где коэффициенты Фурье $\tilde{X}(k)$ представляют собой отсчеты преобразования Фурье вокализованной речи, т. е.

$$\tilde{X}(k) = H_v\left(e^{i\frac{2\pi}{N_p}k}\right) \quad \text{(см. задачу 6.8).}$$

б) Показать, что кратковременное преобразование Фурье от $\tilde{x}(n)$ можно представить в виде

$$\tilde{X}(e^{i\omega}) = \frac{1}{N_p}\sum_{k=0}^{N_p-1} H_v\left(e^{i\frac{2\pi}{N_p}k}\right) W_n(e^{i(\omega-2\pi k/N_p)}),$$

где $W_n(e^{i\omega})$ — преобразование Фурье от $w(n-m)$.

в) Сколько различных значений принимает $X_n(e^{i\omega})$ при фиксированном ω? Для прямоугольного окна

$$w(n) = \begin{cases} 1, & 0 \leqslant n \leqslant N_p - 1; \\ 0, & \textit{в противном случае} \end{cases}$$

найти функцию $W_n(e^{i\omega})$.

г) Для каких значений N_p справедливо равенство

$$\tilde{X}_n\left(e^{i\frac{2\pi}{N_p}k}\right) = H_v\left(e^{i\frac{2\pi}{N_p}k}\right)$$

для прямоугольного окна шириной n?

6.16. Займемся анализом и синтезом сигнала $x(n)=\cos(\omega_0 n)$. Схема для анализа приведена для k-го канала на рис. 3.6.2*а*.

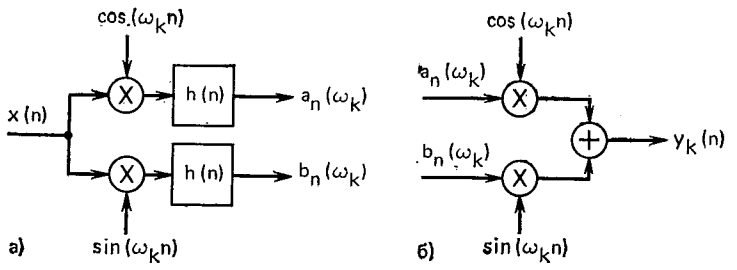


Рис. 3.6.2

а) Определить $a_n(\omega_k)$ и $b_n(\omega_k)$ для заданного входного сигнала.

б) Предположив, что $h(n)$ соответствует узкополосному фильтру нижних частот, упростить полученные выражения для $a_n(\omega_k)$ и $b_n(\omega_k)$ в предположении, что $(\omega_0-\omega_k)$ спадает в полосе фильтра и что $H(e^{i\omega})\approx 1$ для этих частот.

в) Сигналы $a_n(\omega_k)$ и $b_n(\omega_k)$ дают в комбинации мгновенное значение $M_n(\omega_k)$ и производную фазы $\dot{\varphi}_n(\omega_k)$. Определить $M_n(\omega_k)$ и $\dot{\varphi}_n(\omega_k)$ для нашего примера.

г) Показать, что для схемы синтеза, показанной на рис. 3.6.2*б*, выходной сигнал по существу совпадает с входным.

д) Производная фазы $\varphi_n=(\omega_k)$ вычисляется по формуле

$$\dot{\varphi}_n(\omega_k) = \frac{b_n(\omega_k)\dot{a}_n(\omega_k) - a_n(\omega_k)\dot{b}_n(\omega_k)}{[a_n(\omega_k)]^2 + [b_n(\omega_k)]^2}.$$

Найти $\dot{\varphi}_n(\omega_k)$ для нашего примера и сравнить результат с результатом в).

ж) Предположить, что производная фазы из п.д.) вычисляется с помощью первой разности, т. е. $\varphi_n(\omega_k)\approx\frac{1}{T}(a_n(\omega_k)-a_{n-1}(\omega_k))$, где T — период дискретизации во временной области. Найти $\varphi_n(\omega_k)$ по этой формуле и сравнить результат с полученным в п.в). При каких условиях они близки?

7
Гомоморфная обработка речи

7.0. Введение

Одно из основных предположений, сделанных в этой книге, состоит в том, что речевой сигнал трактуется как сигнал на выходе линейной системы с медленно изменяющимися параметрами. Это предположение позволяет считать, что на коротких сегментах речевой сигнал можно рассматривать как сигнал на выходе линейной системы с постоянными параметрами, возбуждаемой либо последовательностью импульсов, либо случайным шумом. Как уже отмечалось, проблема анализа речевого сигнала сводится к измерению параметров модели и оценке изменения этих параметров с течением времени. Поскольку сигнал возбуждения и импульсная характеристика фильтра взаимодействуют через операцию свертки, задача анализа речи может рассматриваться как задача разделения компонент, участвующих в операции свертки. Такая задача иногда называется задачей обратной свертки[1]. В гл. 6 был рассмотрен метод ее решения на основе представления речевого сигнала в виде переменного во времени преобразования Фурье. В данной главе на основе использования теории, изложенной в гл. 6, развивается другой подход к задаче, названный гомоморфной фильтрацией. После краткого введения в общую теорию гомоморфных систем будут рассмотрены различные способы применения методов гомоморфной обратной свертки в области анализа речевых сигналов.

7.1. Гомоморфные относительно свертки системы

Гомоморфные относительно свертки системы удовлетворяют обобщенному принципу суперпозиции. Принцип суперпозиции, если его записать для обычных линейных систем, имеет вид

$$\begin{aligned} L[x(n)] &= L[x_1(n)+x_2(n)] = \\ &= L[x_1(n)]+L[x_2(n)] = \\ &= y_1(n)+y_2(n) = y(n) \end{aligned} \tag{7.1а}$$

и

$$L[ax(n)] = aL[x(n)] = ay(n), \tag{7.1б}$$

где L — линейный оператор. Принцип суперпозиции устанавливает, что если сигнал на входе является линейной комбинацией элементарных сигналов, то и сигнал на выходе будет представлен в виде линейной комбинации соответствующих сигналов. Этот принцип иллюстрируется на рис. 7.1, где символ «+» на входе и выходе

[1] Операцию, обратную свертке (deconvolution), в переводной литературе также называют «разверткой». (*Прим. ред.*)

означает, что аддитивная комбинация сигналов на входе приводит к аддитивной комбинации выходных сигналов.

Как показано в гл. 2, прямым следствием принципа суперпозиции является тот факт, что сигнал на выходе линейной системы может быть представлен в виде дискретной свертки:

$$y(n)=\sum_{k=-\infty}^{\infty} h(n-k)\,(x(k)=h(n)*x(n). \tag{7.2}$$

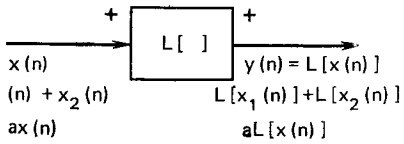


Рис. 7.1. Представление системы, в которой выполняется принцип суперпозиции

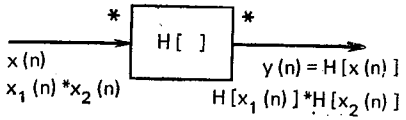


Рис. 7.2. Представление системы, гомоморфной относительно свертки

Символ « * » здесь и далее означает свертку в дискретном времени. По аналогии с принципом суперпозиции для обычных линейных систем определим класс систем, удовлетворяющих обобщенному принципу суперпозиции, в котором сложение заменяется сверткой (легко показать, что свертка обладает такими же алгебраическими свойствами, как и сложение [1]), т. е.

$$\begin{aligned} H[x(n)] &= H[x_1(n)*x_2(n)]= \\ &= H[x_1(n)]*H[x_2(n)]= \\ &= y_1(n)*y_2(n)=y(n). \end{aligned} \tag{7.3}$$

В общем случае возможно сформулировать и уравнение, аналогичное (7.1б), в котором выражено свойство скалярного умножения [2], однако обобщенное скалярное умножение далее не используется. Системы, обладающие свойством (7.3), названы гомоморфными относительно свертки системами. Эта терминология объясняется тем [3], что данные преобразования оказываются гомоморфными преобразованиями линейного векторного пространства. При изображении таких систем (рис. 7.2) операцию свертки представляют в явном виде на входе и выходе системы. Гомоморфный фильтр является гомоморфной системой, обладающей тем свойством, что одна компонента (выделяемая) проходит через эту систему без изменений, а другая — устраняется. В соотношении (7.3), например, если $x_1(n)$ — нежелательная компонента, то необходимо потребовать, чтобы выход, соответствующий $x_1(n)$, представлял собой единичный отсчет, в то время как выход, соответствующий $x_2(n)$, близко совпадал бы с $x_2(n)$. Это полностью аналогично ситуации в линейных системах, где ставится задача выделения сигнала из смеси его с аддитивным шумом.

Важным аспектом теории гомоморфных систем является то, что любая из них может быть представлена в виде каскадного

соединения трех гомоморфных систем, как это изображено на рис. 7.3 для случая систем, гомоморфных относительно свертки [3]. Первый блок преобразует компоненты на входе, представленные в виде свертки, в аддитивную сумму на выходе. Второй блок — обычная линейная система, удовлетворяющая принципам

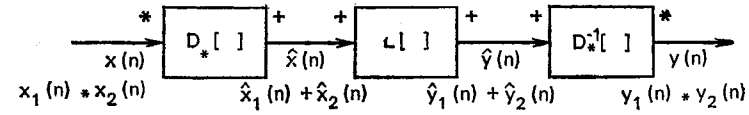


Рис. 7.3. Каноническая форма системы, гомоморфной относительно свертки

суперпозиции в соответствии с (7.1). Третий блок является обратным первому, т. е. преобразует сигналы, представленные в виде суммы, в сигналы, представленные в виде свертки. Важность такого канонического представления заключается в том, что разработка гомоморфной системы сводится к разработке линейной системы. Блок$_*$[ ], называемый характеристическим блоком гомоморфной относительно свертки системы, фиксирован при каноническом представлении, приведенном на рис. 7.3. Очевидно, что обратное преобразование также фиксировано. Характеристическая система для гомоморфной обратной свертки подчиняется обобщенному принципу суперпозиции, в котором операция на входе — свертка, а на выходе — обычное сложение. Свойства характеристической системы определяются выражением

$$\begin{aligned} D_*[x(n)] &= D_*[x_1(n)*x_2(n)]= \\ &= D_*[x_1(n)]+D_*[x_2(n)]= \\ &= \hat{x}_1(n)+\hat{x}_2(n)=\hat{x}(n). \end{aligned} \tag{7.4}$$

Аналогично обратная характеристическая система удовлетворяет соотношению

$$\begin{aligned} D_*^{-1}[\hat{y}(n)] &= D_*^{-1}[\hat{y}_1(n)+\hat{y}_2(n)], = \\ &= D_*^{-1}[\hat{y}_1(n)]*D_*^{-1}[\hat{y}_2(n)]= \\ &= y_1(n)*y_2(n)=y(n). \end{aligned} \tag{7.5}$$

Математическое описание характеристической системы определяется требованиями к выходному сигналу. Если на входе имеется сигнал свертки, то

$$x(n)=x_1(n)*x_2(n) \tag{7.6}$$

и z-преобразование входного сигнала имеет вид

$$X(z)=X_1(z)\,X_2(z). \tag{7.7}$$

Из (7.4) очевидно, что z-преобразование сигнала на выходе системы должно представлять собой сумму z-преобразований компонент. Таким образом, в частотной области характеристическая

система для свертки должна обладать следующим свойством: если на входе имеется произведение компонент, то на выходе должна возникнуть их сумма. Один из подходов к синтезу такой системы представлен на рис. 7.4. Этот подход основан на том,

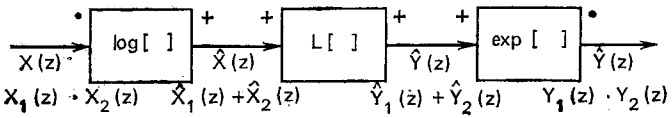


Рис. 7.4. Представление системы, гомоморфной относительно свертки в частотной области

что логарифм произведения равен сумме логарифмов сомножителей, т. е.

$$\hat{X}(z) = \log[X(z)] = \log[X_1(z)\,X_2(z)] = \\ = \log[X_1(z)] + \log[X_2(z)]. \tag{7.8}$$

Если необходимо представлять сигналы во временно́й, а не в частотной области, то характеристическая система примет вид, представленный на рис. 7.5. Аналогичное обратное преобразование показано на рис. 7.6.

