

СПЕКТРАЛЬНЫЙ АНАЛИЗ РЕЧЕВЫХ СИГНАЛОВ
С ПОМОЩЬЮ ЭВМ

Г.Я. Волошин

Методика обмена устной информацией между человеком и вычислительными системами предусматривает использование сложного комплекса параметров речевого сигнала [1]. В настоящее время можно считать установленным, что спектральные параметры являются решающими для автоматического распознавания сигналов речи. Применение ЭВМ для спектрального анализа целесообразно в тех случаях, когда:

а) время, затрачиваемое на расчет спектральной функции, мало по сравнению с общим временем анализа звуков речи на ЭВМ;
б) необходимо использование фильтров с такими временными и частотными характеристиками, которые недостижимы в реальных резонансных системах с потерями;

в) требуется частое изменение параметров фильтра или подбор фильтра с оптимальными параметрами.
При спектральном анализе речевых сигналов на ЭВМ возникает вопрос о методе вычисления спектра и о погрешностях, связанных с особенностями этого метода.

Можно выделить три основных метода расчета энергетического спектра.

1. Непосредственный частотный анализ с помощью преобразования Фурье:

$$S_t(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-j\omega t} dt, \\ \Phi_t^2(\omega) = S_t(\omega) \cdot S_t^*(\omega),$$

где $S_t(\omega)$ - мгновенный комплексный спектр сигнала,

$f(t)$ - исследуемый сигнал,

$\Phi_t^2(\omega)$ - энергетический спектр,

$S_t^*(\omega)$ - спектральная функция, сопряженная с $S_t(\omega)$.

2. Линейное преобразование исходной функции $f(t)$

$$F(t) = \int_{-\infty}^{+\infty} f(t+\theta) \cdot \varphi(\theta) d\theta,$$

где $F(t)$ - преобразованная функция,

$\varphi(\theta)$ - преобразующая функция.

Эта операция эквивалентна пропусканию сигнала через фильтр с параметрами, определяемыми преобразующей функцией [2].

3. Корреляционно-спектральный анализ.

В этом случае расчет производится по формулам:

$$m = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(t) \cdot dt,$$

$$\hat{f}(t) = f(t) - m,$$

$$B(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \hat{f}(t) \cdot \hat{f}(t+\tau) \cdot dt, \quad (1)$$

$$G(\omega) = \lim_{T \rightarrow \infty} \frac{2}{T} \int_0^T B(\tau) \cos \omega \tau d\tau,$$

где m - математическое ожидание,

$\hat{f}(t)$ - центрированная функция,

$B(\tau)$ - функция корреляции,

$G(\omega)$ - спектральная функция.

Естественно предположить, что точность расчета спектральных функций по любому из перечисленных методов примерно одинакова, поэтому ограничимся рассмотрением только корреляционно-спектрального анализа.

В общем случае речевой сигнал представляет собой нестационарный случайный процесс. Для изучения динамики спектральных параметров необходимо получать мгновенные спектры, т.е. спектры сигнала, время наблюдения которого ограничивается интервалом 20 мсек и меньше. Тогда нестационарный случайный сигнал, действующий в промежутке времени T , может быть сопоставлен с квазистационарным сигналом, функция корреляции и

спектральная функция которого совпадают соответственно со средней функцией корреляции и спектром средней мощности нестационарного сигнала [3, 4].

В связи с тем, что время наблюдения T ограничено, при анализе возникают погрешности. Случай, когда $T \gg \tau_{эфф}$, где $\tau_{эфф}$ - интервал корреляции, достаточно освещен в литературе и в статье не рассматривается.

При $T \leq 20$ мсек длительность квазистационарной части голосовых (в частности, гласных) звуков, особенно при полном стиле произношения, много больше T , а тем более периода колебаний основного тона, который не превышает 10 мсек. Это значит, что сигнал практически оказывается гармоническим [5], а потому условие $T \gg \tau_{эфф}$ выполнить невозможно, а это влечет за собой специфические погрешности, значительно превосходящие погрешности при $T \gg \tau_{эфф}$. Поэтому для общей оценки точности корреляционно-спектрального анализа сигналов речи будем рассматривать худший случай, т.е. квазистационарные участки голосовых звуков.

Учитывая, что период колебаний основного тона много меньше длительности квазистационарного участка голосового звука, представим речевой сигнал в идеализированном виде:

$$f(t) = A_0 + \sum_{i=1}^n A_i \sin(\omega_i t + \varphi_i), \quad (2)$$

где A_0 - постоянная составляющая сигнала,
 A_i, ω_i, φ_i - амплитуда, частота и фаза i -й составляющей, соответственно,

$$\omega_i = i \omega_{от},$$

$$i = 1, 2, 3, \dots$$

n - число гармоник основного тона,

$\omega_{от}$ - круговая частота основного тона.

Определим погрешности, возникающие на каждом этапе расчета спектральной функции.

1. Погрешность нахождения постоянной составляющей (математического ожидания). Пусть

$$m \approx \frac{1}{T} \int_0^T f(t) \cdot dt = A_0 + \sum_{i=1}^n \frac{A_i}{\omega_i} [\cos(\omega_i T + \varphi_i) - \cos \varphi_i]. \quad (3)$$

При $T \rightarrow \infty$ второе слагаемое стремится к нулю. Оно составля-

ет абсолютную погрешность δ_i , обусловленную конечностью времени наблюдения T . Итак,

$$\delta_i = \frac{1}{T} \sum_{i=1}^n \frac{A_i}{\omega_i} [\cos(\omega_i T + \varphi_i) - \cos \varphi_i]. \quad (4)$$

Ошибка будет максимальной в случае назальных смычных, так как энергия этих звуков сосредоточена в области нижних частот. Для приближенной оценки погрешности заменим ω_i на ω_{cp} , определяем соотношением

$$\int_0^{\omega_{cp}} G(\omega) d\omega = \int_0^{\infty} G(\omega) \cdot d\omega. \quad (4a)$$

Тогда

$$\delta_i = \frac{1}{T \cdot \omega_{cp}} \sum_{i=1}^n A_i [\cos(\omega_i T + \varphi_i) - \cos \varphi_i].$$

Выражение под знаком $\sum$ в худшем случае может принимать значение полного размаха (т.е. огибающей) колебаний:

$$\delta_i \leq \frac{1}{T \omega_{cp}} A_{огиб}.$$

Относительная погрешность

$$\epsilon_i \leq \frac{1}{T \cdot \omega_{cp}} = \frac{T_{cp}}{2\pi \cdot T}.$$

Для квазистационарных участков назальных смычных $\omega_{cp} \approx 2\pi(400 \frac{1}{сек})$. Отсюда в худшем случае

$$\epsilon_i \leq \frac{4 \cdot 10^{-4}}{T}, \quad (5)$$

где T - время наблюдения в секундах.

