ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ
1969 г. Выпуск 36

УДК 534.78:681.142:62-501.2:519.62

СЖАТИЕ ОПИСАНИЯ СИГНАЛА И ЧЛЕНЕНИЕ РЕЧИ НА ФОНЕМЫ

Б.М. Курилов, Б.П. Гаврилко

При проведении настоящей работы авторы руководствовались, в основном, представлениями о модели восприятия речи человеком, изложенными в [1,2] . Основным предположением в работе было то, что в органе слуха с помощью основной мембраны и волосковых клеток кортиева органа происходит разложение входного сигнала $f(t)$ на два сигнала: сигнал мгновенной амплитуды $a(t)$ и сигнал мгновенной частоты $\rho(t)$.

$$f(t) = a(t) \sin\left[\int_0^t \rho(t)\, dt\right]. \qquad (1)$$

Такое разложение квазигармонического сигнала возможно без потери информации лишь в том случае, если на спектр сигнала наложены некоторые ограничения. Рассмотрим эти ограничения.

Заметим, что значения мгновенной амплитуды и мгновенной частоты могут быть измерены не более одного раза за период функции $f(t)$. То есть ограничения, накладываемые на спектр функции $f(t)$, необходимо рассматривать при условии, что $a(t)$ и $\rho(t)$ определяются по дискретным значениям этой функции один раз за период. В работе [3] на основании теорем Котельникова показано, что в этом случае для функции $f(t)$, имеющей спектр частот от ω_1 до ω_2, должно выполняться неравенство $\omega_2 \leq 2\omega_1$. То есть для представления функции $f(t)$ без потери информации в виде двух функций – мгновенной амплитуды

$a(t)$ и мгновенной частоты $\rho(t)$ – ширина спектра исходной функции должна быть не более одной октавы. Если ширина спектра исходного сигнала более одной октавы, его необходимо предварительно пропустить через набор октавных фильтров $\Phi(j)$. Но для того, чтобы в сигнале полностью отсутствовали частотные составляющие ниже ω_1 и выше ω_2, амплитудно-частотные характеристики октавных фильтров должны быть, вообще говоря, П-образными. Однако конечность длительности сигнала накладывает ограничения на добротность фильтров (или на их постоянную времени τ). Известно, что для того, чтобы сигнал полностью раскачал фильтр, постоянная времени фильтра τ должна быть, по крайней мере, не больше длительности сигнала. Квазигармонический сигнал минимальной длительности, который может присутствовать в j-м фильтре, очевидно, равен длительности одного периода верхней частоты этого фильтра $f_2[j]$

$$t_{min}[j] = \frac{1}{f_2[j]} = \frac{1}{2f_1[j]} . \qquad (2)$$

Тогда, учитывая (2), запишем:

$$\tau[j] \le \frac{1}{2f_1[j]} ,$$

или, немного усиливая это требование, что обычно принято делать в теории расчета фильтров, получим:

$$\tau[j] \le \frac{1}{ef_1[j]} . \qquad (3)$$

В силу неравенства (3) и того, что для октавных фильтров

$$\Delta f[j] = f_1[j] , \qquad (4)$$

получим следующее соотношение:

$$\tau[j] \times \Delta f[j] \le \frac{1}{e}$$

или

$$max\,(\tau \cdot \Delta f) = \frac{1}{e} . \qquad (5)$$

Из (5) следует, что максимально допустимая величина произведения $\tau \cdot \Delta f$ для всех октавных фильтров постоянна и равна $1/e$.

Для одиночного октавного резонансного контура имеем:

$$\tau = \frac{Q}{\pi f_0} ; \qquad f_0 = \sqrt{f_1 \cdot f_2} = f_1\sqrt{2} ,$$

где Q – добротность контура.

Тогда, учитывая (5), запишем:

$$max\ Q = \pi f_0 \tau = \pi\sqrt{2} \cdot \tau \cdot \Delta f = \frac{\pi\sqrt{2}}{e} \simeq 1{,}62 .$$

Этот вывод хорошо согласуется с измерениями эквивалентной добротности основной мембраны слухового анализатора (по данным Бекеши, эквивалентная добротность порядка I,5 ÷ 2).

Далее, зная максимально допустимую величину добротности фильтров, максимизируем их частотно-селективные свойства. Под последними будем понимать степень подавления сигналов вне полосы пропускания.

В работе [4] теоретически показано, что максимальными частотно-селективными свойствами при заданной и постоянной для всех фильтров величине обобщенной расстройки α обладает полосовой фильтр, состоящий из цепочки слабо связанных одиночных контуров и имеющий частотную характеристику вида:

$$\sigma^n = \left[1 + \alpha^2\right]^{n/2} , \qquad (6)$$

где $\alpha \simeq 2Q\frac{\Delta f}{f_0}$ – обобщенная расстройка,

n – число одиночных контуров в полосовом фильтре.