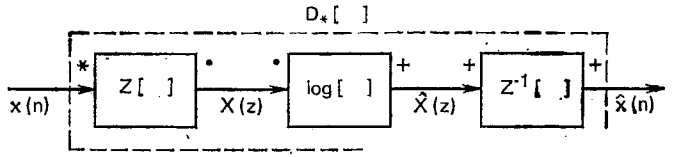


Рис. 7.5. Представление характеристической системы, гомоморфной относительно свертки

Представление прямой и обратной характеристических систем зависит от справедливости соотношения (7.8). Таким образом, логарифм должен быть определен так, чтобы логарифм произведения равнялся сумме логарифмов сомножителей. Это тривиально для действительных положительных величин. Однако в общем случае z-преобразование имеет комплексный характер и вопрос единственности логарифма комплексной случайной величины чрезвычайно важен. С точки зрения вычислений целесообразно рас-

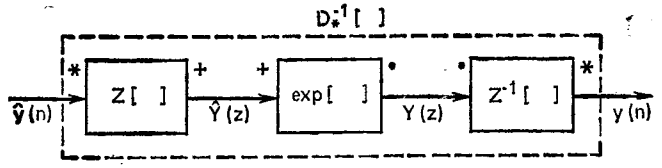


Рис. 7.6. Представление характеристической системы обратной гомоморфной системе

смотреть случай, когда (7.8) справедливо на единичной окружности, т. е. для $z=e^{i\omega}$. Детальное обсуждение проблемы единственности дано в [2]. Для решаемых здесь задач вполне подходит определение логарифма в виде

$$\hat{X}(e^{i\omega}) = \log|X(e^{i\omega})| + i\arg[X(e^{i\omega})]. \tag{7.9}$$

В этом соотношении действительная часть $\log|X(e^{i\omega})|$ не вызывает трудностей. Проблема единственности возникает при определении мнимой части (т. е. $\arg[X(e^{i\omega})]$), которая представляет собой фазовый угол z-преобразования, вычисленного на единичной окружности. В [2] показано, что одним из подходов к решению проблемы единственности является предположение, что фазовый угол представляет собой непрерывную нечетную функцию. В этих условиях уравнение (7.8) справедливо.

С учетом возможности вычисления комплексного логарифма, удовлетворяющего (7.8), обратное преобразование комплексного логарифма преобразования Фурье входного сигнала, являющееся выходом характеристической системы для свертки, имеет вид

$$\hat{x}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{X}(e^{i\omega})\,e^{i\omega n}\,d\omega. \tag{7.10}$$

Выход характеристической системы назван «комплексным кепстром» (термин «кепстр» введен Богертом и др. [4] и является в настоящее время общепринятым для обозначения обратного преобразования Фурье логарифма спектра мощности сигнала; термин «комплексный кепстр» означает, что применяется комплексный логарифм). Термин «кепстр» далее используется для величины

$$c(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|X(e^{i\omega})|\,e^{i\omega n}\,d\omega. \tag{7.11}$$

[Можно показать, что последовательность $c(n)$ представляет собой четную часть комплексного кепстра $\hat{x}(n)$ (см. задачу 7.1).]

Выше была определена характеристическая система для гомоморфной свертки и, таким образом, определена каноническая форма всех гомоморфных систем относительно свертки. Все системы этого класса отличаются только линейной частью. Выбор линейной системы определяется свойствами входного сигнала. Следовательно, для правильного построения линейной системы необходимо прежде всего определить вид и структуру сигнала на выходе характеристической системы, т. е. рассмотреть свойства комплексного кепстра для типичных входных сигналов.

7.1.1. Свойства комплексного кепстра

Для определения свойств комплексного кепстра достаточно рассмотреть случай рационального z-преобразования. Наиболее общая форма преобразования имеет вид

$$X(z)=\frac{Az^r\prod_{k=1}^{M_i}(1-a_k z^{-1})\prod_{k=1}^{M_0}(1-b_k z)}{\prod_{k=1}^{N_i}(1-c_k z^{-1})\prod_{k=1}^{N_0}(1-d_k z)}, \qquad (7.12)$$

где модули величин a_k, b_k, c_k и d_k меньше единицы. Таким образом, сомножители $(1-a_k z^{-1})$ и $(1-c_k z^{-1})$ соответствуют нулям и полюсам внутри единичной окружности, а $(1-b_k z)$ и $(1-d_k z)$ — нулям и полюсам вне единичной окружности. Параметр z^r означает соответствующую задержку во временно́й области. В соответствии с предложением уравнения (7.8) комплексный логарифм $X(z)$ имеет вид

$$\hat{X}(z)=\log[A]+\log[z^r]+\sum_{k=1}^{M_i}\log(1-a_k z^{-1})+\sum_{k=1}^{M_0}\log(1-b_k z)-$$
$$-\sum_{k=1}^{N_i}\log(1-c_k z^{-1})-\sum_{k=1}^{N_0}\log(1-d_k z). \qquad (7.13)$$

Когда (7.13) вычисляется на единичной окружности, легко видеть, что член $\log[e^{i\omega r}]$ вносит вклад только в минимальную часть комплексного логарифма. Поскольку этот член несет информацию только о взаимном расположении во временно́й области, то при вычислении комплексного кепстра он обычно опускается [2]. Таким образом, при обсуждении свойств комплексного кепстра далее этот член не рассматривается. Используя то обстоятельство, что логарифм можно разложить в степенно́й ряд, относительно несложно показать, что комплексный кепстр имеет вид

$$\hat{x}(n)=\begin{cases}\log[A], & n=0,\\ \sum_{k=1}^{N_i}\frac{c_k^n}{n}-\sum_{k=1}^{M_i}\frac{a_k^n}{n}, & n>0,\\ \sum_{k=1}^{M_0}\frac{b_k^{-n}}{n}-\sum_{k=1}^{N_0}\frac{d_k^{-n}}{n}, & n<0.\end{cases} \qquad (7.14)$$

Уравнения (7.14) позволяют выявить ряд важных свойств комплексного кепстра. Прежде всего, комплексный кепстр в общем случае отличен от нуля и бесконечен как для положительных, так и для отрицательных значений n, даже если $x(n)$ удовлетворяет принципу причинности, устойчив и имеет конечную протяженность. Далее видно, что комплексный кепстр является затухающей последовательностью, ограниченной сверху:

$$|\hat{x}(n)|<\beta\alpha^{|n|}/|n|, \quad |n|\to\infty, \qquad (7.15)$$

где α — максимальное абсолютное значение величин a_k, b_k, c_k, d_k; β — постоянный сомножитель.

Если $X(z)$ не содержит нулей и полюсов вне единичной окружности (т. е. $b_k=d_k=0$), то

$$\hat{x}(n)=0, \quad n<0. \qquad (7.16)$$

Такие сигналы называются минимально-фазовыми [5]. Общий результат для последовательности (7.16) состоит в том, что такая последовательность полностью определяется действительной частью преобразования Фурье. Таким образом, для минимально-фазовых систем комплексный кепстр определяется лишь логарифмом модуля преобразования Фурье. Это можно легко показать, если вспомнить, что действительная часть преобразования Фурье представляет собой преобразование Фурье от четной части последовательности, т. е. если $\log|X(e^{i\omega})|$ — преобразование Фурье кепстра, то

$$c(n)=[\hat{x}(n)+\hat{x}(-n)]/2. \qquad (7.17)$$

Используя (7.16) и (7.17) легко показать, что

$$\hat{x}(n)=\begin{cases}0, & n<0,\\ c(n), & n=0,\\ 2c(n), & n>0.\end{cases} \qquad (7.18)$$

Таким образом, для минимально-фазовых последовательностей комплексный кепстр можно получить путем вычисления кепстра и последующего использования (7.18). Другой важный результат для минимально-фазовых систем заключается в том, что комплексный кепстр можно вычислить рекуррентно по входному сигналу [1, 2, 5]. Рекуррентная формула имеет вид

$$\hat{x}(n)=\begin{cases}0, & n<0,\\ \log[x(0)], & n=0,\\ \frac{x(n)}{x(0)}-\sum_{k=0}^{n-1}\left(\frac{k}{n}\right)\hat{x}(k)\frac{x(n-k)}{x(0)}, & n>0.\end{cases} \qquad (7.19)$$

Аналогичные результаты можно получить и тогда, когда $X(z)$ не содержит полюсов и нулей, лежащих внутри единичной окружности. Такие сигналы называют максимально-фазовыми. Для этого случая, как это видно из (7.14),

$$\hat{x}(n)=0, \quad n>0. \qquad (7.20)$$

Совместное использование (7.16) и (7.17) дает

$$\hat{x}(n)=\begin{cases}0, & n>0,\\ c(n), & n=0,\\ 2c(n), & n<0.\end{cases} \qquad (7.21)$$

Как и в случае минимально-фазовых последовательностей, здесь также можно получить рекуррентное соотношение для кепстра:

$$\hat{x}(n)=\begin{cases}\dfrac{x(n)}{x(0)}-\sum\limits_{k=n+1}^{0}\left(\dfrac{k}{n}\right)\hat{x}(k)\dfrac{x(n-k)}{x(0)}, & n<0,\\ \log[x(0)], & n=0,\\ 0, & n>0.\end{cases} \tag{7.22}$$

Важным специальным случаем является случай входного сигнала вида

$$p(n)=\sum_{r=0}^{M}\alpha_r\,\delta(n-rN_p), \tag{7.23}$$

т. е. последовательности импульсов. Преобразование $P(z)$ имеет вид

$$P(z)=\sum_{r=0}^{M}\alpha_r\,z^{-rN_p}. \tag{7.24}$$

Из (7.24) видно, что $P(z)$ представляет собой полином по степеням z^{-N_p}, а не z^{-1}, как это было ранее. Этот полином можно представить как результат произведения $(1-az^{-N_p})$ и $(1-bz^{-N_p})$. Легко видеть, что комплексный кепстр отличен от нуля только для целых значений аргумента, кратных N_p. Например, предположим, что

$$p(n)=\delta(n)+\alpha\,\delta(n-N_p), \tag{7.25}$$

где $0<\alpha<1$. Тогда

$$P(z)=1+\alpha\,z^{-N_p} \tag{7.26}$$

и

$$\hat{P}(z)=\log(1+\alpha\,z^{-N_p})=\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\alpha^n}{n}\,z^{-nN_p}. \tag{7.27}$$

Таким образом, $\hat{p}(n)$ представляет собой импульсную последовательность с периодом

$$p(n)=\sum_{r=1}^{\infty}(-1)^{r+1}\frac{\alpha^r}{r}\,\delta(n-rN_p). \tag{7.28}$$

Как будет видно из результатов § 7.2, тот факт, что комплексный кепстр периодической последовательности импульсов также представляет собой периодическую последовательность импульсов, является чрезвычайно важным для анализа речевых сигналов. Однако перед детальным рассмотрением методов гомоморфной обработки речевых сигналов кратко рассмотрим вопросы применения гомоморфных фильтров для обработки сигналов, подвергнутых операции свертки.

7.1.2. Вычислительные аспекты

Математическое описание характеристической системы и обратного преобразования, представленных на рис. 7.5 и 7.6 соответственно, предполагает применение гомоморфной обработки для сигналов, подвергнутых операции свертки. Если ограничиться рассмотрением абсолютно суммируемых сигналов, то область сходимости z-преобразования будет охватывать единичную окружность, т. е. входная последовательность в этом случае будет иметь преобразование Фурье. В этом случае целесообразно заменить z-преобразование (рис. 7.5 и 7.6) преобразованием Фурье. Другими словами, для важного специального случая последовательностей конечной длины математическое представление характеристической системы относительно свертки имеет вид:

$$X(e^{i\omega})=\sum_{n=0}^{N-1}x(n)\,e^{-i\omega n}; \tag{7.29а}$$

$$\hat{X}(e^{i\omega})=\log[X(e^{i\omega})]=\log|X(e^{i\omega})|+i\arg[X(e^{i\omega})]; \tag{7.29б}$$

$$\hat{x}(n)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\hat{X}(e^{i\omega})\,e^{i\omega n}\,d\omega. \tag{7.29в}$$

Уравнение (7.29а) представляет собой преобразование Фурье входной последовательности, соотношение (7.29б) — комплексный логарифм спектра входного сигнала, а уравнение (7.29в) — обратное преобразование Фурье логарифма спектра входного сигнала. Как уже отмечалось, возникают вопросы единственности этого множества уравнений. Для определения комплексного кепстра необходимо однозначно определить логарифм преобразования Фурье. Полезно потребовать, чтобы комплексный кепстр действительной последовательности также являлся действительной последовательностью. Напомним, что для действительных последовательностей действительная часть преобразования Фурье является четной функцией, а мнимая часть — нечетной. Таким образом, если комплексный кепстр должен быть действительной функцией, то логарифм модуля должен быть четной функцией, а фазу следует определить как нечетную функцию ω. Далее можно показать, что достаточным условием единственности комплексного логарифма является требование, чтобы фаза вычислялась как периодическая функция ω с периодом 2π [1, 2] (эти условия непрерывности необходимы также для существования преобразования Фурье от $\hat{X}(e^{i\omega})$). Алгоритм вычисления фазы разработан и подробно описан в [2, 6].

