Д.т.н., проф. Е.Г. Жиляков, А.С. Белов

О ФИЛЬТРАЦИИ ПАУЗ В РЕЧЕВЫХ ДАННЫХ ДЛЯ РЕАЛИЗАЦИИ В СЛУХОВЫХ АППАРАТАХ

Рассматривается новый алгоритм, предназначенный для обнаружении и удалении пауз при обработке речевых данных в слуховых аппаратах, позволяющий принимать решение о наличии паузы на основе учета отличий в распределении энергетических составляющих отрезков со звуковыми данными (и аддитивным шумом) и отдельно шумов в частотной области.

Введение

В современных цифровых слуховых аппаратах (СА) используются методы обнаружения пауз, основанные на статистическом анализе входного сигнала и распознавании речи от шума, которые контролируют распределение интенсивности входного сигнала на коротких временных интервалах в 15 частотных каналах, вычисляемых с помощью КИХ-фильтров. Оценки долей энергий в этих частотных каналах на основе выходных последовательностей КИХ-фильтров обладают достаточно большими погрешностями, что эквивалентно наличию дополнительных помех.

В связи с этим в статье рассматривается новый алгоритм обнаружения пауз в речевых данных, при их обработке в слуховых аппаратах, в основе создания которого лежит математический аппарат, позволяющий вычислять точные значения долей энергии отрезка сигнала, которая сосредоточена в любом конечном частотном интервале. Это дает возможность адекватно учесть свойство сосредоточенности энергии звуков речи в малом количестве достаточно узких частотных интервалов и принимать решение об отсутствии паузы на основе сравнения попадающих в них долей энергий отрезков со звуковыми данными (и аддитивным шумом) и отдельно шумов.

Математические основы метода

Пусть компоненты вектора $\vec{x} = (x_1,...,x_N)^T$ представляют собой значения некоторого звукового сигнала (функции времени),

которые соответствуют значениям аргумента iΔt, т.е.

$$x_i = x(i\Delta t), \quad i = 1,..., N, \tag{1}$$

где Δt - интервал дискретизации по времени.

Положим далее

$$X(\upsilon) = \sum_{k=1}^{N} x_k e^{-j(k-1)\upsilon}, \tag{2}$$

т.е. $X(\upsilon)$ представляет собой трансформанту Фурье (амплитудный частотный спектр (АМС)) отрезка отсчетов сигнала (вектора), в качестве области определения которой естественно рассматривать (нормированная частота)

$$-\pi \le \upsilon \le \pi, \tag{3}$$

так, что имеет место обратное преобразование (справедливо представление)

$$x_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\upsilon) e^{j(i-1)\upsilon} d\upsilon. \tag{4}$$

Отсюда нетрудно получить равенство Парсеваля

$$\|\vec{x}\|^2 = \sum_{k=1}^{N} x_k^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\upsilon)|^2 d\upsilon, \tag{5}$$

так, что

$$P_r(\vec{x}) = \frac{1}{2\pi} \int_{\upsilon \in V_r} |X(\upsilon)|^2 d\upsilon \tag{6}$$

представляет собой долю энергии отрезка сигнала (евклидовой нормы вектора), соответствующую частотному интервалу

$$V_r = \left[-\upsilon_2^r, -\upsilon_1^r\right) \cup \left[\upsilon_1^r, \upsilon_2^r\right). \tag{7}$$

Подстановка определения (2) в интеграл (6) позволяет получить выражение вида:

$$P_r(\vec{x}) = \vec{x}^T A_r \vec{x}, \tag{8}$$

где $A_r = \{a_{ik}^r\}$,

$$a_{ik}^r = \begin{cases} \dfrac{\sin\left[\upsilon_2^r (i-k)\right] - \sin\left[\upsilon_1^r (i-k)\right]}{\pi (i-k)}, & i \ne k \\ \dfrac{\upsilon_2^r - \upsilon_1^r}{\pi}, & i = k \end{cases} \tag{9}$$

Таким образом, применение нового метода вычислений точных значений долей энергии в частотных интервалах позволяет

максимально повысить достоверность принятия решений о принадлежности анализируемого отрезка паузе.

Иными словами предлагаемый подход позволяет оптимизировать процедуру обнаружения пауз, в том смысле, что при заданной вероятности ложной тревоги максимизируется вероятность правильного обнаружения начала звуков речи за счет адекватного учета сосредоточенности их энергий в нескольких узких частотных интервалах.

Результаты вычислительных экспериментов

Экспериментальные исследования проводятся с целью проверки работоспособности алгоритма сжатия речевых данных за счет обнаружения и кодирования пауз на основе сравнения распределений энергии шума и смеси сигнал + шум в заданных частотных интервалах.