2. Погрешность расчета функции корреляции. Существует несколько способов приближенного вычисления функции корреляции [6]. Приведем некоторые из них:

а) Пусть $T = T_{набл.} - \tau_{max} = const,$

тогда

$$B(\tau) \approx \frac{1}{T} \int_0^T f(t) \cdot f(t+\tau) \cdot dt. \quad (6)$$

В этом случае $0 \leq t + \tau \leq T_{набл.}$

б) Время интегрирования T берется в зависимости от $T = T_{набл.} - \tau$, где $\tau > 0$,

$$B(\tau) \approx \frac{1}{T_{набл.} - \tau} \int_0^{T_{набл.} - \tau} f(t) \cdot f(t+\tau) dt. \quad (7)$$

в) Время интегрирования T берется равным $T_{НАБЛ.} - \tau$, а величина интеграла относится к $T_{НАБЛ.}$.

$$B(\tau) \approx \frac{1}{T_{НАБЛ.}} \int_0^{T_{НАБЛ.}-\tau} f(t) \cdot f(t+\tau) dt. \quad (8)$$

Способ интегрирования для нахождения $B(\tau)$ должен быть аналогичным способу нахождения математического ожидания функции:

$$B(\tau) \approx \frac{1}{T} \int_0^T f(t) \cdot f(t+\tau) \cdot dt.$$

В противном случае не являются дополнительные погрешности, на которых подробно останавливаться не будем (их можно определить, подставив (3) в (7) и (8)).

Допустим, что математическое ожидание определено с достаточно малой погрешностью, которой можно пренебречь. Тогда

$$m = A_0 \text{ и } f(t) = \sum_{i=1}^n A_i \sin(\omega_i t + \varphi_i),$$

$$\begin{aligned} B(\tau) &\approx \frac{1}{T} \int_0^T \sum_{i=1}^n \left\{ A_i \sin(\omega_i t + \varphi_i) \cdot A_i \sin[\omega_i(t+\tau) + \varphi_i] \right\} dt + \\ &+ \frac{1}{T} \int_0^T \sum_{i,j=1}^n \left\{ A_i \sin(\omega_i t + \varphi_i) \cdot A_j \sin[\omega_j(t+\tau) + \varphi_j] \right\} dt + \frac{1}{T} \int_0^T \sum_{i,j=1}^n \left\{ A_j \sin(\omega_j t + \varphi_j) \cdot \right. \\ &\left. \times A_i \sin[\omega_i(t+\tau) + \varphi_i] \right\} dt = \sum_{i=1}^n B_i(\tau) + \sum_{i,j=1}^n B_{ij}(\tau) + \sum_{i,j=1}^n B_{ji}(\tau). \quad (9) \end{aligned}$$

Здесь $i \neq j$;

$B_i(\tau)$ - автокорреляционная функция i -ой составляющей;

$B_{ij}(\tau)$ и $B_{ji}(\tau)$ - взаимные корреляционные функции между i -ой и j -ой составляющими.

Рассмотрим первое слагаемое в выражении (9):

$$\sum_{i=1}^n B_i(\tau) = \frac{1}{T} \int_0^T \sum_{i=1}^n A_i \sin(\omega_i t + \varphi_i) \cdot A_i \sin[\omega_i(t+\tau) + \varphi_i] dt = \quad (10)$$

$$= \sum_{i=1}^n \left\{ \frac{A_i^2}{2} \cos \omega_i \tau \cdot \frac{A_i^2}{\omega_i T} \cdot \frac{\sin \omega_i T \cdot \cos[\omega_i(T+\tau) + 2\varphi_i]}{2} \right\},$$

при $T \rightarrow \infty$

$$\sum_{i=1}^n B_i(\tau) \rightarrow \sum_{i=1}^n \frac{A_i^2}{2} \cos \omega_i \tau.$$

Второе слагаемое в выражении (10) является погрешностью $|\delta_2|$,

обусловленной конечностью времени наблюдения T . Для i -ой составляющей

$$\delta_2 = \frac{A_i^2}{\omega_i T} \frac{\sin \omega_i T \cdot \cos[\omega_i(T+\tau) + 2\varphi_i]}{2}. \quad (11)$$

В худшем случае при $\omega_i = \omega_{max} = \omega_{от.}$, где $\omega_{от.}$ - круговая частота основного тона, выражение для отбрасываемой δ запишется в виде

$$|\delta_2| \leq \frac{A_i^2}{2 \omega_{от.} T}.$$

Относительная погрешность ε_2 (по сравнению с амплитудой колебаний первого слагаемого из выражения (10)) равна

$$|\varepsilon_2| \leq \frac{1}{\omega_{от.} T}.$$

Так как $\omega_{от.} = \frac{2\pi}{T_{от.}}$,

где $T_{от.}$ - период основного тона, то

$$|\varepsilon_2| \leq \frac{0,16 T_{от.}}{T}, \quad (12)$$

где T - время наблюдения.

Рассмотрим второе и третье слагаемые в выражении (9):

$$\begin{aligned} \sum_{i,j=1}^n B_{ij}(\tau) &= \frac{1}{T} \sum_{i,j=1}^n A_i A_j \left\{ \frac{1 - \cos(\omega_i + \omega_j)T}{2(\omega_i + \omega_j)} + \frac{1 - \cos(\omega_i - \omega_j)T}{2(\omega_i - \omega_j)} \right\} (A_i \sin \omega_j \tau + \\ &+ C \cos \omega_j \tau) + \left[\frac{\sin(\omega_i + \omega_j)T}{2(\omega_i + \omega_j)} + \frac{\sin(\omega_i - \omega_j)T}{2(\omega_i - \omega_j)} \right] (B \cos \omega_j \tau + D \sin \omega_j \tau) + \\ &+ \left[\frac{\sin(\omega_i - \omega_j)T}{2(\omega_i - \omega_j)} - \frac{\sin(\omega_i + \omega_j)T}{2(\omega_i + \omega_j)} \right] (A \cos \omega_j \tau + D \sin \omega_j \tau) + \\ &+ \left[\frac{1 - \cos(\omega_i + \omega_j)T}{2(\omega_i + \omega_j)} - \frac{1 - \cos(\omega_i - \omega_j)T}{2(\omega_i - \omega_j)} \right] (D \cos \omega_j \tau - B \sin \omega_j \tau) \}, \quad (13) \end{aligned}$$

где

$i \neq j$;

$$A = \cos \varphi_i \cdot \cos \varphi_j;$$

$$B = \cos\left(\frac{\pi}{2} - \varphi_i\right) \cos\left(\frac{\pi}{2} - \varphi_j\right);$$

$$C = \cos \varphi_i \cdot \cos\left(\frac{\pi}{2} - \varphi_j\right);$$

$$D = \cos\left(\frac{\pi}{2} - \varphi_i\right) \cdot \cos \varphi_j.$$

Выражение для $\sum_{i,j=1}^n B_{ij}(\tau)$ будет аналогичным выражению (I3), если поменять местами индексы i и j .