Соотношения (7.29) записаны в форме, затрудняющей их непосредственное применение, поскольку эти соотношения требуют вычисления интегралов. Однако можно аппроксимировать (7.29) с использованием дискретного преобразования Фурье. Дискретное преобразование Фурье (ДПФ) последовательности конечной

длительности идентично дискретизированному преобразованию Фурье для той же последовательности [5]. Таким образом, алгоритм быстрого преобразования Фурье позволяет быстро вычислить ДПФ [5]. В предложенном подходе вычисления кепстра следует заменить все преобразования Фурье соответствующими дискретными преобразованиями Фурье. Результирующие уравнения имеют вид:

$$X_p(k) = \sum_{n=0}^{N-1} x(n)\, e^{-i\frac{2\pi}{N}kn}, \quad N \leqslant k \leqslant N-1; \tag{7.30а}$$

$$\hat{X}_p(k) = \log [X_p(k)], \qquad 0 \leqslant k \leqslant N-1; \tag{7.30б}$$

$$\hat{x}_p(n) = \frac{1}{N}\sum_{k=0}^{N-1} \hat{X}_p(k)\, e^{i\frac{2\pi}{N}kn}, \quad 0 \leqslant n \leqslant N-1. \tag{7.30в}$$

Уравнение (7.30а) представляет собой обратное дискретное преобразование Фурье логарифма ДПФ последовательности конечной длительности. Индекс p указывает на то, что полученная последовательность не является точно эквивалентной комплексному кепстру, определяемому уравнением (7.29). Это обусловлено тем обстоятельством, что комплексный логарифм, используемый при вычислении ДПФ, является дискретным отображением, и, таким образом, результирующее обратное преобразование представляет собой отображение комплексного спектра, искаженного вследствие эффекта наложения частот [1, 2, 5]. Следовательно, комплексный кепстр, полученный с использованием (7.30), связан с действительным комплексным кепстром соотношением

$$\hat{x}_p(n) = \sum_{r=-\infty}^{\infty} \hat{x}(n+rN). \tag{7.31}$$

Вычислительные операции, необходимые для построения характеристической системы относительно свертки, представлены на рис. 7.7*а*.

Комплексный кепстр, как это было показано выше, основан на вычислении комплексного логарифма, а кепстр в его традицион-

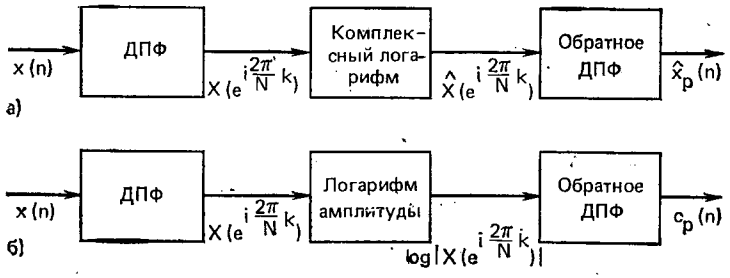


Рис. 7.7. Реализация системы вычисления:
а) комплексного спектра; *б*) кепстра

ном определении основан только на логарифме модуля преобразования Фурье, т. е. кепстр определен в виде

$$c(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log |X(e^{i\omega})|\, e^{i\omega n}\, d\omega, \qquad -\infty < n < \infty. \tag{7.32}$$

Аппроксимация кепстра получается путем вычисления обратного ДПФ логарифма модуля ДПФ входной последовательности:

$$c_p(n) = \frac{1}{N}\sum_{k=0}^{N-1} \log |X_p(k)|\, e^{i\frac{2\pi}{N}kn}, \quad 0 \leqslant n \leqslant N-1. \tag{7.33}$$

Как и ранее, кепстр, полученный с использованием ДПФ, связан с действительным кепстром соотношением

$$c_p(n) = \sum_{r=-\infty}^{\infty} c(n+rN). \tag{7.34}$$

На рис. 7.7*б* показано, как с помощью ДПФ и обратного ДПФ осуществить вычисления, приводящие к (7.34).

Вследствие эффекта наложения частот, присущего дискретному преобразованию Фурье при вычислении кепстра, часто требуется использовать как можно большее значение N. Как показано в [1, 2, 5, 6], большое значение N (т. е. большая частота дискретизации преобразования Фурье) необходимо также и для точного вычисления комплексного логарифма. Однако существование быстрого преобразования Фурье (БПФ) делает возможным использование $N = 512$ или более.

Недавно был предложен иной подход к вычислению кепстра [7] последовательности конечной длительности без нежелательных искажений за счет эффекта наложения частот. Основная идея заключается в непосредственном использовании (7.14) для определения комплексного кепстра через положение нулей (корней) полинома z-преобразования. Этот метод предполагает точное и эффективное вычисление корней полинома сравнительно высокой степени (в задачах обработки речи иногда степень полинома достигает 500). Однако если корни определены с достаточной точностью, то теоретически комплексный кепстр свободен от эффекта наложения частот, присущего вычислительным методам, основанным на реализациях конечной длительности. В [7] приводятся хорошие результаты, полученные для тестовых случаев.

В данном параграфе обсуждались математические и вычислительные аспекты гомоморфных относительно свертки систем. Однако здесь не содержится обсуждение различных частных вопросов и отдельных тонкостей таких систем, поскольку они достаточно полно отражены в [1, 2, 5—7]. Ниже рассмотрено применение гомоморфных относительно свертки систем при анализе речевых сигналов.

7.2. Комплексный кепстр речи

Модели сигналов, с одной стороны, и методы анализа во временной области — с другой, можно объединить и эффективно использовать в теории гомоморфной фильтрации речи. Вспомним, что модель речеобразования обязательно состоит из линейной системы с медленно изменяющимися во времени параметрами и сигнала возбуждения в виде последовательности импульсов или белого шума. Поэтому короткий сегмент вокализованного речевого сигнала целесообразно рассматривать как результат воздействия сигнала возбуждения в виде последовательности импульсов на линейную систему с постоянными параметрами. Аналогично короткий сегмент невокализованного сигнала можно представить как результат возбуждения линейной системы с постоянными параметрами случайным шумом. Короткий сегмент вокализованной речи можно представить в виде

$$s(n) = p(n) * g(n) * v(n) * r(n) = p(n) * h_v(n) = \sum_{r=-\infty}^{\infty} h_v(n - r N_p), \quad (7.35)$$

где $p(n)$ — периодическая импульсная последовательность с периодом N_p отсчетов и $h_v(n)$ — импульсная характеристика линейной системы, отражающая эффект формы источника возбуждения $g(n)$, импульсную характеристику речевого тракта $v(n)$ и импульсную характеристику излучения $r(n)$. Аналогично для невокализованного сегмента сигнала получаем

$$s(n) = u(n) * v(n) * r(n) = u(n) * h_u(n), \quad (7.36)$$

где $u(n)$ — сигнал возбуждения в виде случайного шума; $h_u(n)$ — импульсная реакция системы, объединяющая воздействие речевого тракта и излучения. Для случая вокализованной речи передаточная функция линейной системы имеет вид

$$H_v(z) = G(z) V(z) R(z). \quad (7.37)$$

Для невокализованной речи получаем

$$H_u(z) = V(z) R(z). \quad (7.38)$$

Кратко рассмотрим природу различных компонент в (7.37) и (7.38). Из результатов, приведенных в гл. 3, следует, что передаточная функция речевого тракта имеет вид

$$V(z) = \frac{A z^{-M} \prod_{k=1}^{M_i} (1 - a_k z^{-1}) \prod_{k=1}^{M_o} (1 - b_k z)}{\prod_{k=1}^{N_i} (1 - c_k z^{-1})}. \quad (7.39)$$

Для вокализованной речи кроме носовых звуков адекватная модель содержит только полюсы, т. е. $a_k = 0$, $b_k = 0$ для всех k. Для носовых

звуков и невокализованной речи необходимо рассматривать как полюсы, так и нули. Некоторые нули передаточной функции могут лежать вне единичного круга. Для устойчивости системы все ее полюсы должны располагаться внутри единичного круга. Таким образом, поскольку $v(n)$ действительно, полюсы и нули могут возникать лишь в виде комплексно-сопряженных пар. Эффект излучения, результатом которого, как это показано в гл. 3, является подъем в области высоких частот, может быть приближенно смоделирован как

$$R(z) \approx 1 - z^{-1}. \quad (7.40)$$

Наконец, для вокализованной речи импульс возбуждения имеет конечную протяженность. Таким образом, $G(z)$ будет иметь вид

$$G(z) = B \prod_{k=1}^{L_i} (1 - \alpha_k z^{-1}) \prod_{k=1}^{L_o} (1 - \beta_k z), \quad (7.41)$$

где нули α_k и β_k могут быть как внутри, так и вне единичной окружности.

Используя рассмотренные выше модели и результаты 7.1.2, приступим к анализу комплексного кепстра короткого сегмента речевого сигнала (детальное исследование содержится в [8]). Для вокализованного речевого сигнала полный вклад речевого тракта, источника возбуждения и излучения в общем случае может оказаться неминимально-фазовым, что приводит к ненулевым значениям кепстра в области отрицательного времени. Из (7.14) следует, что комплексный кепстр быстро затухает с ростом n. Кроме того, отметим, что вклад в комплексный кепстр от периодического возбуждения проявится в наличии импульсов в точках, кратных периоду возбуждения. Пример анализа (рис. 7.8) иллюстрирует основные особенности вокализованного речевого сигнала. На рис. 7.8*а* показан сегмент вокализованного сигнала, взвешенный с окном Хемминга. На рис. 7.8*б* представлен логарифм модуля дискретного преобразования Фурье. В этой функции имеется периодическая компонента, обусловленная периодическим характером входного сигнала. На рис. 7.8*в* представлен разрывной характер главного значения фазы, а на рис. 7.8*г* — фазовая кривая, лишенная разрывов. Результат преобразования Фурье в комплексный кепстр кривых рис. 7.8*б* и 7.8*г* представлен на рис. 7.8*д*. Отметим наличие пиков в положительном и отрицательном времени и быстрое затухание компонент в области малых времен, что обусловлено совместным воздействием речевого тракта, источника возбуждения и излучением. Кепстр, являющийся обратным преобразованием Фурье логарифма амплитуды модуля спектра, показан на рис. 7.8*е*. В данном случае сохранены все основные особенности комплексного кепстра, как это и предполагалось, поскольку он является четной частью комплексного кепстра.

Последовательность графиков рис. 7.8 показывает, как можно использовать гомоморфную фильтрацию для анализа речевого

сигнала. Прежде всего отметим, что импульс в кепстре, обусловленный квазипериодическим возбуждением, оказывается отделенным от остальных компонент. Это приводит к соответствующей системе гомоморфной фильтрации речевого сигнала, представленной на рис. 7.9. Сегмент речевого сигнала взвешивается с некоторым окном, т. е. кепстр вычисляется так, как это обсуждалось в 7.1.3, и требуемые компоненты кепстра выделяются с использованием «окна по кепстру» $l(n)$. Этот вид фильтрации иногда называют «частотно-инвариантной линейной фильтрацией». В результате взвешенный комплексный кепстр подвергается обратному преобразованию для получения требуемых компонент. Это показано на рис. 7.10. На рис. 7.10*а* и *б* показаны логарифм модуля и фаза, полученные в процессе использования процедуры обратного преобразования в случае, когда

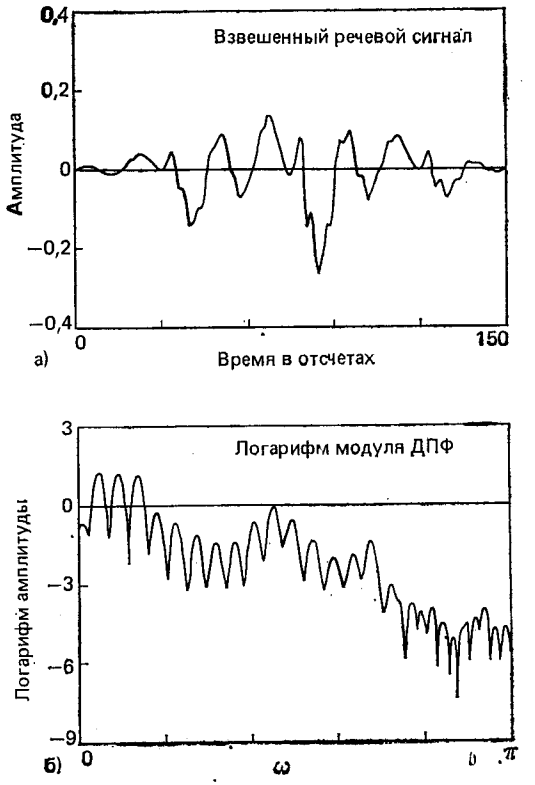
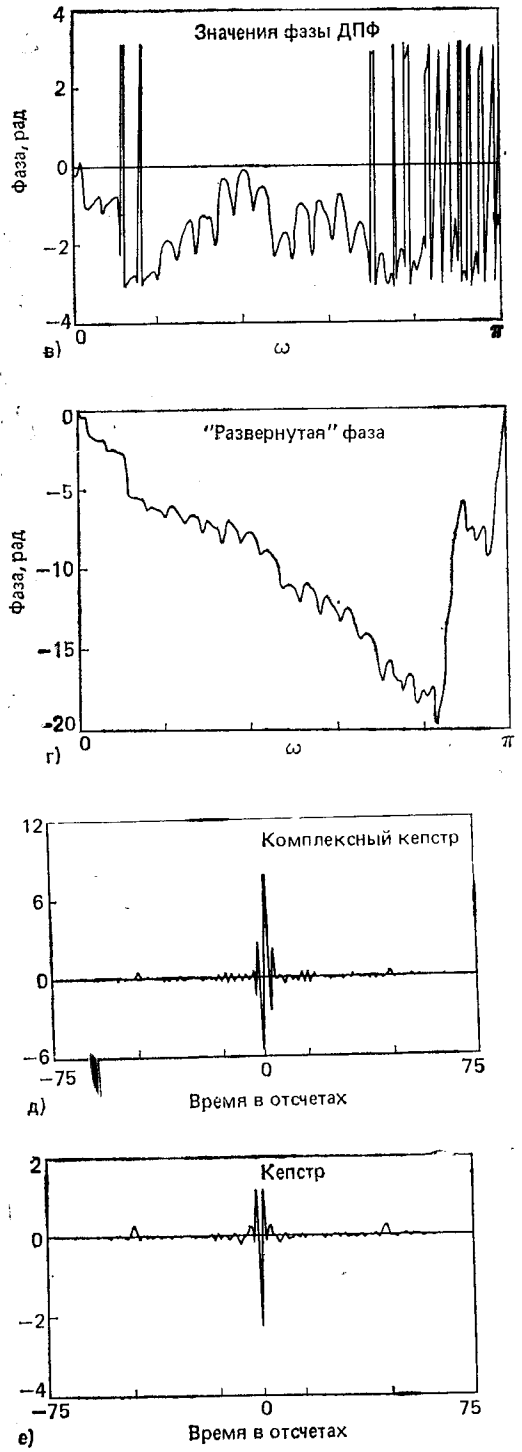