В частности, оцениваются вероятности правильного обнаружения пауз (ПОП) и ложного обнаружения пауз (ЛОП).

В основе вычислительных экспериментов по обработке отрезков речевых сигналов лежит разработанный алгоритм сжатия за счёт кодирования пауз.

I. Формируем основную гипотезу H_0: анализируемый отрезок $\vec{x} = (x_1,..., x_n)^T$ длительностью N отсчетов принадлежит паузе в речевых сообщениях, используется решающее правило:

Если

$$S = \max(P_r / P_r^{\Pi}) > h_\alpha, \; r = 1,...., R; \tag{10}$$

то гипотеза отвергается.

Здесь: R – общее количество частотных интервалов;

P_r^{Π} – оценка математического ожидания доли энергии сигнала в паузе в соответствующем частотном интервале;

h_α – порог, удовлетворяющий условию

$$\int_0^{h_\alpha} W(S) dS \ge 1 - \alpha, \tag{11}$$

где: $1-\alpha$ – выбранный уровень правильного обнаружения пауз (α - вероятность ложных тревог); W – функция плотности вероятности решающей функции S;

II. Также как в реальных информационных технологиях, для оценивания $P_r^{П}$ и h_a используется этап обучения. На этапе обучения осуществляется:

1) Вычисление среднего значения энергии сигнала заведомо относящегося к паузе по формуле

$$P_r^{П} = \sum_{k=1}^{N_y} (P_r)_k^{П} / N_y \qquad (12)$$

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.

$(P_r)_k^{П}$ -энергия анализируемого отрезка относящегося к паузе в r-том частотном интервале, вычисляемая по формуле (8).

2) На основе неравенства Чебышева проводится итерация по определению порога, обеспечивающего заданный уровень вероятности ложной тревоги

$$h_a \le \overline{S}_{П} + a * D_{П} / \sqrt{\alpha}. \qquad (13)$$

где: h_a - порог, α - заданный уровень вероятности ложной тревоги, a-коэффициент, значение которого больше двух, а его конкретная оценка проводится в процессе обучения по следующему принципу: $\beta_m = \dfrac{\sum_{k=1}^{N_y} sig(S_k^{П} - h_a^m)}{N_y}$, $sig(x) = 1, x > 0;\ sig(x) = 0, x \le 0.$

Если $\left|\alpha - \beta_m \le \alpha^2\right|$, то $h_a = h_a^m$ и прекратить итерации. В противном случае при $\alpha > \beta_m$ положить $a_{m+1} := (1 - \alpha \times a_m) \times a_m$, если же выполняется неравенство $\alpha < \beta_m$, то положить $a_{m+1} := (1 + \alpha \times a_m) \times a_m$, положить m=m+1 и продолжить итерации.

До этого при использовании обучающей выборки относящихся к паузе данных вычисляются оценки математических ожиданий вида (12). Затем вычисляются оценки математического ожидания и дисперсии решающей функции:

$$\overline{S}_{П} = \sum_{k=1}^{N_y} (S_k^{П}) / N_y, \qquad (14)$$

До этого при использовании обучающей выборки относящихся к паузе данных вычисляются оценки математических ожиданий вида (12). Затем вычисляются оценки математического ожидания и дисперсии решающей функции:

$$\overline{S}_{П} = \sum_{k=1}^{N_y} (S_k^{П}) / N_y, \qquad (14)$$

$$D_{П}^2 = \sum_{k=1}^{N_y} (S_k^{П})^2 / N_y - \overline{S}_{П}^2. \qquad (15)$$

Здесь символ $S_k^{П}$ означает значение решающей функции на k-том анализируемом отрезке, заведомо относящемся к паузе данных.

Для сокращения объема вычислительных работ используются свойства собственных векторов и матриц:

1) Для каждого из частотных интервалов вычисляются матрицы $\mathbf{A}_r$ и соответствующие наборы собственных векторов и чисел

$$\lambda_{kr} \vec{q}_{kr} = \mathbf{A}_r \vec{q}_{kr}, \quad k = 1,2,...,J.$$
$$J = 2\left[\frac{N}{2R}\right] + 2, r = 1,..., R.$$

2) На основе собственных значений матриц $\mathbf{A}_r$ формируется матрица $\mathbf{AA}$

$$AA = \begin{pmatrix} \sqrt{L_1} Q_1^T \\ \sqrt{L_2} Q_2^T \\ \dots \\ \sqrt{L_R} Q_R^T \end{pmatrix}, \qquad (16)$$

где $Q_r = (\vec{q}_{1r},...,\vec{q}_{Jr})$ - матрица собственных векторов, $L_r = diag(\lambda_{1r},...,\lambda_{Jr})$ - диагональная матрица собственных чисел матрицы $\mathbf{A}_r$.