Таким образом, исходный сигнал $f(t)$ можно представить в виде:

$$f(t) = \sum_{j=1}^{m} a_j(t) \sin\left[\int_0^t \rho_j(t)\,dt\right] , \qquad (7)$$

где $a_j(t)$ и $\rho_j(t)$ – мгновенная амплитуда и мгновенная частота на выходе полосового октавного фильтра, имеющего частотную характеристику (6) и временную характеристику, удовлетворяющую выражению (5); m – число полосовых октавных фильтров на заданном диапазоне частот исходного сигнала.

Следует отметить, что представление исходного сигнала в виде (7) является п о л н ы м в силу второй и четвертой теорем Котельникова и м и н и м а л ь н ы м в том смысле, что уменьшение числа полос системы фильтров или частоты квантования модулирующих функций в общем случае приведет к потере информации.

Однако, при таком представлении исходного сигнала сжатия

описания этого сигнала не происходит из-за того, что квантование модулирующих функций ведется из расчета максимально возможной частоты этих функций независимо от их истинной мгновенной частоты. Для того, чтобы квантование модулирующих функций велось в зависимости от их мгновенной частоты, преобразуем выражение (7).

Прежде всего заметим, что функция $a_j(t)$ в (7) представляется в виде временной последовательности:

$$a_j(t) = a_j(t_0);\ a_j(t_1);\ \ldots;\ a_j(t_i);\ \ldots,$$

которую можно записать как

$$a_j(t) = a_j(t_0);\ a_j(t_1) + \frac{\Delta a_j(t_1)}{\Delta t_1}\cdot\Delta t_1;\ \ldots \quad (8)$$
$$\ldots;\ a_j(t_{i-1}) + \frac{\Delta a_j(t_i)}{\Delta t_i}\cdot\Delta t_i;\ \ldots.$$

Тогда без потери информации о модулирующей функции $a_j(t)$ можно брать лишь такие дискретные отсчеты этой функции на временной оси, для которых

$$\frac{|\Delta a_j(t_i)|}{\Delta t_i} > 0$$

или, задавшись некоторым уровнем точности ε представления функции $a_j(t)$, такие отсчеты, для которых

$$\frac{|\Delta a_j(t_i)|}{\Delta t_i} > \varepsilon.$$

Тогда

$$a_j(t) = a_j(t_0), t_0;\ a_j(t_\kappa), t_\kappa;\ \ldots$$
$$\ldots;\ a_j(t_e), t_e;\ \ldots. = A_j(t).$$

Проводя аналогичное преобразование для $\rho_j(t)$, запишем (7) в виде:

$$f(t) = \sum_{j=1}^{m} A_j(t)\sin\left[\int_0^t R_j(t)\,dt\right]. \quad (9)$$

Такое представление исходной функции $f(t)$ по-прежнему является полным, а сжатие описания происходит без потери информации (или с допустимыми потерями ε) и зависит от свойств модулирующих функций. Это сжатие будет особенно эффективным, если модулирующие функции сигнала имеют постоянные значения на больших участках.

Далее, для сокращения динамического диапазона представления функций $A_j(t)$ и $R_j(t)$ проведем их логарифмирование,

сместив каждую из функций на +I.

$$\ln[A_j(t)+1] = A_j^*(t);\ \ln[R_j(t)+1] = R_j^*(t).$$

Тогда выражение (9) можем представить в виде:

$$f(t) = \sum_{j=1}^{m}[e^{A_j^*(t)}-1]\sin\left\{\int_0^t[e^{R_j^*(t)}-1]\,dt\right\}. \quad (10)$$

И, наконец, для представления исходной функции $f(t)$ на временной оси в виде сегментов, содержащих или не содержащих сигнал хотя бы в одной из частотных полос, введем функцию сегментации $\Psi(A_j^*, t)$ по следующему закону:

$$\Psi(A_j^*, t) = \begin{cases} 1, \text{ если } A_j^*(t) = \lambda, \\ 0, \text{ если } A_j^*(t) \neq \lambda. \end{cases} \quad (11)$$

Здесь λ – некоторое пороговое значение. Тогда, разбивая исходный сигнал $f(t)$ по временной оси на сегменты, находящиеся между единичными значениями функции $\Psi(A_j^*, t)$, и вводя обозначения

$$B_\xi = \sum_{j=1}^{m}[e^{A_j^*(t)}-1]\sin\left\{\int_0^t[e^{R_j^*(t)}-1]\,dt\right\}\Big|_{t-t_\alpha}^{t-t_\beta}$$
$$(\xi = 1,2,3,\ldots),$$