Рис. 7.8. Гомоморфный анализ вокализованной речи:
а) взвешенный речевой сигнал; *б*) логарифм модуля кратковременного преобразования Фурье; *в*) значения фазы; *г*) «развернутая» фаза; *д*) комплексный кепстр; *е*) кепстр

$$l(n)=\begin{cases}1, & |n|<n_0,\\ 0, & |n|\geqslant n_0,\end{cases} \tag{7.42}$$

где n_0 выбрано меньшим, чем период основного тона N_0. Соответствующий выходной сигнал показан на рис. 7.10*в*. (Заметим, что постоянный фазовый сдвиг π радиан был устранен при вычислении кепстра.) Этот сигнал аппроксимирует импульсную реакцию $h_v(n)$, определяемую (7.35). Если выбрать $l(n)$ таким образом, чтобы восстановить компоненты возбуждения, т. е.

$$l(n)=\begin{cases}0, & |n|<n_0,\\ 1, & |n|\geqslant n_0,\end{cases} \tag{7.43}$$



получим результаты (рис. 7.10*г*, *д*, *е*), вычисленные для логарифма модуля, фазы и выходного сигнала. Выходной сигнал аппроксимирует импульсную последовательность возбуждения, амплитуды которой затухают в соответствии с весами окна Хемминга, примененного при взвешивании входного сигнала.

Для полноты иллюстраций применения гомоморфного анализа к обработке речевого сигнала рассмотрим случай анализа невокализованного сегмента речевого сигнала, показанного на рис. 7.11. На рис. 7.11*а* представлен речевой сигнал, взвешенный с окном Хемминга. На рис. 7.11*б* изображен логарифм модуля спектра, а на рис. 7.11*в* — кепстр. Отметим случайные флуктуации в логарифме модуля спектра. Это связано с тем, что возбуждение в данной ситуации случайно и преобразование Фурье короткого сегмента содержит случайную компоненту. В этом случае результаты малочувствительны к вычислению фазы. Сказанное подтверждается рис. 7.11*в*, на котором отсутствуют пики возбуждения, возникавшие в случае вокализованного сигнала, однако область малых времен в кепстре содержит информацию о $H_u(e^{i\omega})$. Это видно из рис. 7.11*г*, где показан ло-

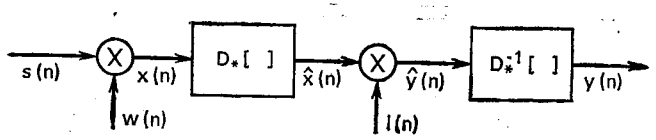
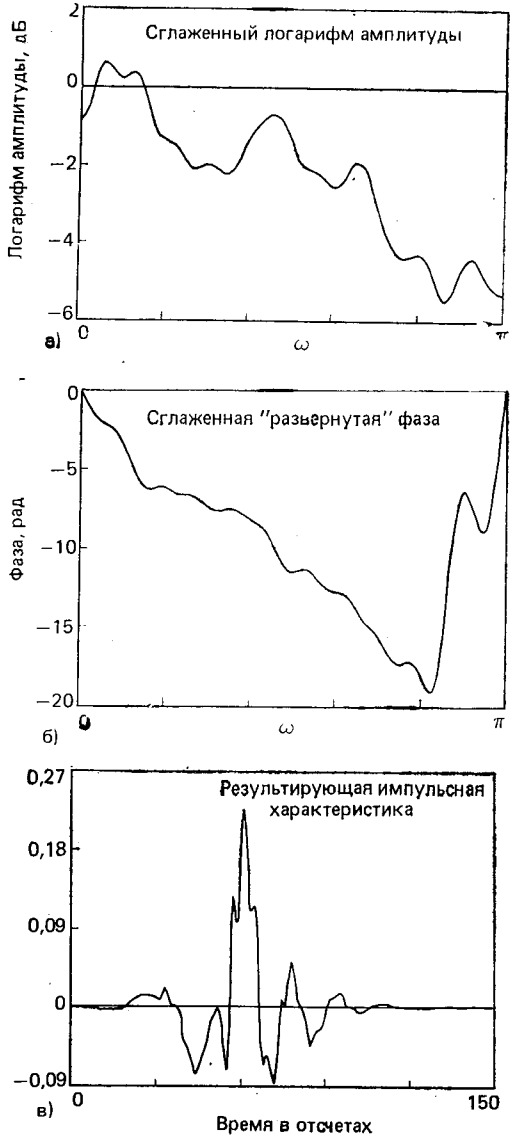


Рис. 7.9. Реализация системы гомоморфной фильтрации речи

тарифм модуля, полученный применением кепстрального окна (7.42) к кепстру 7.11*в*.

Предшествующее обсуждение и примеры показывают, что с помощью гомоморфной фильтрации можно выделить ряд важных компонент речевого сигнала. Тем не менее этот подход не используется достаточно широко, поскольку в ряде речевых приложений полное разложение сигнала не требуется. Чаще сталкиваются с необходимостью оценки таких параметров, как период основного тона и частоты формант. Для этих целей кепстральный анализ весьма эффективен. Таким образом, для большинства задач обработки речи можно избежать обременительных вычислений фазы. Отметим, например, сравнивая 7.8*е* и 7.11*в*, что кепстр позволяет отделять вокализованную речь от невокализованной и, кроме того, период основного тона для вокализованного речевого сигнала хорошо просматривается на кепстральных диаграммах. Частоты формант также можно определить с использованием логарифма модуля передаточной функции речевого тракта, которая вычисляется по кепстру с помощью кепст-



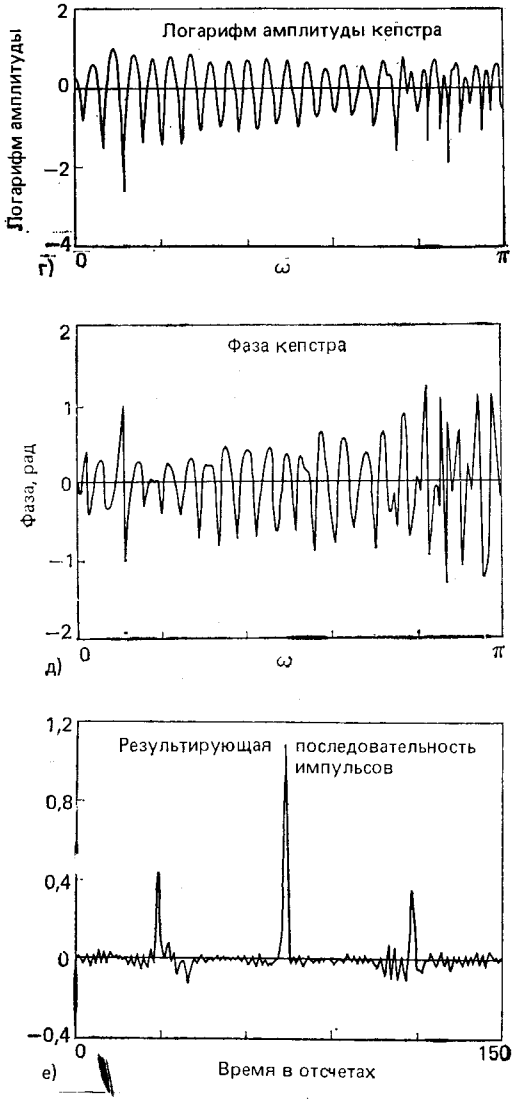


Рис. 7.10. Гомоморфная фильтрация вокализованной речи:
а) и *б*) оценка амплитуды и фазы $H_v(e^{j\omega})$; *в*) оценка импульсной характеристики; *г*) и *д*) оценка амплитуды и фазы $P(e^{j\omega})$; *е*) оценка $p(n)$

рального окна (7.42). В последующих разделах главы кепстральный подход применяется для оценивания формантных частот и периодов основного тона, а также для построения вокодерной системы передачи речевого сигнала.

7.3. Оценивание основного тона

Рисунки 7.8*е* и 7.11*в* позволяют сделать вывод о возможности оценивания периода основного тона с использованием гомоморфной обработки. Было отмечено, что для вокализованного сегмента речи пик в кепстре возникает при задержке, соответствующей периоду основного тона. Для невокализованного сегмента такие пики в кепстре не возникают. Это свойство кепстра может быть использовано для классификации вокализованный/невокализованный и для периода основного тона на вокализованной речи.

Метод оценивания основного тона на основе кепстрального анализа, разработанный для нереального масштаба времени, чрезвычайно прост. Кепстр, полученный в соответствии с результатами 7.1.3, исследуется с целью отыскания пика в области возможных значений основного тона. Если пик в кепстре превышает порог, то сегмент классифицируется как вокализованный, а координата пика дает хорошую оценку периоду основного

тона. Если максимум кепстра не превышает порога, то сегмент классифицируется как невокализованный. Изменение во времени типа возбуждения и периода основного тона можно оценить с использованием зависящего от времени кепстра, что достигается на основе вычисления зависящего от времени преобразования Фурье. Обычно кепстр вычисляется 1 раз через каждые 10—20 мс, поскольку в нормальной речи параметры возбуждения не изменяются быстрее.

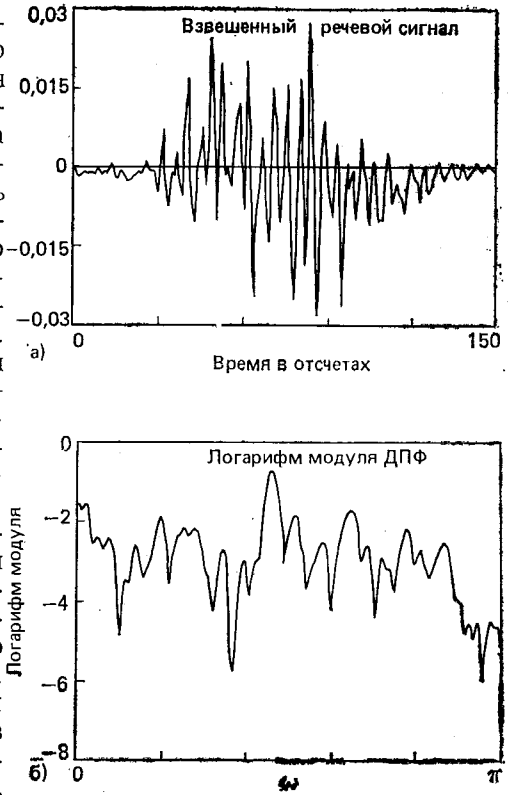
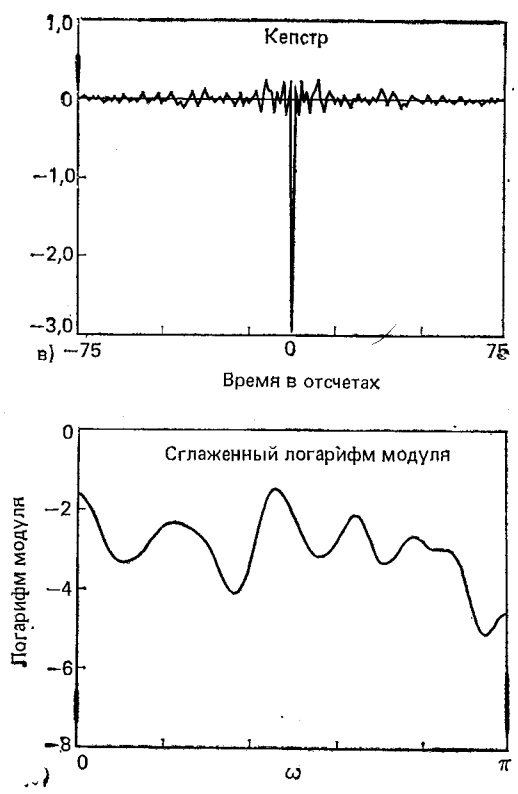