Как видно из (I3), с увеличением времени наблюдения T коррелятивные связи между гармоническими составляющими затухают. При $T \rightarrow \infty$ автокорреляционная функция суммы гармонических колебаний равна сумме автокорреляционных функций составляющих. При ограниченном времени наблюдения имеют место взаимно корреляционные связи между гармоническими составляющими, вносящие искажения в функцию автокорреляции сложного периодического сигнала. Для количественной оценки этих искажений рассмотрим один из частных случаев.

Пусть в выражении (I3) $A = 1$. Тогда $B = C = D = 0$ и

$$\sum_{i,j=1}^n B_{ij}(\tau) = \sum_{i,j=1}^n [B_{ij}(\tau) + B_{ji}(\tau)] = \sum_{i,j=1}^n \frac{A_i A_j}{T} \left\{ \frac{1 - \cos(\omega_i + \omega_j)T}{2(\omega_i + \omega_j)} + \frac{1 - \cos(\omega_i - \omega_j)T}{2(\omega_i - \omega_j)} \right\} \times \sin \omega_i \tau + \left[\frac{\sin(\omega_i - \omega_j)T}{2(\omega_i - \omega_j)} - \frac{\sin(\omega_i + \omega_j)T}{2(\omega_i + \omega_j)} \right] (\cos \omega_i \tau + \cos \omega_j \tau) + \left[\frac{-\cos(\omega_i + \omega_j)T + 1}{2(\omega_i + \omega_j)} - \frac{1 - \cos(\omega_i - \omega_j)T}{2(\omega_i - \omega_j)} \right] \sin \omega_j \tau \quad (I4)$$

где $i \neq j$.

В процессе суммирования перебираются все i и j , т.е. рассматриваются случаи $i > j$ и $i < j$. В результате четные функции, которые делятся на величину 2 ($\omega_i - \omega_j$), скомпенсируются, и выражение (I4) примет вид:

$$\sum_{i,j=1}^n B_{ij}(\tau) = \sum_{i,j=1}^n \frac{A_i A_j}{T} \left\{ \frac{1 - \cos(\omega_i + \omega_j)T}{2(\omega_i + \omega_j)} (\sin \omega_i \tau + \sin \omega_j \tau) + \left[\frac{\sin(\omega_i - \omega_j)T}{2(\omega_i - \omega_j)} - \frac{\sin(\omega_i + \omega_j)T}{2(\omega_i + \omega_j)} \right] (\cos \omega_i \tau + \cos \omega_j \tau) \right\}, \quad (I5)$$

где $i \neq j$.

Это выражение описывает затухающую (с увеличением T) колебательную функцию.

Для определенности положим $A_i = A_j$. Так как для речевых сигналов $n \gg 1$, то можно считать, что ω_i и ω_j проходят одни и те же значения, т.е. можно их считать равными для всех членов, за исключением члена

$$\frac{\sin(\omega_i - \omega_j)T}{2(\omega_i - \omega_j)}$$

Тогда

$$\sum_{i,j=1}^n B_{ij}(\tau) \approx \sum_{i,j=1}^n \frac{2A_i^2}{T} \left\{ \frac{1 - \cos 2\omega_i T}{2\omega_i} \sin \omega_i T + \left[\frac{\sin(\omega_i - \omega_j)T}{2(\omega_i - \omega_j)} - \frac{\sin 2\omega_i T}{2\omega_i} \right] \cos \omega_i \tau \right\}.$$

Чтобы оценить верхний предел погрешности δ_j , допустим, что

$$\sin \omega_i T = \cos \omega_i T \approx 0,7; \quad 1 - \cos 2\omega_i T = 2;$$

$$\sin(\omega_i - \omega_j)T = 1; \quad \sin 2\omega_i T = 0,$$

так как $\cos 2\omega_i T = -1$; ω_i заменим на ω ср. в смысле (4а), а $(\omega_i - \omega_j)$ - на $0,7 \omega_{cp}$. (Здесь коэффициент 0,7 появляется ввиду того, что в разности $(\omega_i - \omega_j)$ вес нижних частот увеличивается по линейному закону).

В результате

$$|\delta_j| \leq \frac{1,4 \cdot 10^{-3}}{T} \sum_{i=1}^n \frac{A_i^2}{2}, \quad (I6)$$

а относительная погрешность ε_j равна

$$|\varepsilon_j| \leq \frac{1,4 \cdot 10^{-3}}{T}, \quad (I7)$$

где T - время наблюдения в секундах;

ε_j - величина безразмерная, так как в формуле (I7) в числителе подставлено значение величины T ср. в сек. Заметим, что эта погрешность определена для худшего случая, поэтому математическое ожидание погрешности меньше, чем величина, определяемая по формуле (I7).

Рассмотрение других частных случаев привело к результатам, близким к (I7).

3. Погрешность расчета спектра.

Допустим, что функция корреляции рассчитана с достаточно малой погрешностью, тогда

$$B_i(\tau) = \frac{A_i^2}{2} \cos \omega_i \tau, \quad (I8)$$

и

$$G(\omega) \approx \frac{2}{\pi} \int_0^{\tau_0} \sum_{i=1}^n B_i(\tau) \cos \omega \tau \cdot d\tau, \quad (I9)$$

где τ_0 - интервал времени, для которого рассчитана функция корреляции. Подставив (I8) в (I9), после соответствующих преобразований, получим:

Для частот $\omega = \omega_i$

$$G(\omega) = \sum_{i=1}^n \frac{A_i^2}{\pi} \left(\frac{\tau_0}{2} + \frac{\sin 2\omega_i \tau_0}{4\omega_i} \right), \quad (20)$$

при $\tau_0 \rightarrow \infty$

$$G(\omega) \rightarrow \sum_{i=1}^n \frac{A_i^2 \tau_0}{2\pi} \quad (20a)$$

При ограниченном τ_0 второе слагаемое вносит погрешность в расчет $G(\omega)$.

Для i -й составляющей абсолютная погрешность δ_4 равна

$$\delta_4 = \frac{A_i^2}{4\pi\omega_i} \sin 2\omega_i \tau_0.$$

Эта погрешность убывает с ростом частоты. Относительная погрешность

$$\varepsilon_4 = \frac{\delta_4 \cdot 2\pi}{A_i^2 \tau_0} = \frac{\sin 2\omega_i \tau_0}{2\omega_i \tau_0}. \quad (21)$$

Огибающую находим, положив $\sin 2\omega_i \tau_0 = \pm 1$,

$$|\varepsilon_4| \leq \frac{1}{2\omega_i \tau_0}.$$

Наибольшим искажениям подвергаются низкочастотные составляющие спектра. Худший случай: $\omega_i = \omega_{\min} = \omega_{OT} = \frac{2\pi}{T_{OT}}$. Отсюда

$$|\varepsilon_4| \leq \frac{0.08 T_{OT}}{\tau_0}. \quad (22)$$

Для частот $\omega \neq \omega_i$

$$G^*(\omega) = \frac{1}{\pi} \sum_{i=1}^n A_i^2 \left[\frac{\sin(\omega_i - \omega) \tau_0}{2(\omega_i - \omega)} + \frac{\sin(\omega_i + \omega) \tau_0}{2(\omega_i + \omega)} \right]. \quad (23)$$

