, необходимое на обра-

ботку сигнала. Оценка времени проводилась для речевого сигнала длительностью 24000 отсчетов. Сигнал разбивался на окна одинаковой длины длиной 128 отсчетов. Время обработки оценивалось стандартными средствами программного приложения MatLab. Сравнение времени обработки одного и того же сигнала описанными выше алгоритмами показало, что при выбранных параметрах время, требуемое на обработку речевого сигнала, составляет одинаковую величину.

Таким образом, все эти алгоритмы имеют одинаковое время обработки речевого сигнала для выявления местоположения пауз. Следовательно, выбор алгоритма обнаружения пауз для систем реального времени может основываться только по значению вероятности ошибочного принятия решения. Анализ алгоритмов демонстрирует, что наилучшие показатели имеет алгоритм с обучением на основе субполосной обработки.

Литература

1. Шелухин О.И. и Лукьянцев Н.Ф. Цифровая обработка и передача речи. Под ред. О.И. Шелухина. М., Радио и связь, 2000. 456 с с илл.

2. Росляков А.В. и др. IP-телефония. М.; Радио и связь, 2003. 252 с.:

3. Герасимов, А.В. и др. Применение метода модифицированного линейного предсказания к задачам выделения акустических признаков речевых сигналов. – "Радиотехника и электроника", 2005, т. 50, № 10.

4. Коротаев Г.А. Некоторые аспекты линейного предсказания при анализе речевого сигнала. – "Зарубежная радиоэлектроника", 1991, № 7.

5. Жиляков Е.Г., Белов С.П. и Прохоренко Е.И. Методы обработки речевых данных в информационно-телекоммуникационных системах на основе частотных представлений. Белгород, 2007. 136 с.

6. Белов А.С. Разработка математических моделей и алгоритмов анализа и синтеза звуковых сигналов в цифровых слуховых аппаратах:. Автореф. дисс. на соискание ученой степени к.т.н. Белгород, 2009. 22 с.

7. Жиляков Е.Г. и Белов А.С. О фильтрации пауз в речевых данных для реализации в слуховых аппаратах. - «Вопросы радиоэлектроники», сер. ЭВТ, 2008, вып. 1, с.143-152.

Статья поступила 12.10.2010