Рис. 7.11. Гомоморфный анализ невокализованной речи:
а) взвешенный речевой сигнал; *б*) логарифм модуля кратковременного преобразования Фурье; *в*) кепстр; *г*) оценка $H_u(e^{j\omega})$

На рис. 7.12 и 7.13 показан пример, полученный Ноллом [9], который первым описал процедуру оценивания периода основного тона на основе кепстра. На рис. 7.12 показана серия логарифмических спектров и соответствующих им кепстров для мужского голоса. Кепстры, изображенные на данном рисунке, представляют собой квадратный корень из $c(n)$, как они были определены выше. В этом примере частота дискретизации на входе составляла 10 кГц. Окно Хемминга протяженностью 40 мс (400 отсчетов) перемещалось с шагом 10 мс, т. е. логарифмические спектры слева и кепстры справа вычислялись через каждые 10 мс. Как следует из рис. 7.12, первые семь 40-миллисекундных интервалов соответствуют невокализованному сигналу, а остальные интервалы — вокализованной речи, причем период основного тона возрастает с течением времени, т. е. частота основного тона падает. На рис. 7.13 показан пример анализа женского голоса. В этом случае речевой сигнал, соответствующий последовательности кепстров и спектров, оказывается вокализованным в начале и невокализованным в конце. Легко видеть, что в конце вокализованного сегмента период основного тона удваивается, что нередко бывает по окончании вокализованного сегмента. Сравнение рис. 7.12 и 7.13 показывает, что для женского голоса частота основного тона гораздо выше, чем для мужского.



Эти два примера, хорошо иллюстрирующие результаты анализа основного тона для речевых сигналов, могут привести к предположению о том, что на основе гомоморфного анализа можно построить очень простой и эффективный алгоритм выделения основного тона и классификации речи на вокализованную/невокализованную. К сожалению, как это зачастую бывает при анализе речи, имеется ряд практических вопросов и трудностей, которые должны быть решены при разработке алгоритма на основе кепстра. Нолл [9] описал одну из возможных схем анализа речи на основе кепстра. Но имеется и ряд других схем, которые успешно используются для этих целей. Вместо того чтобы описывать здесь детально каждую из известных схем, целесообразнее остановиться на тех неизбежных трудностях, которые возникают при построении кепстральных анализаторов основного тона.

Во-первых, наличие выброса в кепстре в диапазоне 3—20 мс очень точно указывает на то, что данный сегмент является вокализованным. Однако отсутствие пика или наличие слабого пика не означает, что данный сегмент является невокализованным. Амплитуда или даже просто существование пика в кепстре зависит от целого ряда факторов, включая длину окна, используемого для взвешивания входного сигнала, и формантной структуры самого сигнала. Легко показать (см. задачу 7.10), что наибольшая амплитуда пика в кепстре равна единице. Это достигается только в случае абсолютного совпадения периодов основного тона. Это, конечно, совершенно не достижимо в реальном случае, даже если использовать прямоугольное временно́е окно, включающее целое число периодов. Прямоугольные временны́е окна применяются весьма редко вследствие худших результатов, даваемых ими при оценивании спектра. В случае, например, окна Хемминга очевид-

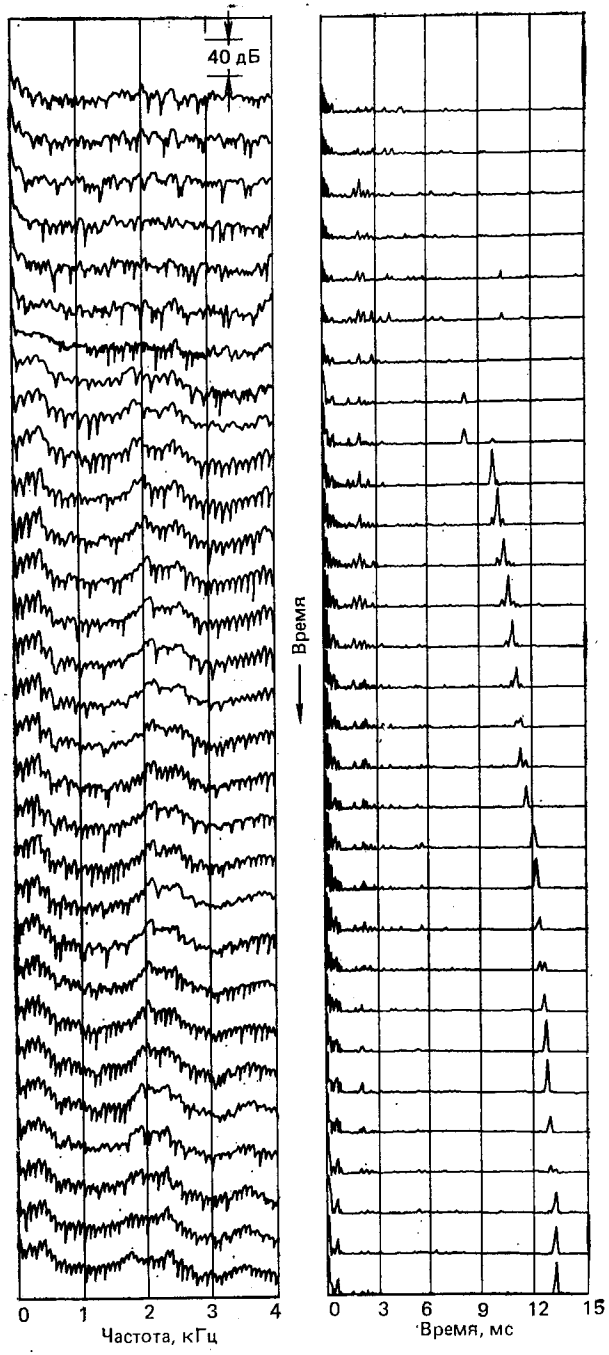


Рис. 7.12. Набор логарифмов спектра и кепстров для мужского голоса [9]

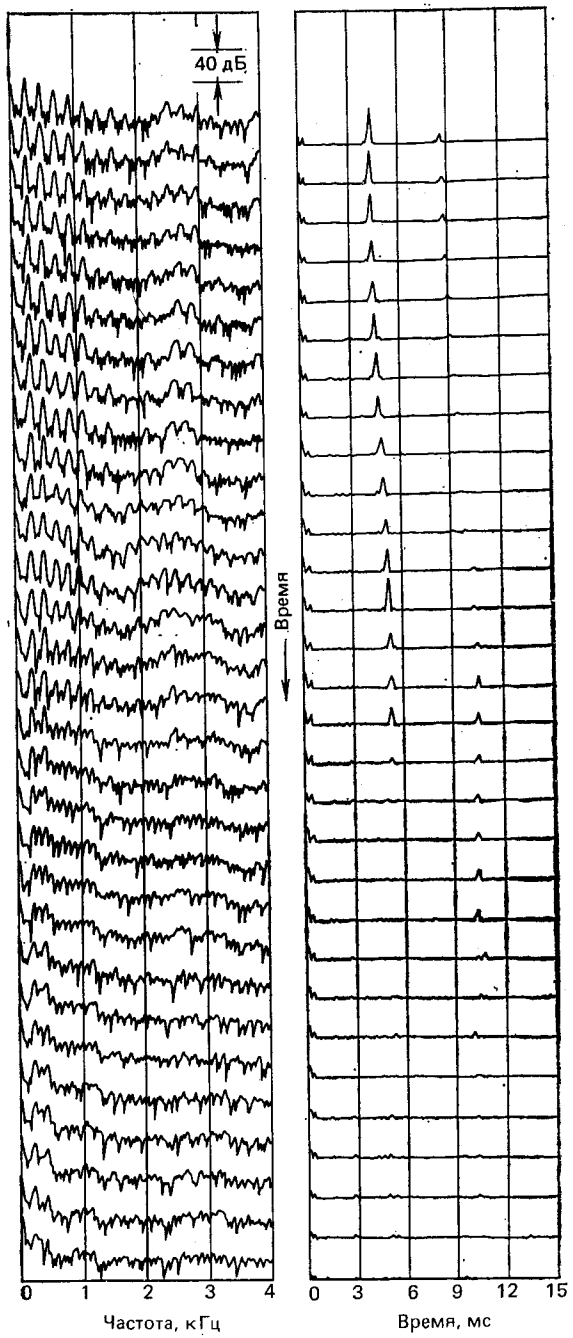


Рис. 7.13. Набор логарифмов спектра и кепстров для женского голоса

но, что как протяженность окна, так и его относительное расположение по отношению к речевому сигналу будут оказывать значительное влияние на величину наибольшего пика в кепстре.

Как крайний случай предположим, что окно имеет протяженность менее двух периодов основного тона. Очевидно, что при этом трудно ожидать точного оценивания периодичности по спектру или кепстру сигнала. Таким образом, протяженность окна может оказаться такой, что с учетом уменьшения амплитуды данных к границам выборки, по крайней мере, два периода основного тона пропадут во взвешенных данных. Для мужской речи с низкой частотой основного тона требуется окно порядка 40 мс. Для голосов с более высокой частотой основного тона требуются пропорционально меньшие окна. Желательно, конечно, выбирать окно настолько малым, насколько это возможно, чтобы избежать значительных изменений параметров сигнала на протяжении используемого сегмента. Чем длиннее окно, тем значительнее изменения параметров в пределах окна и тем больше отклонение от принятой модели анализа. Один из способов выбора окна, при котором оно было и не слишком длинным и не слишком коротким, состоит в адаптации длины окна с учетом предшествующих (или возможно среднего значения) оценок периодов основного тона [10, 11].

Другая причина, по которой сигнал может сильно отличаться от описываемого моделью, заключается в чрезмерном ограничении полосы. Ярким примером подобной неадекватности может служить синусоидальный сигнал. В логарифме спектра такой сигнал даст только один пик. Поскольку в спектре нет периодических колебаний, в кепстре не будет пиков. В речевом сигнале вокализованные сегменты обычно очень узкополосны с плохо выраженной гармонической структурой на частотах выше нескольких сотен герц. В этом случае пики в кепстре отсутствуют. К счастью, для любых значений периода, за исключением самых коротких, область, в которой возникают пики в кепстре, не содержит других компонент, кроме основного тона. Таким образом, для определения положения импульса основного тона можно использовать достаточно низкий порог (порядка 0,1).

При правильно подобранной протяженности окна на входе положение и амплитуда импульса кепстра обеспечивают в большинстве случаев хорошую оценку периода основного тона и классификации тон/шум. В тех случаях, когда кепстральный анализ не позволяет точно ответить на вопрос о наличии импульсов основного тона и значении периода, для вынесения окончательного решения можно привлечь дополнительную информацию о виде функции среднего числа переходов через нуль, энергии сигнала или улучшить оценки с помощью сглаживания [11]. Дополнительная логика при реализации устройств оценивания на основе кепстра требует усложнения аппаратуры. Эта часть общей схемы выделения основного тона вносит незначительную долю в общие вычислительные затраты, но вместе с тем оказывается весьма полезной.

7.4. Оценивание формант

На основе примеров § 7.2 можно сделать вывод, что часть кепстра в области малых времен в основном содержит информацию о речевом тракте, источнике возбуждения и излучении, в то время как в области больших времен заключена информация о сигнале возбуждения. Это свойство использовалось для классификации сегментов и оценивания периода основного тона путем использования части кепстра в области больших времен. Примеры § 7.2 указывают также и на метод получения отклика речевого тракта на основе кепстра. Действительно, отметим, что «сглаженные» логарифмы модуля (см. рис. 7.10*а* и *г*) получаются путем взвешивания кепстра. Эти сглаженные спектры отражают резонансную структуру речевого сигнала, т. е. пики в спектре соответствуют формантным частотам. Это означает, что последние можно оценить по положению максимумов в «кепстрально сглаженном» логарифмическом спектре.