Чтобы найти затухание при разных расстройках $\Delta\omega = \omega_i - \omega$, разделим выражения для i -й составляющей из (23) на таковое из (20a). Тогда погрешность примет вид:

$$\varepsilon_5 = f(\omega, \tau_0) = \frac{\sin(\omega_i - \omega) \tau_0}{(\omega_i - \omega) \tau_0} + \frac{\sin(\omega_i + \omega) \tau_0}{(\omega_i + \omega) \tau_0}. \quad (24)$$

В физическом смысле $f(\omega, \tau_0)$ представляет собой аналитическое выражение характеристики машинного фильтра. Так как $\omega_i - \omega = \Delta\omega$, а $\omega_i + \omega = 2\omega_i - \Delta\omega$, то

$$f(\omega, \tau_0) = \frac{\sin \Delta\omega \tau_0}{\Delta\omega \tau_0} + \frac{\sin(2\omega_i - \Delta\omega) \tau_0}{(2\omega_i - \Delta\omega) \tau_0}. \quad (25)$$

В области верхних частот, т.е. при $2\omega_i \gg \Delta\omega$,

$$f(\omega, \tau_0) \approx \frac{\sin \Delta\omega \tau_0}{\Delta\omega \tau_0}. \quad (26)$$

Таким образом, параметры фильтра определяются временем интегрирования.

Выражение для огибающей

$$|\varepsilon_5|_r = |f(\omega, \tau_0)| \leq \frac{1}{\Delta\omega \tau_0}. \quad (27)$$

Здесь $\Delta\omega$ выбирается равной расстоянию по оси частот между соседними спектральными составляющими. Для речевого сигнала

$$\Delta\omega = \omega_{OT} = \frac{2\pi}{T_{OT}},$$

откуда

$$|\varepsilon_5|_r \leq \frac{0.16 \cdot T_{OT}}{T}. \quad (28)$$

В общем случае

$$\varepsilon_5 = \frac{\sin \Delta\omega \tau_0}{\Delta\omega \tau_0} + \frac{\sin(2\omega_i - \Delta\omega) \tau_0}{(2\omega_i - \Delta\omega) \tau_0};$$

в низкочастотной части спектра $\Delta\omega = \omega_{OT}$, $\omega_i = 2\omega_{OT}$. Тогда

$$\varepsilon_5 = \frac{\sin \omega_{OT} \tau_0}{\omega_{OT} \tau_0} + \frac{\sin 3\omega_{OT} \tau_0}{3\omega_{OT} \tau_0}; |\varepsilon_5|_r^* \leq \frac{1.2}{\omega_{OT} \tau_0}. \quad (29)$$

Ближайшая спектральная составляющая вносит погрешность $|\varepsilon_5|_r^*$, а 2-я, 3-я и последующие составляющие вносят погрешности соответственно:

$$|\varepsilon_5|_2^* \approx \frac{1.2}{2\omega_{OT} \tau_0}, |\varepsilon_5|_3^* \approx \frac{1.2}{3\omega_{OT} \tau_0} \text{ и т.д.}$$

При $2\pi\omega_{OT}\tau_0 > \pi$, т.е. 20 мсек $> T > 10$ мсек, где $T = 2\tau_0$ (наиболее часто встречающийся на практике случай), погрешности, вносимые четными и нечетными составляющими имеют противоположные знаки.

Общая погрешность для нижних частот (учитываются погрешности, вносимые только более высокочастотными составляющими):

$$|\varepsilon_5| = |\varepsilon_5|_1^* - |\varepsilon_5|_2^* + |\varepsilon_5|_3^* - |\varepsilon_5|_4^* + \dots = \frac{1.2}{\omega_{OT} \tau_0} \left[1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots \right] = \frac{1.2}{\omega_{OT} \tau_0} \ln 2 \approx \frac{13.2 \cdot 10^{-4}}{\tau_0}. \quad (29a)$$

При рассмотрении формул (4), (II), (I3), (2I) и (25) видно, что все погрешности носят колебательный, затухающий во времени характер, т.е. периодически принимают значение, равное нулю.

Это происходит при соблюдении, например, условий

$$T \cdot \omega_{от} = 2\kappa\pi,$$

где $\kappa = 1, 2, 3, \dots$,

$$\tau_o = \frac{T}{2}.$$

и При $K=1$ можно получить огибающую спектра, близкую к идеальной (начиная с частот 1 кГц и выше, где соблюдается условие $2\omega_i \gg \Delta\omega$), но при условии соблюдения требований теоремы Котельникова, т.е. когда число спектральных линий достаточно для восстановления спектральной функции $G(\omega)$ с помощью ряда:

$$\sum_{i=1}^n G(\omega_i) \cdot \frac{\sin(\omega_i - \omega)\tau_o}{(\omega_i - \omega)\tau_o}.$$

При увеличении K появляются пики на частотах, соответствующих гармоническим составляющим, и при $K \rightarrow \infty$ спектр стремится к линейчатому (рис. 1). При выборе времени интегрирования от t до $t+T$ приходится обрабатывать участок функции от t до $t+(T+\tau_o)$ в связи с выбранным способом расчета функции корреляции.

При несоблюдении условия $T = KT_{от}$ появляются погрешности, оценить которые можно по формулам (5), (12), (17), (22), (28) и (29).

Результаты расчета спектра синтезированного сигнала с известным спектром для случаев 1) $\varphi_i = 0$ и 2) $\varphi_i = 30^\circ$ приведены на рис. 2. При $T = \frac{1}{K} T_{от}$ сдвиг фазы не вызывает изменения спектра, так как погрешность, определяемая фазами составляющих (см. (13), (11)), равна нулю. При $T \neq \frac{1}{K} T_{от}$ сдвиг фазы приводит к изменению спектра, что вызывается погрешностями расчета из-за ограничения времени наблюдения.

В ы в о д ы. Приближенные значения погрешностей, возникающих на различных этапах расчета спектральной функции речевых сигналов, определяются по следующим формулам:

Этап расчета	m	$B_i(\tau)$	$\sum_{i=1}^n B_{\delta i}(\tau)$	$G(\omega_i)$	$\Delta G(\omega \neq \omega_i)$
Погрешность	$ k \leq \frac{4 \cdot 10^{-4}}{T}$	$ E_2 \leq \frac{0,16 T_{от}}{T}$	$ E_3 \leq \frac{1,4 \cdot 10^{-3}}{T}$	$ E_4 \leq \frac{0,08 T_{от}}{T}$	$ E_5 \leq \frac{0,13 T_{от}}{T_o}$

Здесь T , τ_o и $T_{от}$ - в секундах. В формулах для E_1 и E_3 в числителе подставлено численное значение величины $T_{от}$ и $T_{от}$.

Для подсчета суммарной погрешности необходимо проделать дополнительные преобразования.

