УДК 510.62

В. Д. НАГОРНЫЙ, Г. А. ПРАСОЛ,
С. Ю. ШАБАНОВ-КУШНАРЕНКО

О НОРМАЛИЗАЦИИ СЛИТНОЙ РЕЧИ ПО СРЕДНЕМУ УРОВНЮ ЗВУКОВОГО ДАВЛЕНИЯ

Известно, что орган слуха человека нечувствителен к медленным (инфразвуковым) колебаниям давления воздуха. Такие колебания возникают постоянно и являются серьезной помехой при автоматическом распознавании слитной речи. Наше ухо, по-видимому, каким-то образом стабилизирует среднее значение диаграммы, не меняя ее формы. Этим достигается нормализация слитной речи по среднему уровню звукового давления. В статье рассматривается математическая модель такой нормализации и предлагается способ ее технической реализации.

Пусть $x(t)$ — исходная, а $y(t)$ — нормализованная по среднему уровню диаграмма звукового давления; t — время. Простейшее преобразование, обеспечивающее нормализацию среднего уровня функции $x(t)$ имеет вид

$$T\frac{dy}{dt}+y=T\frac{dx}{dt}. \qquad (1)$$

Здесь T — постоянная времени, с которой осуществляется нормализация. Оператор $y(t)=F(x(t))$, задаваемый дифференциальным уравнением (1), примем в качестве математической модели нормализации среднего уровня диаграммы звукового сигнала органом слуха человека. Пример действия оператора F показан на рис. 1. На диаграмме *а* имеется скачок среднего уровня звукового давления $x(t)$. Возникшее отклонение среднего

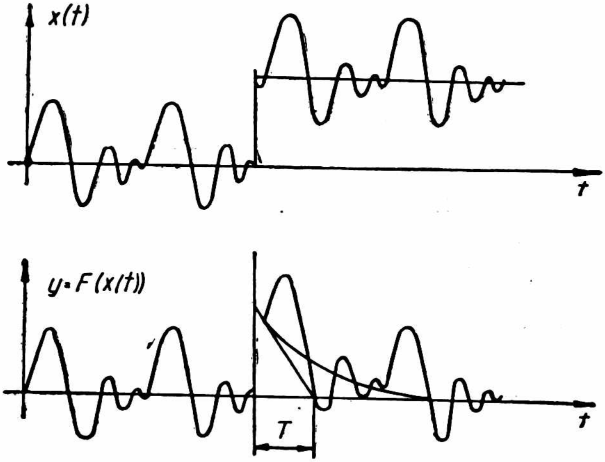

Рис. 1. Пример действия нормализатора среднего уровня диаграммы звукового давления

уровня от нулевого значения постепенно исчезает в нормализованном сигнале $y(t)$, причем тем скорее, чем меньше постоянная времени T (диаграмма *б*). Вопреки рисунку, постоянная времени T должна иметь гораздо большее значение, иначе в процессе нормализации исказится форма акустической диаграммы. Для диаграмм звукового давления речевых сообщений характерно медленное изменение среднего уровня. В этом случае при правильном выборе параметра T нормализатор будет эффективно стабилизировать среднее значение диаграммы на нулевом уровне. Скачкообразное изменение среднего уровня на диаграмме рис. 1 принято чисто условно, только для того, чтобы пояснить принцип действия модели нормализации и роль ее параметра T.

Чтобы осуществить экспериментальную проверку модели нормализации, нужно предварительно создать аппаратуру, формирующую нормированный сигнал $y(t)$ по заданному сигналу $x(t)$. В этом случае, предъявляя испытуемому поочередно звуки $x(t)$ и $y(t)$, можно выяснить, звучат они для него одинаково или нет. В случае положительного ответа можно экспериментально определить наименьшее (критическое) значение $T_{кр}$ постоянной времени T, обеспечивающее равенство звучания сигналов $x(t)$ и $y(t)$. Схема экспериментальной установки изображена на рис. 2. Звук $x(t)$, задаваемый экспериментатором, от микрофона *1* или иного источника электрического сигнала, повторяющего диаграмму звукового давления, поступает на вход нормализатора уровня 2. В ответ нормализатор уровня на своем выходе формирует в соответствии с моделью F колебание электрического потенциала $y(t)$. С помощью ключа *3* и телефона *4* звуки $x(t)$ и $y(t)$ поочередно предъявляются испытуемому, который на слух устанавливает факт их равенства или неравенства.