Рассмотрим модель речеобразования, представленную на рис. 7.14. Эта чрезвычайно упрощенная модель описывает вокализо-

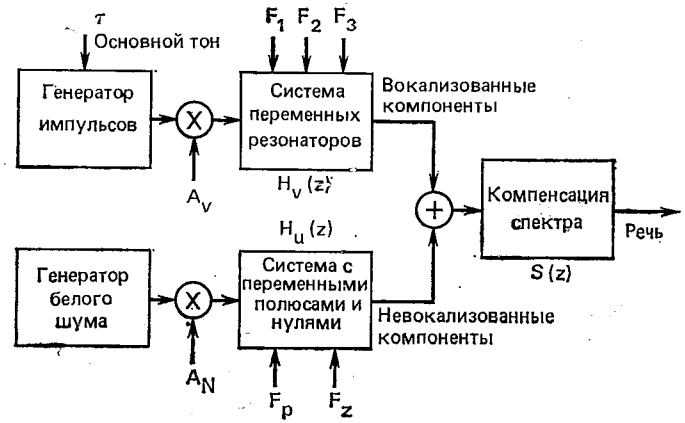


Рис. 7.14. Цифровая модель речеобразования

ванный речевой сигнал с помощью периода и амплитуды основного тона и трех первых формант, а невокализованную речь с помощью амплитудного значения и положения единственного нуля и полюса. Для компенсации свойств в области высоких частот используется дополнительный неперестраиваемый фильтр. Все перечисленные параметры, конечно, изменяются с течением времени. Метод оценивания этих параметров основан на вычислении кепстрально сглаженного логарифма спектра через каждые 10—20 мс. По кепстру производится анализ на вокализованность сегмента и определяется положение максимумов. Если сегмент вокализован, то по кепстру определяются период основного тона и первые три формантные частоты, которые вычисляются по систе-

ме логических правил, учитывающих применяемую модель [11, 12]. В случае невокализованного сегмента полюс определяется в точке максимума спектральной плотности, а нуль — в том месте, где сохраняется относительная амплитуда между максимумом и минимумом [12].

Иллюстрация использования метода оценивания периода основного тона и формантных частот для вокализованной речи представлена на рис. 7.15. Слева показана последовательность кепстров, вычисленных через каждые 20 мс, справа — последова-

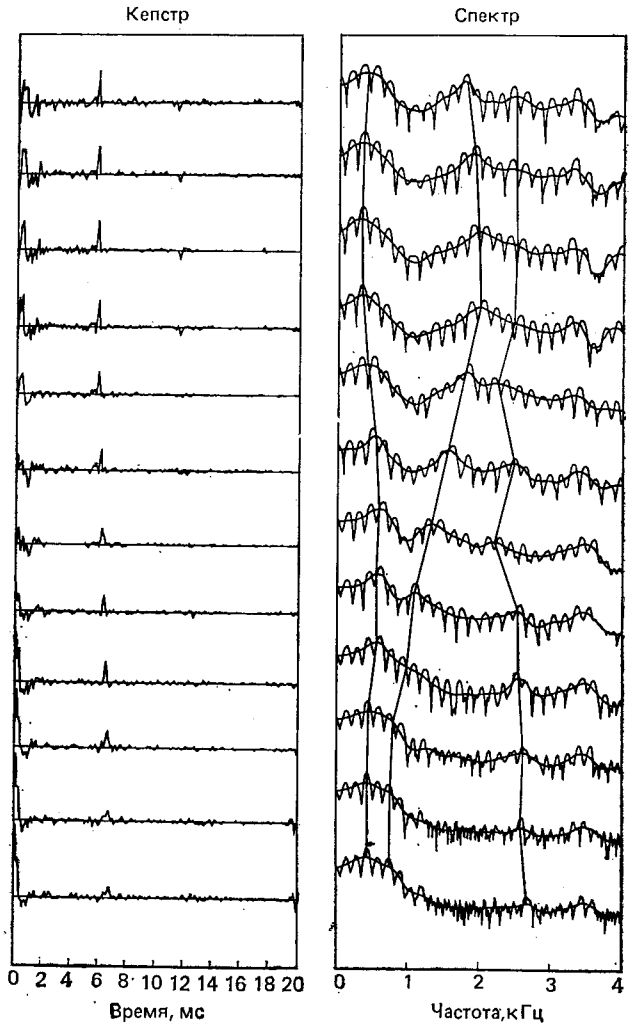


Рис. 7.15. Автоматическое оценивание формантных траекторий по кепстрально сглаженному логарифму спектра

тельность логарифмов спектров с соответствующими сглаженными спектральными оценками, полученными на основе кепстра. Линиями соединены максимумы, выделенные с использованием алгоритма [11] для трех первых формант. Из рис. 7.15 видно, что две первые форманты сблизились настолько, что они не являются уже двумя отдельными максимумами. Эту ситуацию можно обнаружить и отделить пики, если вычислить z-преобразование $H_v(z)$ по контуру, проходящему вблизи полюсов. Вычисление производится с помощью алгоритма спектрального анализа, названного острым z-преобразованием (CZT) [13]. Пример улучшенного разделения показан на рис. 7.16.

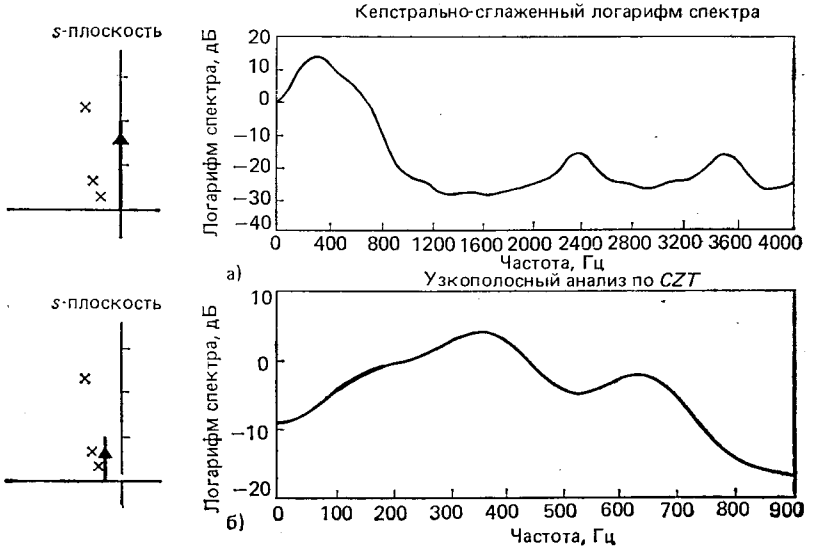


Рис. 7.16. Улучшение разрешения по частоте с помощью алгоритма CZT [11]

Другой подход к оцениванию формант с использованием кепстра предложил Оливье [14], который применил итеративную процедуру, напоминающую метод анализа через синтез, для определения положения полюсов передаточной функции, подгоняемой к сглаженному спектру по критерию минимума среднего квадрата ошибки.

Речевой сигнал можно синтезировать по формантам и периоду основного тона, как это было описано выше, путем простого управления моделью (см. рис. 7.14) с оцененными параметрами. В этом случае для вокализованной речи передаточная функция модели имеет вид

$$V(z)=\prod_{k=1}^{4}\frac{1-2e^{-\alpha_k T}\cos(2\pi F_k T)+e^{-2\alpha_k T}}{1-2e^{-\alpha_k T}\cos(2\pi F_k T)z^{-1}+e^{-2\alpha_k T}z^{-2}}. \qquad (7.44)$$

Это соотношение описывает каскадно соединенные цифровые ре-

зонаторы с единичным коэффициентом усиления на нулевой частоте так, что амплитуда речевого сигнала определяется амплитудой управления A_v. Первые три формантные частоты F_1, F_2 и F_3 изменяются во времени, в то время как F_4 зафиксирована на частоте 4000 Гц и $T=0{,}0001$ с (т. е. частота дискретизации 10 кГц). Полосы формант α_k также зафиксированы на уровне средних для речевого сигнала значений. Неадаптивный фильтр для компенсации влияния источника возбуждения и излучения имеет следующую передаточную функцию:

$$S(z)=\frac{(1-e^{-aT})(1+e^{-bT})}{(1-e^{-aT}z^{-1})(1+e^{-bT}z^{-1})}, \qquad (7.45)$$

где коэффициенты a и b выбраны так, чтобы обеспечивать хорошее приближение спектра. Целесообразно выбрать a и b равными 400π и 5000π соответственно. Более точные значения коэффициентов для данного диктора могут быть получены с использованием спектра, усредненного за большой интервал времени.

Для невокализованного речевого сигнала влияние речевого тракта моделировалось линейной системой с передаточной функцией

$$V(z)=\frac{(1-2e^{-\beta T}\cos(2\pi F_p T)+e^{-2\beta T})(1-2e^{-\beta T}\cos(2\pi F_z T)z^{-1}+\rightarrow}{(1-2e^{-\beta T}\cos(2\pi F_p T)\ z^{-1}+e^{-2\beta T}\ z^{-2})(1-2e^{-\beta T}\times\rightarrow}$$
$$\frac{\rightarrow+e^{-2\beta T}\ z^{-2})}{\rightarrow\times\cos(2\pi F_z T)+e^{-2\beta T})},$$

где F_p взята как максимальное значение сглаженного логарифма спектра на частотах выше 1000 Гц, а F_z удовлетворяет эмпирической формуле

$$F_z=(0{,}0065F_p+4{,}5-\Delta)(0{,}014F_p+28). \qquad (7.46)$$

Здесь

$$\Delta=20\log_{10}|H[e^{i2\pi F_p T}]|-20\log_{10}|H(e^{i0})| \qquad (7.47)$$

обеспечивает сохранение соответствующего соотношения амплитуд [12]. То, что такая относительно простая модель отражает все основные свойства спектра речевого сигнала, иллюстрируется на рис. 7.17 и 7.18, где сравниваются сглаженные логарифмы спектров и результаты, даваемые моделью, определяемой рис. 7.14 и соотношениями (7.44)—(7.47) как для вокализованных, так и для невокализованных речевых сигналов соответственно. В качестве примера речевого сигнала, синтезированного с использованием описанной модели, может служить сигнал, представленный на рис. 7.18. В верхней части рисунка изображены траектории параметров, построенные по речевому сигналу, спектрограмма которого представлена на рис. 7.19*б*. На рис. 7.19*в* показана спектрограмма синтезированного сигнала, полученного с использованием модели рис. 7.14 на основе параметров рис. 7.19*а*. Очевидно, что в синтезированном сигнале хорошо сохранились основные

черты исходной речи. Фактически, несмотря на чрезвычайно грубый способ описания, синтетическая речь очень разборчива и сохраняет многие черты индивидуальности диктора. В действительности период основного тона и формантные частоты оцениваются

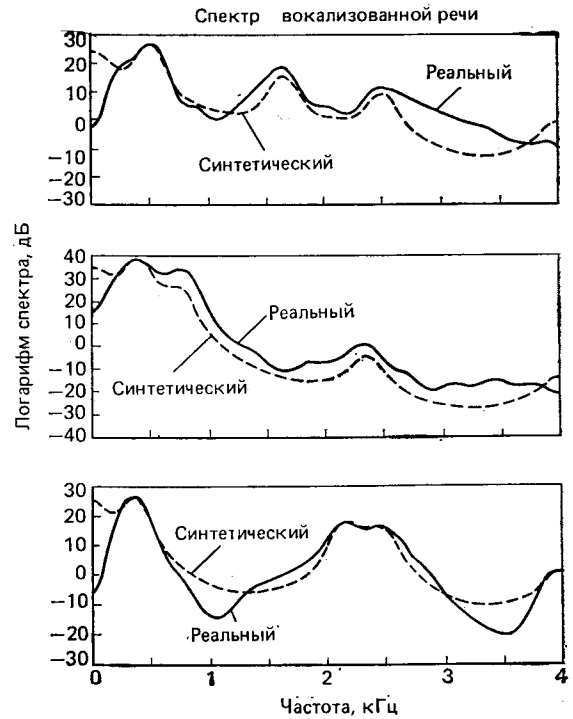


Рис. 7.17. Сравнение кепстрально сглаженного спектра и спектра модели для вокализованного сигнала

с использованием алгоритма, основанного на результатах обширных исследований в области верификации диктора (см. гл. 8 и § 9.2).

Ценное свойство рассмотренного представления заключается в возможности получения очень малых требуемых скоростей передачи. Полная система анализа—синтеза, основанная на этом представлении (формантный вокодер), показана на рис. 7.20. Параметры модели оценивались 100 раз в секунду и фильтровались для устранения шума. Частота дискретизации понижалась до удвоенной частоты среза фильтра и затем параметры квантовались. При синтезе каждый параметр интерполировался с целью получения частоты дискретизации 100 Гц и использовался в синтезаторе так, как это показано на рис. 7.14.

Для выявления подходящего множества параметров были проведены аудиторные испытания. Анализатор и синтезатор соединялись друг с другом для получения образцов сигнала. Затем

параметры подвергались фильтрации с помощью фильтров нижних частот и определялась такая полоса фильтра, при которой отсутствуют слышимые различия между синтезированными сигналами с отфильтрованными и неотфильтрованными параметрами.