1. Погрешность расчета $B(\tau)$, обусловленная погрешностью E_1 ,

$$\Delta B_{E_1}(\tau) \approx \frac{1}{T_o} \int_0^T \{f(t)E_1 + f(t+\tau) \cdot E_1 + E_1^2\} dt,$$

$$|E_2| = \frac{\Delta B_E(\tau)_{max}}{B_i(\tau)_{max}} \approx \frac{E_1^2(T+2)}{T} \approx \frac{32 \cdot 10^{-8}}{T^3}, \quad (29б)$$

так как $2 \gg T_{сек}$.

2. Суммарная погрешность расчета функции корреляции:

$$|E_{B(\tau)}| = |E_2| + |E_2| + \ln 2 |E_3|,$$

где $\ln 2$ появляется по тем же причинам, что и при расчете $|E_3|$ (см. вывод формулы (29а)).

Для $T_{от} \leq 10$ мсек

$$|E_{B(\tau)}| \leq \frac{2,56 \cdot 10^{-3}}{T} + \frac{32 \cdot 10^{-8}}{T^3}. \quad (29в)$$

3. Погрешность расчета $G(\omega)$, обусловленная погрешностью $E_{B(\tau)}$,

$$|\Delta G(\omega)|_{отн} = \frac{1}{T_o} \int_0^{T_o} E_{B(\tau)} \cos \omega \tau d\tau \leq \frac{|E_{B(\tau)}|}{\omega T_o} \sin \omega T_o$$

Огибающая погрешности $|\Delta G(\omega)|_{отн}$ при $T_o = \frac{T}{2}$ описывается выражением

$$|\Delta G(\omega)|_{отн} \leq \frac{8 \cdot 10^{-6}}{T^2} + \frac{10^{-9}}{T^4}. \quad (29г)$$

4. Суммарная погрешность расчета спектральной функции $|E_{G(\omega)}|$ при $T \leq 10$ мсек будет

$$|E_{G(\omega)}| = |E_4| + |E_5| + |\Delta G(\omega)|_{отн} \leq \frac{4,2}{T} + \frac{8}{T^2} + \frac{1000}{T^4}, \quad (29д)$$

где T - время наблюдения в миллисекундах.

Эта формула даёт верхний предел погрешности для $f_{от} = 100$ гц. Общее выражение для погрешности $E_{G(\omega)}$ справедливо для $T \geq 1,5 T_{от}$, так как в этом случае вершины функции $\frac{\sin x}{x}$ лежат на кривой $\frac{1}{x}$.

О частоте квантования речевого сигнала. Как известно, теорема отсчетов Котельникова применима к функциям с ограниченным спектром. Тогда частота отсчетов $f_{кв} = 2F_{max}$ где F_{max} — верхняя частота спектра сигнала.

При ограничении времени наблюдения спектр становится теоретически бесконечным. В этом случае можно ориентироваться на верхнюю частоту сигнала с определенной погрешностью, определяемой соотношением

$$\varepsilon_6 = \frac{\int_{\omega_{max}}^{\infty} G(\omega) \cdot d\omega}{\int_{-\infty}^{\infty} G(\omega) \cdot d\omega} \quad (30)$$

Иначе говоря, погрешность определяется относительной энергией сигнала, приходящейся на частоты $\omega > \omega_{max}$. В этом случае для восстановления функции с помощью ряда Котельникова с заданной погрешностью нужно выполнить условие

$$f_{кв} = 2f_{max}, \text{ где } f_{max} = \frac{\omega_{max}}{2\pi} \quad (30a)$$

Рассмотрим вопрос о взаимосвязи ω_{max} и времени наблюдения для машинного фильтра. Пусть имеется синусоидальный нормированный по амплитуде сигнал с частотой ω_i , который мы анализируем в течение времени τ_0 . Спектральная функция такого сигнала имеет вид

$$G(\omega)_{отн} = \frac{\sin(\omega - \omega_i)\tau_0}{(\omega - \omega_i)\tau_0} + \frac{\sin(\omega + \omega_i)\tau_0}{(\omega + \omega_i)\tau_0} \quad (31)$$

Полная энергия сигнала

$$G_{общ} = \int_{-\infty}^{\infty} G(\omega) d\omega = \int_{-\infty}^{\infty} \left[\frac{\sin(\omega - \omega_i)\tau_0}{(\omega - \omega_i)\tau_0} + \frac{\sin(\omega + \omega_i)\tau_0}{(\omega + \omega_i)\tau_0} \right] d\omega.$$

Проведя замену переменных $x = (\omega - \omega_i)\tau_0$, $y = (\omega + \omega_i)\tau_0$, получим

$$G_{общ} = \frac{1}{\tau_0} \left[\int_{-\omega_i\tau_0}^{\omega_{max}\tau_0} \frac{\sin x}{x} dx + \int_{\omega_i\tau_0}^{\omega_{max}\tau_0} \frac{\sin x}{x} dx + \int_{-\omega_i\tau_0}^{\omega_{max}\tau_0} \frac{\sin y}{y} dy + \left(\int_{\omega_i\tau_0}^{\omega_{max}\tau_0} \frac{\sin y}{y} dy \right) \right].$$

Первое и четвертое слагаемые взаимно уничтожаются, так как функция $\frac{\sin x}{x}$ четная.

$$G_{общ} = \frac{\pi}{\tau_0} \quad (32)$$

Напомним, что формула (31) справедлива для $\omega \neq \omega_i$. Энергия сигнала на этих частотах убывает при увеличении τ_0 . Определим относительную погрешность следующим образом:

$$\varepsilon_6 = \frac{G_{общ} - G_{акт}}{G_{общ}}, \quad (33)$$

где $G_{акт}$ — энергия сигнала, сосредоточенная в активной полосе. Под активной полосой подразумевается полоса частот от 0 до ω_{max} , в которой сосредоточена подавляющая часть энергии сигнала [7].

Соотношение (30) можно переписать в виде:

$$\varepsilon_6 = \frac{\int_{\omega_{max}}^{\infty} G(\omega) \cdot d\omega}{G_{общ}}, \text{ где } G(\omega) = \begin{cases} G(\omega) & \text{при } \omega > \omega_{max}, \\ 0 & \text{при } \omega < \omega_{max}. \end{cases}$$

Ограничивая спектр сигнала частотой ω_{max} , заменяем реальный сигнал идеализированным со спектральной функцией.

$$G_2(\omega) = \begin{cases} G(\omega) & \text{при } \omega \leq \omega_{max}, \\ 0 & \text{при } \omega > \omega_{max}. \end{cases}$$

Отсюда $G_1(\omega) = G(\omega) - G_2(\omega)$,

где $G(\omega)$ — спектральная функция исходного сигнала;

$G_2(\omega)$ — спектральная функция восстановленного сигнала;

$G_1(\omega)$ — разностная спектральная функция.

$$G_1(\omega) = \frac{2}{\pi} \int_0^{\tau_0} B_1(\tau) \cos \omega \tau \cdot d\tau = \frac{2}{\pi} \int_0^{\tau_0} B(\tau) \cos \omega \tau \cdot d\tau - \frac{2}{\pi} \int_0^{\tau_0} B_2(\tau) \cos \omega \tau \cdot d\tau = \frac{2}{\pi} \int_0^{\tau_0} [B(\tau) - B_2(\tau)] \cos \omega \tau \cdot d\tau.$$

Отсюда $B_1(\tau) = B(\tau) - B_2(\tau)$.