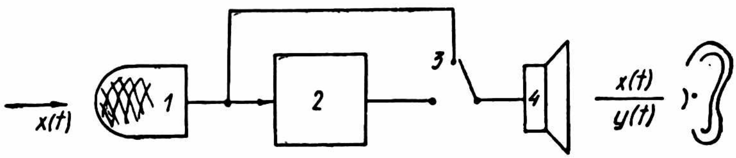

Рис. 2. Схема экспериментальной установки

Нормализатор звука будем строить из цифровых элементов смешанного типа, предназначенных для обработки чисел, представленных и параллельными, и последовательными кодами. Принцип построения таких элементов описан в работе [1, с. 112÷121]. Для создания нормализатора уровня потребуется элемент, умножающий n-разрядный двоичный код числа x на коэффициент 2^p, где p — фиксированное целое число. Если p — положительное, элемент сдвигает код числа x на p разрядов вправо, если отрицательное — влево. В результате получаем двоичный код числа $y=2^p x$.

На рис. 3, *а*, показана схема устройства, умножающего число $x=0,\ x_1,\ x_2 \ldots x_n$ на 2^p при отрицательном p, на рис. 3, *б* — при положительном. Символы $x_1,\ x_2,\ \ldots,\ x_n$ обозначают двоичные разряды дробной части кода числа x. Значения переменной x должны находиться в пределах $0<x<1$. На рис. 3, *в* изображено условное обозначение описываемого элемента.

На рис. 4, *а* дано обозначение суммирующего элемента, складывающего n-разрядные коды чисел x и y и формирующего результат снова в виде n-разрядного двоичного кода суммы z. Для правильной работы сумматора необходимо, чтобы и слагаемые, и сумма находились в пределах от 0 до 1. На рис. 4, *б* представлено условное обозначение преобразователя n-разрядного двоичного кода числа x в последовательный частотно-импульсный код того же числа. Принцип действия преобразователя описан в работе [1, с. 118, 150]. На рис. 4, *в* приведено обозначение формирователя n-разрядного двоичного кода интеграла z разности чисел x и y, выполняющего операцию

$$z = z_0 + c\int_0^t (x-y)\,dt. \tag{2}$$

Код ε_z числа $z\,(0 < z < 1)$ формируется в m-разрядном (если $m > n$) или n-разрядном (если $m \leqslant n$) двоичном регистре интегратора. В начальный момент дискретного времени ($i = 0$) в регистр заносится код ε_{z_0}. Числа x и y поступают в ин-

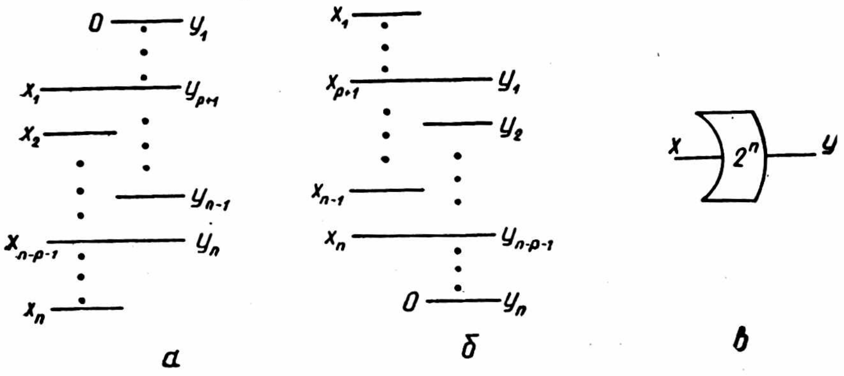

Рис. 3. Логическая схема множительного устройства

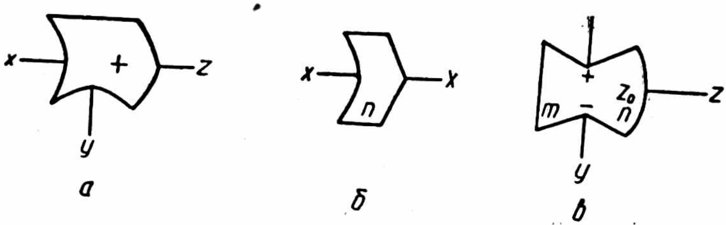

Рис. 4. Сумматор n-разрядных кодов

тегратор в форме последовательных кодов δ_x и δ_y. Если в i-й момент дискретного времени $\delta_x(i) = 1$ и $\delta_{(y)}(i) = 0$, то к коду ε_z в момент i добавляется число 2^{-m}, т. е. единица m-го разряда:

$$\varepsilon_z(i+1) = \varepsilon_z(i) + 2^{-m}. \tag{3}$$

Если $\delta_x(i) = 0$ и $\delta_y(i) = 1$, то вычитается число 2^{-m}:

$$\varepsilon_z(i+1) = \varepsilon_z(i) - 2^{-m}. \tag{4}$$

Если же $\delta_x(i) = \delta_y(i)$, то значение кода ε_z в момент i не меняется:

$$\varepsilon_z(i+1) = \varepsilon_z(i). \tag{5}$$

Найдем формулу для определения масштабного множителя c интегратора. Это можно сделать, положив в равенстве (2) $x(t) \equiv 1$, $y(t) \equiv 0$, $z_0 \equiv 0$. Тогда

$$z(t) = ct. \tag{6}$$

С другой стороны, согласно только что описанному принципу действия интегратора имеем:

$$\varepsilon_z(p) = \varepsilon_{z_0} + \sum_{i=1}^{p} \frac{\delta_x(i) - \delta_y(i)}{2^m} \tag{7}$$

Поскольку в данном случае $\varepsilon_{z_0} = 0$, $\delta_x(i) \equiv 1$, $\delta_y \equiv 0$, формула (7) запишется в виде

$$\varepsilon_z(p) = \frac{p}{2^m}. \tag{8}$$

Полагая, что момент дискретного времени p соответствует моменту t физического времени, имеем

$$z(t) = \varepsilon_z(p). \tag{9}$$

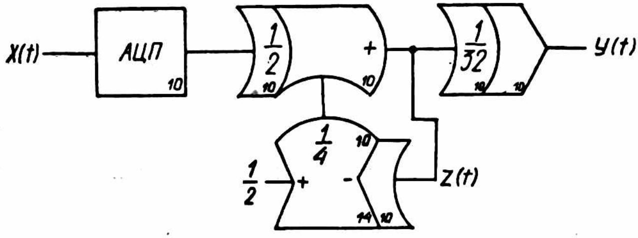

Рис. 5. Схема устройства для нормализации звуковой диаграммы

Приравнивая правые части равенств (6) и (8), получаем искомую формулу

$$c = \frac{1}{2^m T_T}, \tag{10}$$

где T_T — тактовый интеграл, равный

$$T_T = \frac{t}{p}. \tag{11}$$

На рис. 5 изображена схема, составленная из только что описанных элементов. Она реализует уравнение (1). Символами X, Y, Z обозначены машинные переменные. Переменные X, Y связаны с переменными x, y уравнения (1) соотношениями

$$x = 2X - 1; \tag{12}$$

$$y = 128Y - 2. \tag{13}$$

Полагаем, что сигналы x и y изменяются в пределах от -1 до $+1$, тогда, согласно (12) и (13), переменная X будет изменяться в пределах от 0 до 1, а переменная Y — в пределах от 1/128 до 3/128. Диаграмма звука в виде электрического напряжения $X(t)$, меняющегося во времени t, поступает на аналого-цифровой преобразователь. Последний формирует с частотой

20 кгц 10-разрядный параметрический код, изменяющийся в пределах от 0 до 1 и копирующий исходное колебание с точностью около 0,1 %.