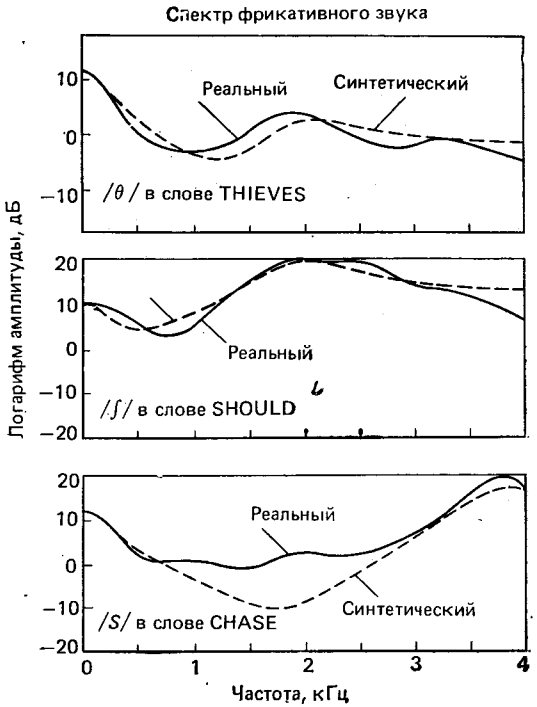


Рис. 7.18. Сравнение кепстрально сглаженного спектра и спектра модели для невокализованного сигнала

Обнаружено, что частота фильтра может быть выбрана равной 16 Гц без заметных различий в качестве. Полученные параметры затем дискретизировались с частотой около 33 Гц (прореживание 3 : 1). Затем был проведен эксперимент по определению требуемой скорости передачи. Формантные параметры и период основного тона квантовались с использованием линейного квантователя (подобранного для каждого параметра), а амплитудные параметры — с использованием логарифмического квантователя. Результаты данного эксперимента при анализе их с точки зрения качества восприятия представлены в табл. 7.1. При использовании частоты дискретизации 33 Гц и данных табл. 7.1 обнаружено, что для полностью вокализованной фразы качество сигнала по сравнению с синтезированным сигналом без кватования не снижается до скоростей порядка 600 бит/с. (Отметим, что для адекватного описания переходов тон/шум требуется передача соответствующего признака с частотой 100 Гц.)

На рис. 7.21*а* показан пример траектории параметров, оцененных по исходному речевому сигналу с частотой 100 раз в секунду. На рис. 7.21*б* представлены те же параметры после сглаживания с использованием КИХ-фильтра нижних частот с полосой

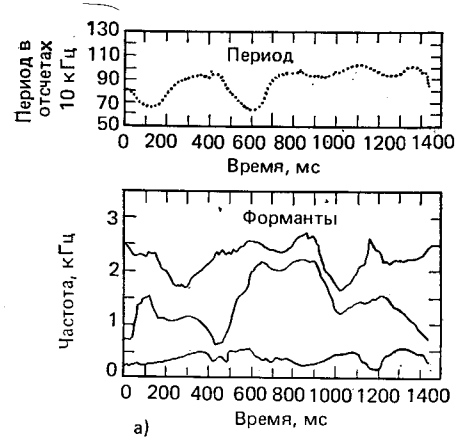
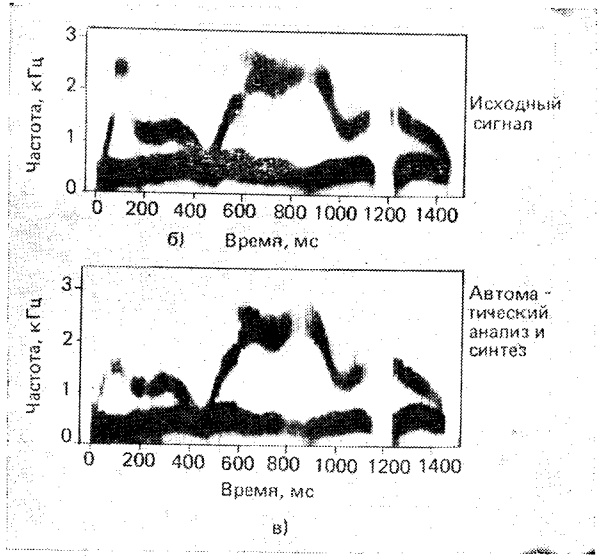


Рис. 7.19. Автоматический анализ и синтез фразы «We were away a year ago»:
а) период основного тона и формантные траектории, построенные на ЭВМ; *б*) широкополосные спектрограммы исходного сигнала; в) широкополосная спектрограмма синтетической речи [11]



16 Гц. На рис. 7.21*в* показаны траектории параметров после прореживания, квантования и интерполяции с коэффициентом 3. Хотя между трактрнями во всех трех случаях и имеется видимое различие, но различие между образцами по восприятию незначительно или вовсе отсутствует. Это представление сигнала использовано для экспериментов по синтезу речи в системах машинного ответа [16] (см. 9.1.3).

Таблица 7.1

Результаты анализа формантного вокодера по слуховому восприятию

Параметр	Необходимое количество двоичных единиц на отсчет
τ	6
F_1	3
F_2	4
F_3	3
$\log [A_V]$	2

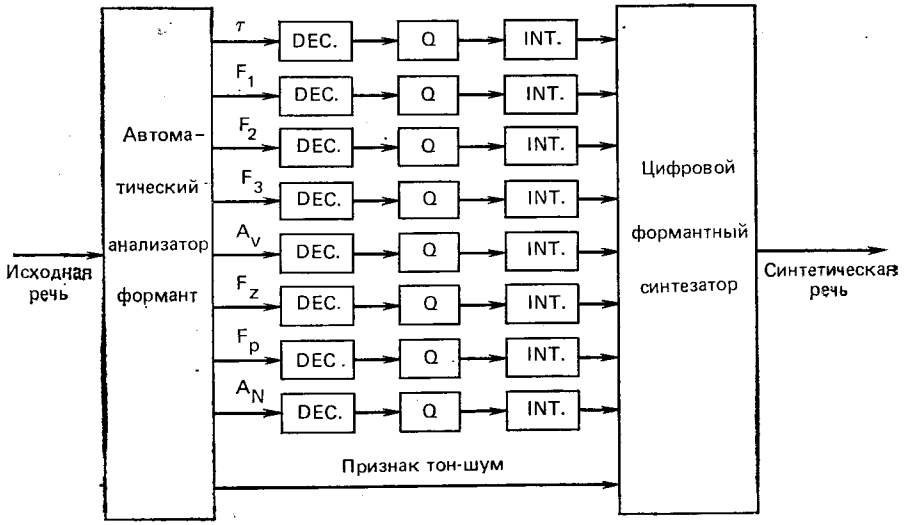


Рис. 7.20. Структурная схема формантного вокодера: DEC. — прореживание; INT. — интерполяция

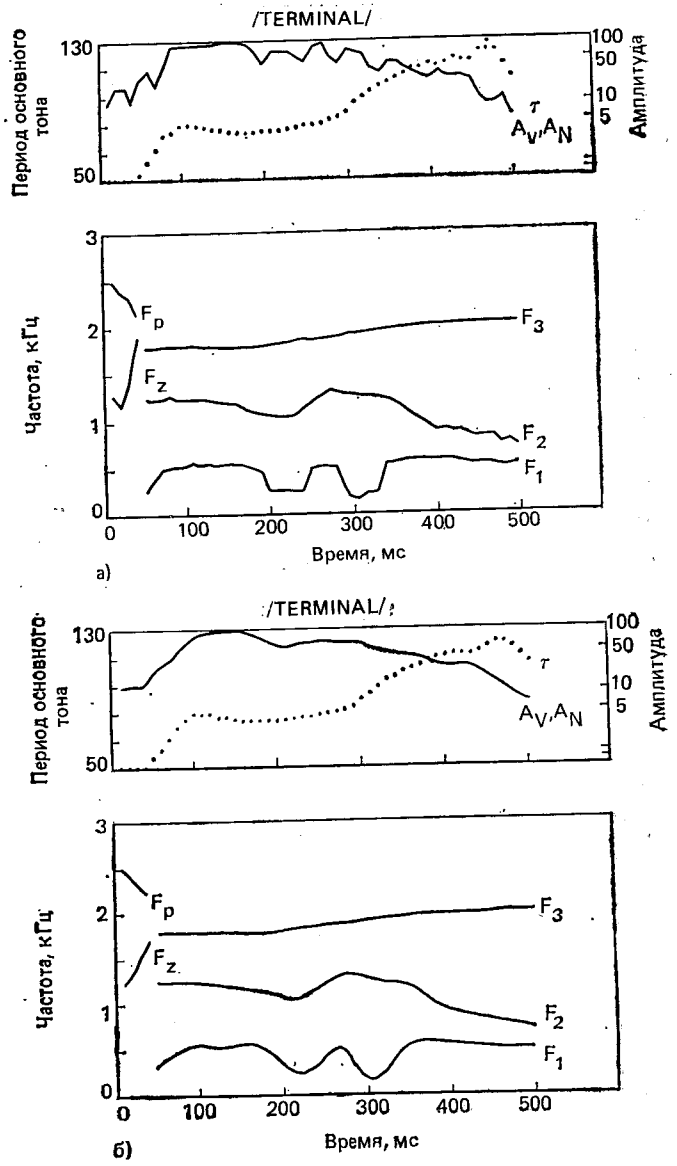
7.5. Гомоморфный вокодер

Как было показано выше, текущая гомоморфная обработка речевого сигнала приводит к весьма удобному описанию, где основные параметры сигнала отделены друг от друга, т. е. информация о сигнале возбуждения расположена в области больших времен, а информация о речевом тракте и форме импульса возбуждения — в области малых времен кепстра. Зависящий от времени комплексный спектр фактически содержит ту же информацию, что и текущий спектр сигнала, который, в свою очередь (см. гл. 6), является точным описанием речевого сигнала. Кепстральное представление, однако, не использует информации о фазе сигнала, содержащейся в преобразовании Фурье, и поэтому кратковременный кепстр не позволяет единственным образом описать речевое колебание. Тем не менее на основе кепстра можно оценить формантные частоты, период основного тона и классифицировать сигнал как вокализованный или невокализованный. Кепстр используется также для непосредственного описания речи в системах, называемых гомоморфными вокодерами [17].

В гомоморфном вокодере кепстр вычисляется 1 раз через каждые 10—20 мс. Период основного тона и признак тон/шум оцениваются по кепстру, а компоненты кепстра в области малых времен (примерно первые 30 отсчетов) квантуются и кодируются для передачи или хранения. По квантованным отсчетам кепстра в области малых времен в синтезаторе восстанавливается импульсная реакция $h_v(n)$ или $h_u(n)$ и вычисляется свертка с функцией возбуждения, восстановленной в синтезаторе по информации об основном тоне, признаке тон/шум и соответствующих амплитудах. Этот алгоритм представлен на рис. 7.22. На рис. 7.22*а* показан анализатор. Кепстр вычисляется в соответствии с 7.1.3. Затем с помощью кепстрального окна выделяется область малых времен. В [17] при моделировании использовались первые 26 отсчетов кепстра. Полный кепстр использовался также для выделения информации и об основном тоне и признаке тон/шум в соответствии с результатами § 7.3. Информация о сигнале возбуждения совместно с квантованными значениями кепстра использовалась для цифрового представления сигнала и передавалась по каналу 50—100 раз в секунду. Для синтеза входного сигнала по кепстральному описанию вычислялась импульсная реакция. Для того чтобы понять, как это делалось, вспомним, что кепстр — это четная функция времени и поэтому для построения кепстра достаточно знать лишь его часть, локализованную в области положительного времени.

Преобразование Фурье части кепстра в области малых времен приводит к логарифму передаточной функции, описывающей совместное влияние речевого тракта, формы импульса возбуждения и излучения. Однако фаза в данном случае равна нулю. В схеме рис. 7.22*б* преобразование Фурье изменяется для получения действительного четного преобразования, обратное преобразование которого представляет собой «импульсную характеристику», являющуюся четной функцией. Импульсная характеристика, полученная таким образом по кепстру (см. рис. 7.8*е*), показана на рис. 7.23*а*. Эту импульсную характеристику можно свернуть с последовательностью импульсов, отстоящих друг от друга на период основного тона для вокализованной речи, и с равноотстоящей последовательностью импульсов случайной полярности для невокализованных сегментов. (В [17] расстояние между импульсами для невокализованного сигнала превосходило единицу для уменьшения объема вычислений.)