Таким образом, ε_6 представляет собой относительную энергию сигнала, имеющего автокорреляционную функцию, равную разности автокорреляционных функций исходного и восстановленного сигналов. Итак,

$$\varepsilon_6 = \frac{1}{\pi} \left[\int_0^{\omega_{max}} \frac{\sin(\omega - \omega_i)\tau_0}{(\omega - \omega_i)\tau_0} d\omega + \int_0^{\omega_{max}} \frac{\sin(\omega + \omega_i)\tau_0}{(\omega + \omega_i)\tau_0} d\omega \right].$$

После замены переменных

$$\varepsilon_6 = \frac{1}{\pi} \left[\int_{-\omega_i\tau_0}^{\omega_{max}\tau_0} \frac{\sin x}{x} dx + \int_{\omega_i\tau_0}^{\omega_{max}\tau_0} \frac{\sin y}{y} dy \right] = \frac{1}{\pi} \left[\int_0^{\omega_{max}\tau_0} \frac{\sin x}{x} dx + \int_0^{\omega_{max}\tau_0} \frac{\sin y}{y} dy \right] \quad (34)$$

Зная ω_i , τ_0 и допустимую ε_6 , можно по таблицам интегрального синуса $Si(x)$ определить ω_{max} и $f_{кв} = \frac{\omega_{max}}{\pi}$.

Известно, что при $x \gg \pi$, $\int_0^x \frac{\sin x}{x} dx \approx \frac{\pi}{2}$. Это условие соблюдается для сигналов речи в области верхних

частот (а именно они определяют частоту отсчетов). Действительно, $(\omega - \omega_i) \tau_0$ имеет величину порядка π и выше. Тогда для

$$f_i \geq 5 \text{ кГц} \quad (\omega_{\max} + \omega_i) \tau_0 \geq 200 \pi.$$

Отсюда

$$\varepsilon_6 \approx \frac{1}{2} - \frac{1}{\pi} \int_0^{(\omega_{\max} - \omega_i) \tau_0} \frac{\sin x}{x} dx. \quad (35)$$

При $(\omega_{\max} - \omega_i) \tau_0 \gg 1$

$$\varepsilon_6 \approx \frac{1}{\pi} \cdot \frac{\cos(\omega_{\max} - \omega_i) \tau_0}{(\omega_{\max} - \omega_i) \tau_0}. \quad (36)$$

Для примера положим

$$\omega_i = 2\pi(6 \text{ кГц}); \quad \omega_{\max} = (6,3 \text{ кГц}) 2\pi; \quad \tau_0 = 10 \text{ мсек}.$$

Тогда

$$(\omega_{\max} - \omega_i) \tau_0 \approx 19 \gg 1.$$

Погрешность (36) носит периодический затухающий во времени характер.

Огибающая этой функции

$$|\varepsilon_6| \leq \frac{1}{\pi(\omega_{\max} - \omega_i) \tau_0} = \frac{1}{2\pi^2(f_{\max} - f_i) \tau_0}. \quad (37)$$

Практически под f_i следует понимать верхнюю граничную частоту фильтра, стоящего на входе системы, то есть $f_i = F_{\max}$, по терминологии соотношения $f_{\text{кв}} = 2F_{\max}$.

Тогда

$$|\varepsilon_6| \leq \frac{0,1}{\tau_0(f_{\text{кв}} - 2F_{\max})}, \quad (38)$$

$$f_{\text{кв}} \geq 2F_{\max} + \frac{0,1}{\tau_0 |\varepsilon_6|}. \quad (39)$$

Таким образом, частота квантования определяется не только параметрами сигнала ($F_{\max}$), но и методом его обработки (временем наблюдения и способом расчёта спектральной функции).

Пример: пусть $F_{\max} = 5 \text{ кГц}$, $|\varepsilon_6| = 0,01$,

тогда при $\tau_0 = 10 \text{ мсек}$ $f_{\text{кв}} \geq 11 \text{ кГц}$,

$\tau_0 = 20 \text{ мсек}$ $f_{\text{кв}} \geq 10,5 \text{ кГц}$.

Мы рассмотрели случай одной синусоидальной составляющей.

В реальных условиях вид спектра в области частот $f > F_{\max}$ определяется не только составляющей $F_{\max}$, но и соседними составляющими. Тогда для $\omega_{\text{от}} \tau_0 > \pi$ в правой части выражения (38) появится множитель $\ln 2$ по тем же причинам, что и при вы-

воде формулы (29а). Таким образом, для речевых сигналов имеет место:

$$|\varepsilon_6| \leq \frac{\ln 2 \cdot 0,1}{\tau_0(f_{\text{кв}} - 2F_{\max})} \quad (39а)$$

и

$$f_{\text{кв}} \geq 2F_{\max} + \frac{0,07}{\tau_0 |\varepsilon_6|}. \quad (39б)$$

Из формулы (35) видно, что погрешность носит колебательный характер с прохождением через нуль. Истолковать это можно так: имеются такие особые точки, определяющие частоту квантования, попадание в которые обеспечивает восстановление сигнала, эквивалентного исходному в том смысле, что исходный и восстановленный сигналы имеют одинаковую энергию, максимумы их спектральных функций приходятся на одни и те же частоты и величины этих максимумов равны. На рис.3 изображены спектр исходного синусоидального сигнала при ограниченном времени наблюдения (рис.3а) и случай попадания в первую и вторую особые точки (рис.3б и рис.3в, соответственно).

При увеличении номера особой точки

$$f_{\text{кв}} = 2F_{\max} + \frac{0,1}{\tau_0 |\varepsilon_6|}.$$

Чтобы попасть в эти особые точки, нужно точно знать $F_{\max}$, т.е. в конечном счёте частоту основного тона. Тогда можно сильно сократить обрабатываемый массив за счёт уменьшения $f_{\text{кв}}$. В речевых сигналах $f_{\text{от}}$ не является постоянной величиной, так как зависит от диктора и является носителем определенной информации, поэтому реализовать уменьшение $f_{\text{кв}}$ весьма трудно. Это применимо для спектров с постоянным шагом по оси частот между соседними составляющими.

Для синусоидального сигнала $\sin \omega_i t$ I-я особая точка $(\omega_{\max} - \omega_i) \tau_0 \approx 1,92$, откуда

$$f_{\text{кв}} \approx 2F_{\max} + \frac{0,611}{\tau_0}. \quad (40)$$

В случае сложного сигнала (т.е. при наличии нескольких спектральных составляющих) получить $|\varepsilon_6| = 0$ невозможно, так как особые точки разных составляющих не совпадают между собой.