3) Для определения приближенных значений долей энергии отрезка сигнала вычисляется вектор

$$\vec{yy} = AA\vec{x} = \begin{pmatrix} \vec{y}_1 \\ \vec{y}_2 \\ \dots \\ \vec{y}_R \end{pmatrix}, \qquad (17)$$

и суммы квадратов компонент соответствующих подвекторов

$$\hat{P}_r = \sum_{k=1}^{J} (y_{kr})^2 . \qquad (18)$$

$$\vec{y}_r = \sqrt{L_r} Q_r^T \vec{x} , r = 1,\ldots,R ,$$

что является оценкой доли энергии сигнала в r-том частотном интервале отрезка речевого сигнала.

III. Для оценки обоснованности выбора h_α используется этап диагностической проверки:

1) Вычисление оценки вероятности правильного обнаружения паузы ($\alpha_{пол}$).

Вычисление оценки вероятности правильного обнаружения паузы проводился на участке сигнала заведомо относящегося к паузе по формуле

$$\alpha_{пол} = \frac{D_{нел}}{k2 - k1};$$

где $D_{нел}$ - количество значений решающей функции не превышающих порог,

$k1$ - номер отсчета начала паузы,

$k2$ - номер отсчета конца паузы.

2) Вычисление оценки вероятности ложного обнаружения паузы ($\alpha_{лоп}$).

Вычисление оценки вероятности ложного обнаружения паузы проводился на участке сигнала заведомо относящегося к речи по формуле

$$\alpha_{пол} = \frac{D_n}{n2 - n1};$$

где D_n - количество значений решающей функции превышающих порог,

$n1$ - номер отсчета начала речи,

$n2$ - номер отсчета конца речи.

В качестве эмпирических данных были использованы отрезки речевых файлов. Длина анализируемого отрезка выбрана равной N=20, 60, 200, 1000.

Область определения трансформант Фурье дискретных

сигналов $[0,\pi]$ разбивается на R одинаковых частотных интервалов, таких что $\upsilon_{2r} - \upsilon_{1r} = \Delta\upsilon = \text{const}$, причем такие R, что M является целым числом (N=MR).

В ходе экспериментов для всех значений N используется различное разбиение оси частот на R частотных интервалов, а именно:

при N=20 R=2, 5, 10,
при N=60 R=2, 6, 10, 15, 30,
при N=200 R=2, 20, 25, 50, 100,
при N=1000 R=2, 10, 50, 100.

В табл. 1 представлены предварительно вычисленное значение порога при заданном уровне вероятности ложной тревоги, оценка вероятности правильного и ложного обнаружения пауз на участке сигнала в 100000 отсчетов при заданных N и R.

Таблица 1

Оценка вероятности правильного ($\alpha_{пол}$) и ложного ($\alpha_{лоп}$) обнаружения пауз

№ эксперимента	N	R	$\alpha_{пол}$	$\alpha_{лоп}$
1	20	2	0,98781	0,0032
2	20	5	0,98781	0,0026
3	20	10	0,98320	0,0013
4	60	2	0,98261	0,0012
5	60	6	0,98801	0,0006
6	60	10	0,98441	0,0006
7	60	15	0,99101	0,0007
8	60	30	0,98381	0,0007
9	200	2	0,99001	0,0000
10	200	20	0,99601	0,0000
	200	25	0,99201	0,0022
12	200	50	0,99801	0,0000
13	200	100	0,99801	0,0000
14	1000	2	0,99001	0,0000
15	1000	10	0,98001	0,0000
16	1000	50	0,98001	11
17	1000	100	0,98001	0,0000

Для иллюстрации полученных результатов на рис. 1-4 изображены границы пауза/речь и речь/пауза (на рисунках, заранее на слух определяемая граница, отображена вертикальной линией).

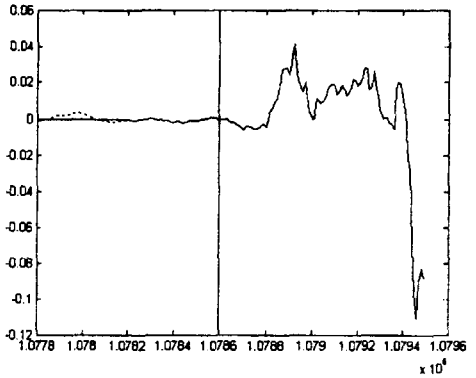

Рис. 1
Граница пауза/звук, определенная при использовании параметров N=60, R=2

При использовании значений параметров N=60, R=2,40 отсчетов паузы определяются как речь, т.к. анализируемый отрезок достаточно велик и захватывает как речь, так и паузу перед речью. Звук при этом не искажается, но ухудшается степень сжатия речи.