Далее сигнал поступает на схему, состоящую из шести элементов, работающих на тактовой частоте 5 МГц. Сигнал умножается на 1/2, после чего диапазон его изменения сужается до пределов 0—0,5. Элемент суммирования прибавляет к сигналу в начальный момент времени величину 1/4, в результате чего на его выходе формируется сигнал $Z(t)$, изменяющийся в диапазоне 0,25—0,75. Если среднее значение сигнала $Z(t)$ постоянно поддерживается на уровне 0,5 (это соответствует отсутствию постоянной составляющей у сигнала $x(t)$), то второе слагаемое, поступающее на нижний вход суммирующего элемента, постоянно сохраняет значение 1/4. Если же среднее значение сигнала $x(t)$ смещается от нулевого уровня в сторону положительных или отрицательных величин, то в действие вступают преобразователь с интегрирующим элементом, которые, будучи включены в цепь обратной связи, регулируют величину второго слагаемого сумматора таким образом, чтобы среднее значение сигнала $Z(t)$ постоянно сохранялось на уровне 0,5.

Сигнал $Y(t)$ формируется путем умножения функции $Z(t)$ на число 1/32 и преобразования результата в последовательный двоичный код. Схеме нормализатора уровня соответствует следующая система машинных уравнений:

$$\frac{1}{2}X + \frac{1}{4} + c\int_0^t \left(\frac{1}{2} - Z\right) dt = Z, \tag{14}$$

$$Y = \frac{1}{32}Z. \tag{15}$$

Дифференцируя по t левые и правые части равенства (14) и (15), имеем

$$\frac{1}{2}\frac{dX}{dt} + c\left(\frac{1}{2} - Z\right) = \frac{dZ}{dt}; \tag{16}$$

$$\frac{dY}{dt} = \frac{1}{32}\frac{dZ}{dt}. \tag{17}$$

Заменяя в (16) Z и $\frac{dZ}{dt}$ по (15) и (17), получаем уравнение, связывающее машинные переменные X и Y:

$$\frac{1}{2}\frac{dX}{dt} + c\left(\frac{1}{2} - 32Y\right) = 32\frac{dY}{dt}. \tag{18}$$

Переходя в последнем уравнении от машинных к моделируемым переменным x и y с помощью зависимостей (12) и (13), приходим к уравнению

$$\frac{dy}{dt} + cy = \frac{dx}{dt}. \tag{19}$$

Сравнивая уравнения (1) и (19) и используя (10), находим выражение для постоянной времени нормализатора уровня:

$$T = \frac{1}{c} = 2^m T_T. \tag{20}$$

Тактовой частоте 5 МГц соответствует тактовый интервал $T_T = 0{,}2$ мкс. Постоянную времени принимаем равной $T = 1$ с. Отсюда число разрядов интегратора

$$m = \frac{\lg \frac{T}{T_т}}{\lg 2} = \frac{\lg \frac{1}{0{,}2 \cdot 10^{-6}}}{0{,}3} = \frac{6 \lg 5}{0{,}3} = \frac{6 \cdot 0{,}7}{0{,}3} = 14.$$

В остальных решающих элементах нормализатора уровня используем 10-разрядные двоичные коды. Максимально возможный интервал между соседними импульсами в последовательном коде сигнала $Y(t)$ составляет $T_{\max} = 128 \cdot T_T = 128 \cdot 0{,}2 = 25{,}6$ мкс, что менее 50 мкс. Этим гарантируется отсутствие помех в виде тональных звуков в формируемом нормализатором уровня звуковом сигнале $y(t)$ [2, с. 263]. Минимально возможный интервал равен

$$T_{\min} = \frac{128}{3} T_T = \frac{128 \cdot 0{,}2}{3} = 8{,}5 \text{ мкс}.$$

Наибольшая относительная погрешность интервалов между импульсами составляет

$$\frac{T_T}{T_{\min}} 100 = \frac{0{,}2}{8{,}5} \cdot 100 = 2{,}4\ \%,$$

что не выходит за допустимый порог 3 %. Этим гарантируется отсутствие шумов квантования в звуке $y(t)$, формируемом нормализатором уровня [2, с. 252].

Список литературы: 1. *Маленченко З. Ю.* Математические модели некоторых функций слухового восприятия и их приложения в технике обработки слитной речи: Дис.... канд. техн. наук.— Х., 1984.— 131 с. 2. *Бондаренко М. Ф.* Математические модели морфологических и фонетических отношений и их применение для автоматизации обработки речевых сообщений: Дис. ... д-ра техн. наук.— Х., 1984.— 296 с.

Поступила в редколлегию 04.12.84.