В этом случае при $|\varepsilon_6| \leq 2\%$ $f_{\text{кв}} \geq 2F_{\max} + \frac{0,611}{\tau_0}, \quad (41)$

а при $|\varepsilon_6| \leq 0,9\%$ $f_{\text{кв}} \geq 2F_{\max} + \frac{1,56}{\tau_0}. \quad (42)$

Для сравнения приведем значения f_{kv} , выведенные из (39), при $|\varepsilon_0| \leq 2\%$

$$f_{kv} \geq 2F_{max} + \frac{5}{\tau_0}, \quad (43)$$

при $|\varepsilon_0| \leq 0,9\%$

$$f_{kv} \geq 2F_{max} + \frac{10}{\tau_0}. \quad (44)$$

Особенно значительное уменьшение f_{kv} достигается при малом τ_0 .

Пример: пусть имеется сигнал с верхней частотой, равной F_{max} , время наблюдения $\tau_0 = \frac{4}{F_{max}}$. Тогда из (41) и (42) получим соответственно

$$\text{для } |\varepsilon_0| \leq 2\% \quad f_{kv} = 2,15 F_{max},$$

$$\text{для } |\varepsilon_0| \leq 0,9\% \quad f_{kv} = 2,39 F_{max}.$$

Из (43) и (44) получим соответственно

$$\text{для } |\varepsilon_0| \leq 2\% \quad f_{kv} = 3,25 F_{max},$$

$$\text{для } |\varepsilon_0| \leq 0,9\% \quad f_{kv} = 4,5 F_{max}.$$

В особые точки можно попасть, оставив f_{kv} постоянной, но изменив τ_0 . Эту методику можно использовать для значительного уменьшения времени наблюдения при жестко заданной частоте квантования.

Этот путь более приемлем для обработки сигналов на ЭВМ, так как f_{kv} обычно постоянна и задается входным преобразователем, а τ_0 можно менять в процессе решения задачи.

В общем случае частота квантования при заданной погрешности $|\varepsilon_0|$ определяется по формулам (39), (39б).

В заключение отметим, что спектральный анализ речевых сигналов с помощью ЭВМ в отличие от аналоговой аппаратуры позволяет учесть погрешности и соответствующим подбором времени наблюдения уменьшить их до требуемого уровня.

Л и т е р а т у р а

1. Загоруйко Н.Г. Об обмене устной информацией между человеком и вычислительными системами. Сб. "Вычислительные системы", Изд. ИМ СО АН СССР, IO, 1963.
2. Бриллиэн Л. Наука и теория информации. ФМ, 1962.
3. Железнов Н.А. Энергетические характеристики и интервал корреляции стохастических сигналов, в частности - речевых сигналов. Вопросы статистики речи (материалы совещания ЛГУ им. А.А.Иданова, 1958 г.).
4. Железнов Н.А. Интервал временной корреляции сигналов и его связь с параметрами спектра мощности. Электросвязь, № 15, 1961 г.
5. Харкевич А.А. Спектры и анализ. ФМ, 1962 г.
6. Кутин Б.Н. О вычислении корреляционной функции стационарного случайного процесса по экспериментальным данным. Автоматика и телемеханика, т. XVIII, № 3, 1957 г.
7. Агеев Д.В. Активная полоса частотного спектра функции времени. Труды Горьковского политехнического института, № I, 1955 г.

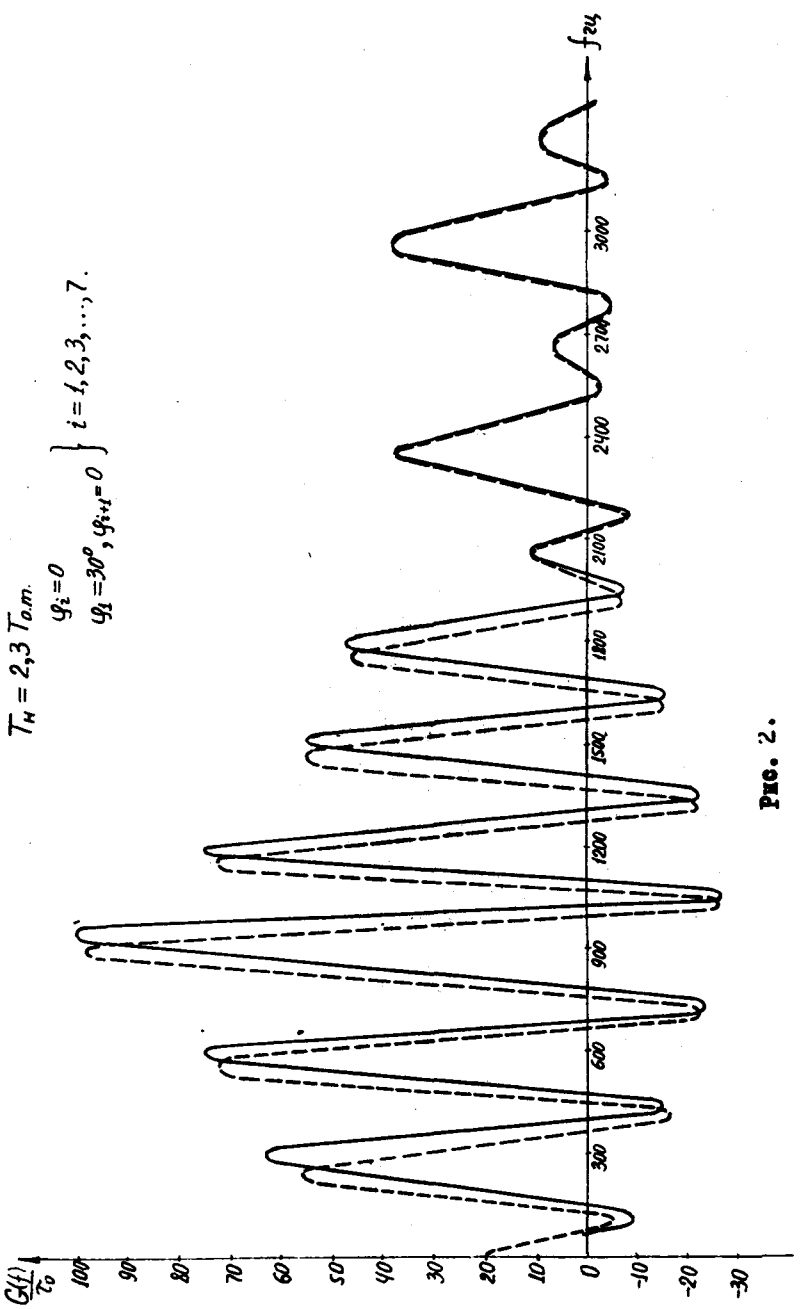


Рис. 2.