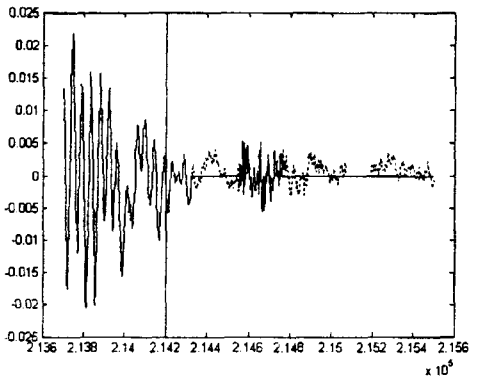

Рис. 2
Граница звук/пауза, определенная при использовании параметров N=60, R=2.

При использовании значений параметров N=60, R=2,180 отсчетов паузы определяются как речь. Звук при этом не искажается, но некоторые короткие участки паузы (например, отсчеты с 214600 по 214800) определяются как речь, что создает «треск» при воспроизведении и ухудшает степень сжатия речи.

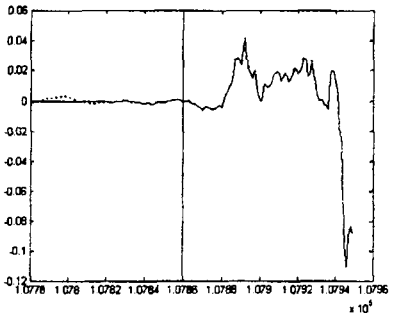

Рис. 3
Граница пауза/звук, определенная при использовании параметров N=60, R=10.

При использовании значений параметров N=60, R=10, 40 отсчетов паузы определяются как речь, т.к. анализируемый отрезок достаточно велик и захватывает как речь, так и паузу перед речью. Звук при этом не искажается, но ухудшается степень сжатия речи.

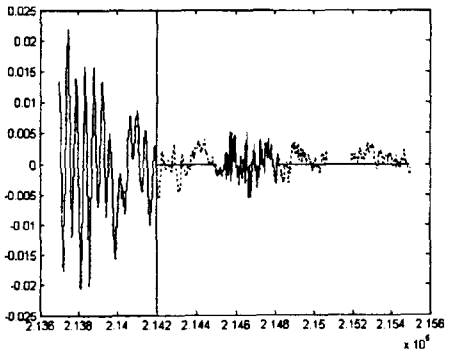

Рис. 4
Граница звук/пауза, определенная при использовании параметров N=60, R=10

При использовании значений параметров N=60, R=10, граница паузы определяется точно, но некоторые короткие участки паузы (например, отсчеты с 214600 по 214800) определяются как речь, что создает «треск» при воспроизведении и ухудшает степень сжатия речи.

Заключение

Предлагаемый алгоритм сжатия речевых данных за счет обнаружения и кодирования пауз на основе сравнения распределений энергии шума и смеси сигнал + шум в заданных частотных интервалах обладает высокой работоспособностью.

При всех использованных сочетаниях N и R вероятность правильного обнаружения пауз не менее 0,98, а ложного обнаружения пауз - не превосходит 0,005. По результатам вычислительных экспериментов рекомендуется использовать длины анализируемых отрезков N=60 при количестве частотных интервалов R=10, т.к. при этом адекватно учитываются узость частотных интервалов, где сосредоточена энергия речевых сигналов, и объем вычислительных работ.

Литература

1. Жиляков Е.Г., Белов С.П. и Прохоренко Е.И. Уменьшение объема битового представления речевых данных на основе нового метода удаления пауз. - Вопросы радиоэлектроники", сер. ЭВТ, 2007, вып. 2, с. 82-92.

2. Жиляков Е.Г., Белов С.П. и Прохоренко Е.И. Вариационные методы частотного анализа звуковых сигналов. – "Труды учебных заведений связи", 2006, вып. 174, с. 163-172.

3. Белов С.П. и др. Способ обнаружения пауз в речевых сигналах и устройство его реализующее. Положит. решение о выдаче Патента на изобретение по заявке №2006138374/09 (041799) от 30.10.2006.

Статья поступила 15.10.2007