преобразуем (10) к виду:

$$f(t) = B_1, B_2, B_3, \ldots, B_\xi, \ldots \quad (12)$$

Здесь следует отметить, что положение сегментов B_ξ на временной оси будет зависеть как от свойств самого сигнала, так и от значения параметров системы $f_1[j]$, $f_2[j]$, $\tau[j]$, $G[j]$, m и λ, причем между параметрами $f_1[j]$, $f_2[j]$, $\tau[j]$, $G[j]$ и m выше получена взаимосвязь исходя из критерия максимума частотно-селективных свойств системы и минимума потери информации об исходной функции $f(t)$. Таким образом, свободными параметрами системы остались $f_1[1]$ и λ.

Цель настоящей работы заключается в том, чтобы, варьируя свободные параметры системы описания речевого сигнала, получить сегментные отрезки B_ξ, максимально приближающиеся к фонемам.

Для достижения этой цели был проведен эксперимент с речевыми сигналами на ЭВМ "БЭСМ-6".

Речевой сигнал вводился в ЭВМ через 9-разрядный аналого-цифровой преобразователь с частотой квантования 20 кгц и затем пропускался через гребенку цифровых математических фильтров [5] . На выходе фильтров вычислялись значения функций $A_j^*(t)$ и $R_j^*(t)$ и выдавались на печать в виде графиков. По полученным на ЭВМ "БЭСМ-6" значениям функций вручную вычислялось значение функции сегментации $\Psi(A_j^*, t)$. Функция $R_j^*(t)$ для членения не использовалась.

В качестве речевого материала было взято 23 русских слова в исполнении трех дикторов (двое мужчин и одна женщина) , десять английских слов для одного диктора (мужской голос) и десять японских слов для одного диктора (женский голос). Общее число фонем составляло 410. В эксперименте варьировались не только значения свободных параметров системы f_1 [I] и λ , но и значения всех остальных параметров. Из соображений не принципиального характера были взяты математические фильтры, аппроксимированные по Баттерворту [5] . Частотный диапазон речи разбивался на полосы от 3 до 15. Временные характеристики фильтров варьировались для значения произведения $\tau \cdot \Delta f$ от I/2 до I/I0; значение λ – во всем динамическом диапазоне. Результаты членения речевого потока на фонемы оценивались методом экспертной оценки (невыделение фонемы "Р" не считалось ошибкой). Наилучший результат членения 99,2% (четыре ошибки) был получен при значениях параметров (см. табл.I): f_1[I], равном 75 гц, λ , равном I,2 от среднего уровня шума на участке "молчания", и остальных параметров системы, близких к расчетным.

В заключение следует отметить, что экспертная оценка результатов членения не является, вообще говоря, объективной ; действительно, объективной оценкой качества членения может служить надежность распознавания вычлененных фонем.

Таблица I

№ фильтра (j)	Граничные частоты f_1 и f_2	Полоса пропускания Δf	Постоянная времени τ
I.	75 ÷ I50 гц	75 гц	7 мсек
2.	I50 ÷ 300 гц	I50 гц	3,5 мсек
3.	300 ÷ 600 гц	300 гц	I,75 мсек
4.	600 ÷ I200 гц	600 гц	0,87 мсек
5	I200 ÷ 2400 гц	I200 гц	0,43 мсек
6	2400 ÷ 4800 гц	2400 гц	0,22 мсек
7	4800 ÷ 9600 гц	4800 гц	0,II мсек

ЛИТЕРАТУРА

I. Л.В. БОНДАРКО и др. Модель восприятия речи человеком. Изд-во "Наука", Новосибирск, I968 г.
2. Распознавание слуховых образов. Под ред. Н.Г. ЗАГОРУЙКО и Г.Я. Волошина. Изд-во "Наука" , Новосибирск, I966.
3. И.Т. ТУРБОВИЧ, О.А. ПЕТРОВ. Об одном методе полного описания одномерных образов совокупностью простых функций (применительно к речевым сигналам).-В сб. "Опознавание образов". Теория передачи информации. М., Изд-во "Наука", I965, стр. 25.
4. В.С. ПЕРЕБЕРЗЕВ-ОРЛОВ. Об одном способе улучшения качества радиоприема. – В сб. "Опознавание образов", Теория передачи информации. Изд-во "Наука", М., I965, стр. I29.
5. В.С. ЛОЗОВСКИЙ и др. Отчет по теме "Ф-7I" НПО "Факел", Новосибирск, I968.

Поступила в редакцию
I0.I.I969г.