$$\left. \begin{array}{l} q_i = 0 \\ q_i = 30^\circ, q_{i+1} = 0 \end{array} \right\} i = 1, 2, 3, \dots, 7.$$

$$T_H = 2, 3 T_{0.m.}$$

ИРЯКОЖЕНЕ

$$T = K T_{0.m.}$$

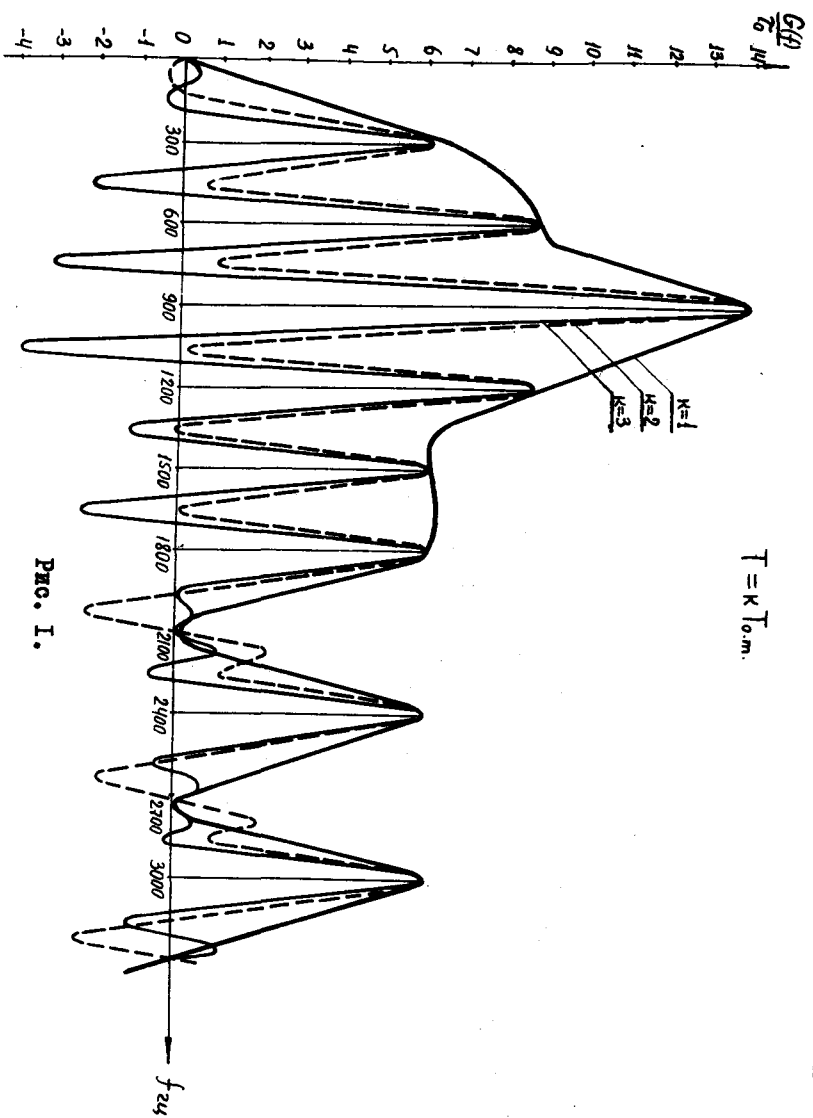


Рис. 1.

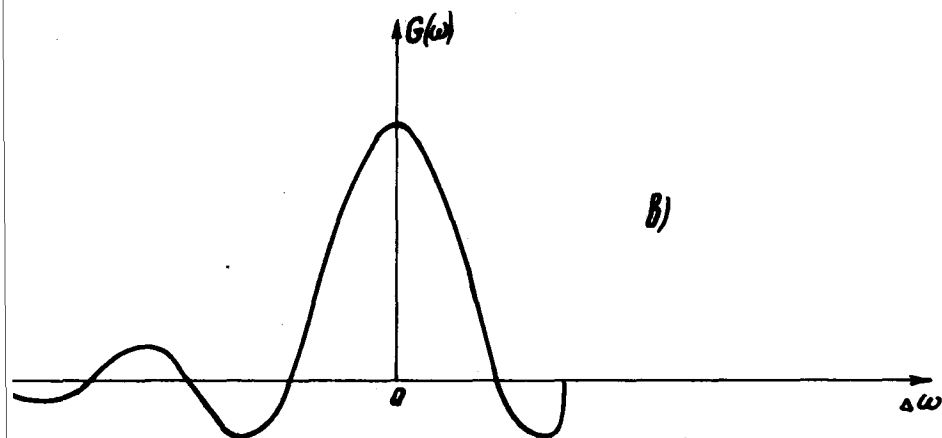
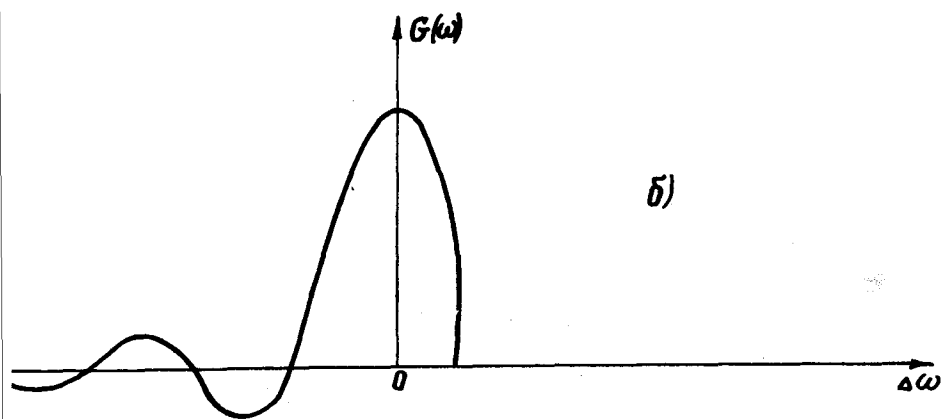
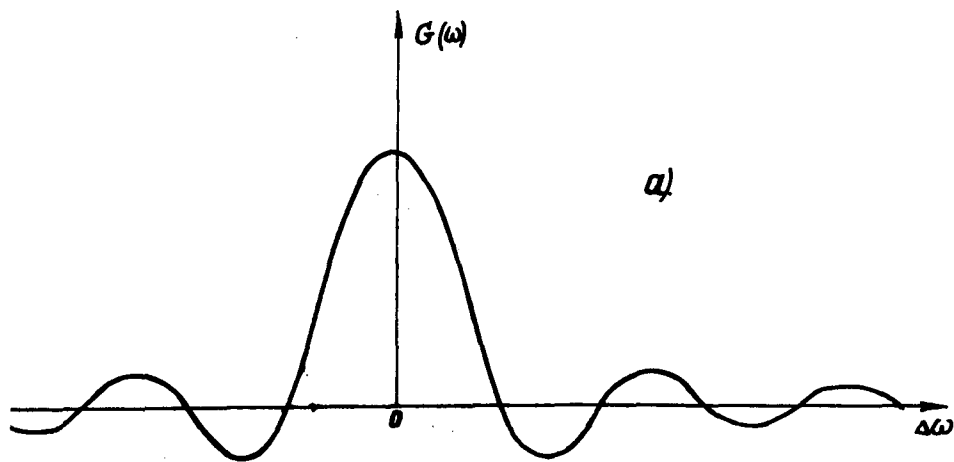


Рис. 3.