По логарифмическому спектру можно получить и минимально-фазовую импульсную характеристику, для чего следует использовать кепстральное окно вида



$$l(n)=\begin{cases}1, & n=0,\\ 2, & 0<n\leqslant n_0,\\ 0, & \textit{в противном случае.}\end{cases} \qquad (7.48)$$

Результат преобразования, приводящий к минимально-фазовой характеристике, показан на рис. 7.23*б* [5]. Импульсная характе-

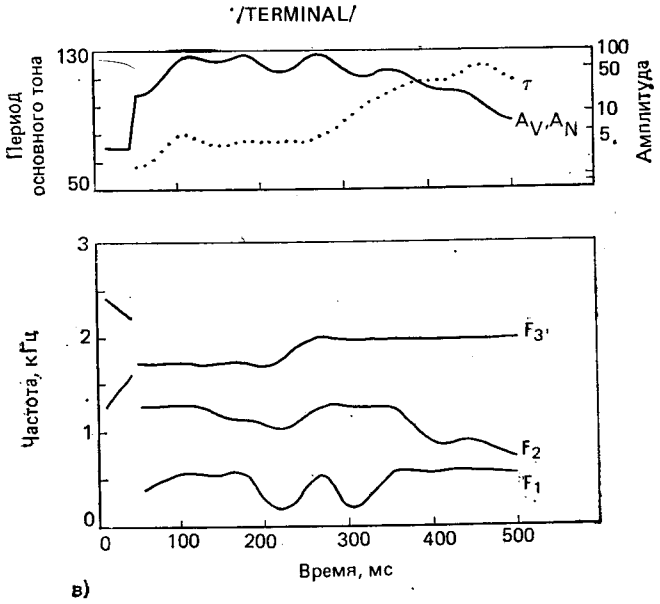


Рис. 7.21. Иллюстрация процесса квантования управляющих сигналов формантного вокодера:
а) исходные данные; *б*) сглаженные данные; *в*) квантованные и сглаженные данные

ристика, изображенная на рис. 7.23*б*, имеет такой же логарифм преобразования Фурье, как и исходная (рис. 7.23*а*). Оппенгейм [17] рассмотрел также случай максимально-фазового восстанов-

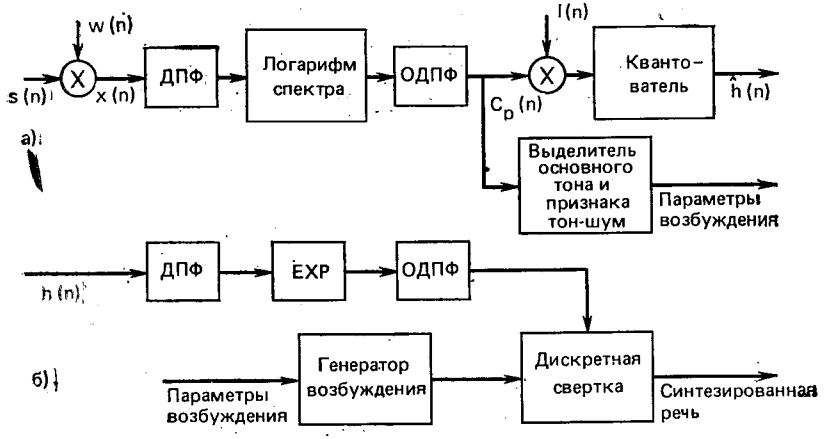


Рис. 7.22. Структурная схема гомоморфного вокодера:
а) анализатор; *б*) синтезатор

ления импульсной реакции, т. е.

$$l(n)=\begin{cases}1, & n=0,\\ 2, & -n_0\leqslant n<0,\\ 0, & \textit{в противном случае.}\end{cases} \tag{7.49}$$

Этот случай для нашего примера представлен на рис. 7.23*в*. Тесты на восприятие показали, что минимально-фазовое описание является наиболее предпочтительным. Это вполне естественно вследствие того, что минимально фазовый сигнал наиболее соответствует речевому сигналу.

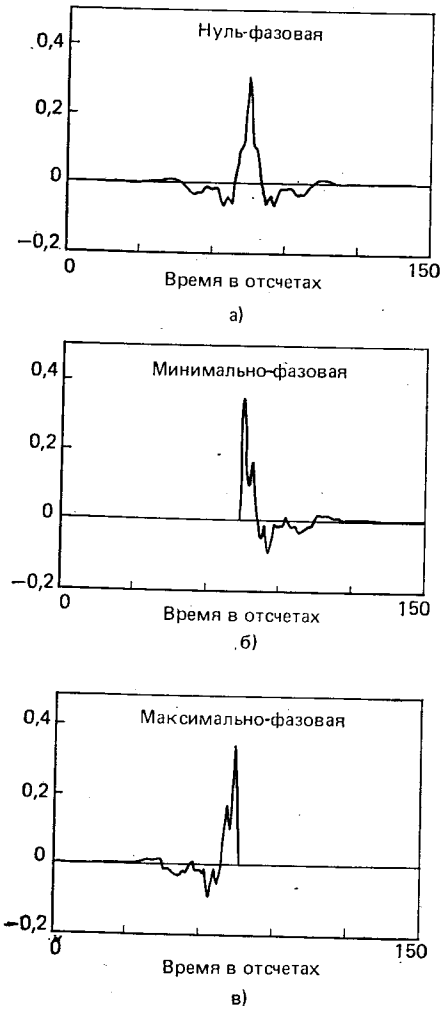


Рис. 7.23. Импульсная характеристика, вычисленная по кепстру:
а) нуль-фазовая; *б*) минимально-фазовая; *в*) максимально-фазовая

Гомоморфный вокодер с 26 значениями кепстра, квантованными с частотой 50 Гц, обеспечивает «очень высокое качество и натуральность речевого сигнала» [17]. Последующие исследования показали, что при преобразовании кепстральной информации перед квантованием скорость передачи может быть значительно понижена [18]. Другие исследования показали, что для повышения эффективности кепстральных методов целесообразно применять адаптацию протяженности временно́го окна, используемого при вычислении спектра сигнала [19].

Гомоморфный вокодер, как и любые вокодерные системы, в которых пытаются разделить параметры речи на сигнал возбуждения и параметры речевого тракта, позволяет достигнуть малой скорости передачи и дополнительной гибкости при обработке речи ценой усложнения в описании и потерь в качестве. Данная система обладает тем преимуществом, что кепстр, требующий для своего вычисления наибольших затрат, позволяет оценить как параметры речевого тракта, так и параметры возбуждения. Данный метод наиболее привлекателен, если имеется возможность использования БИС для вычисления ДПФ.

7.6. Заключение

В данной главе рассмотрены основные методы гомоморфной обработки сигналов применительно к речи. Основная идея гомоморфной обработки заключается в разделении или обратной свертке сегмента речевого сигнала на компоненты, представляющие импульсную характеристику и источник возбуждения. Это достигается путем линейной фильтрации обратного преобразования Фурье логарифма спектра сигнала, т. е. кепстра сигнала. Рассмотрены вычислительные аспекты применения гомоморфной обработки речи. В заключительной части главы изложены некоторые основные методы оценивания параметров сигнала на основе гомоморфной модели.

Задачи

7.1. Комплексный кепстр последовательности является обратным преобразованием Фурье комплексного логарифма спектра

$$\hat{X}(e^{i\omega})=\log|X(e^{i\omega})|+i\arg[X(e^{i\omega})].$$

Показать, что кепстр $c(n)$, определенный как обратное преобразование Фурье логарифма модуля, является четной частью $\hat{x}(n)$, т. е. показать, что $c(n)=\hat{x}(n)+\hat{x}(-n)/2$.

7.2. Рассмотрим полюсную модель речевого тракта в виде

$$V(z)=\frac{1}{\prod_{k=1}^{q}(1-c_k z^{-1})(1-c_k^{*} z^{-1})},$$

где $c_k=r_k e^{i\theta_k}$.

Показать, что соответствующий кепстр имеет вид

$$\hat{v}(n)=2\sum_{k=1}^{q}\frac{(r_k)^n}{n}\cos(\theta_k n).$$

7.3. Рассмотрим полюсную модель, описывающую речевой тракт, форму импульса возбуждения и излучение в виде

$$H(z)=\frac{G}{1-\sum_{k=1}^{p}\alpha_k z^{-k}}.$$

Предположим, что все полюса лежат внутри единичной окружности. Используя (7.22), получим рекурсивное соотношение между комплексным кепстром, $\hat{h}(n)$ и коэффициентами $\{\alpha_k\}$ (как комплексный кепстр $1/H(z)$ связан с $\hat{h}(n)$?).

7.4. Рассмотрим минимально-фазовую последовательность $x(n)$ *конечной длины* с кепстром $\hat{x}(n)$ и последовательность $y(n)=\alpha^n x(n)$ с комплексным кепстром $\hat{y}(n)$.

а) Если $0<\alpha<1$, то как $\hat{y}(n)$ связан с $\hat{x}(n)$?

б) Как следует выбрать α, чтобы $y(n)$ уже не был минимально-фазовым?

в) Как следует выбрать α, чтобы $y(n)$ был максимально-фазовым?

7.5. Показать, что если $x(n)$ — минимально-фазовый, то $x(-n)$ — максимально-фазовый.

7.6. Рассмотрим последовательность $x(n)$ с комплексным кепстром $\hat{x}(n)$. Преобразование $\hat{x}(n)$ имеет вид

$$\hat{X}(z)=\log[X(z)]=\sum_{m=-\infty}^{\infty}\hat{x}(m)z^{-m},$$

где $X(z)$ — z-преобразование $x(n)$.

Преобразование $\hat{X}(z)$ дискретизировано в N равностоящих точках на единичной окружности:

$$\hat{X}_p(k) = \hat{X}\left(e^{j\frac{2\pi}{N}k}\right), \quad 0 \leqslant k \leqslant N-1.$$

Используя ДПФ, вычисляем

$$\hat{x}_p(n) = \frac{1}{N}\sum_{k=0}^{N-1} \hat{X}_p(k)\, e^{j\frac{2\pi}{N}kn}, \quad 0 \leqslant n \leqslant N-1,$$

что может служить аппроксимацией комплексного кепстра.

а) Выразить $\hat{X}_p(k)$ через действительный кепстр $\hat{x}(m)$.

б) Подставить выражение п.а) в обратное преобразование Фурье для $\hat{x}_p(n)$ и показать, что $\hat{x}_p(n) = \sum_{r=-\infty}^{\infty} \hat{x}(n+rN)$.

7.7. Рассмотрим последовательность

$$x(n) = \delta(n) + \alpha\delta(n - N_p).$$

а) Определить комплексный кепстр. Изобразить ваш результат.

б) Изобразить кепстр $c(n)$ для $x(n)$.

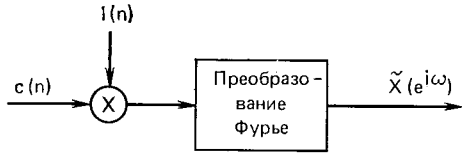


Рис. 3.7.1

в) Предположим, что по (7.30) вычислено приближение $x_p(n)$. Изобразить $x_p(n)$ для $0 \leqslant n \leqslant N-1$ в случае $N_p = N/6$. Что, если N не делится нацело на N_p?

г) Повторить п.в) для кепстральной аппроксимации $c_p(n)$ при $0 \leqslant n \leqslant N-1$ с использованием (7.33).

д) Если наибольшее значение кепстральной аппроксимации $c_p(n)$ используется для определения N_p, то как велико должно быть N для того, чтобы избежать ошибок?

7.8. Для сглаживания логарифма модуля спектра сигнала его кепстр часто взвешивают и преобразование Фурье имеет вид рис. 3.7.1.

а) Написать выражение, связывающее $X(e^{j\omega})$ с $\log|X(e^{j\omega})|$ и $L(e^{j\omega})$, где $L(e^{j\omega})$ — преобразование Фурье $l(n)$.

б) Какое кепстральное окно следует использовать для сглаживания функции $\log|X(e^{j\omega})|$?

в) Сравнить применение прямоугольного окна и окна Хемминга в качестве кепстральных окон.

г) Какова должна быть протяженность кепстрального окна и почему?

7.9. Рассмотрим сегмент вокализованного сигнала, который можно представить в виде $s(m) = p(m) * h_v(m)$, где $p(m) = \sum_{r=-\infty}^{\infty} \delta(m - rN_p)$. При вычислении комплексного кепстра (или кепстра) первый шаг заключается в умножении $s(m)$ на окно $w(m)$ для выделения сегмента $x_n(m) = s(m)w(n-m)$ входных данных для гомоморфной обработки.

а) Определить условия, при которых можно аппроксимировать $x_n(m)$ в виде $x(m) = p_n(m) * h_v(m)$, где $p_n(m) = p(m)w(n-m)$.

б) Для специального случая $n=0$ определить преобразование $p_0(m)$ через z-преобразование $w(m)$.

в) Выразить комплексный кепстр $\hat{p}_0(m)$ через $\hat{w}(m)$.

7.10. В задаче 7.9 показано, что периодичность взвешенного сегмента вокализованной речи может быть приближенно представлена выражением $p_n(m) = p(m)w(n-m)$, где $p(m) = \sum_{r=-\infty}^{\infty} \delta(m - rN_p)$. В этой задаче исследуется влияние

положения окна на комплексный кепстр $\hat{p}_n(m)$. Предположим, что имеется окно Хемминга вида

$$w(m) = \begin{cases} 0{,}54 - 0{,}46\cos(2\pi m/(2N_p)), & 0 \leqslant n \leqslant 2N_p; \\ 0, & \textit{в противном случае.} \end{cases}$$

а) Изобразить $p_n(m)$ как функцию m для $n = 3N_p/4;\ 9N_p/8;\ 5N_p/4;\ 3N_p/2$.

б) Для каждого из перечисленных выше случаев составить выражение для $p_n(m)$ и показать, что соответствующее z-преобразование имеет вид $P_n(z) = \alpha_1 z^{N_p} + \alpha_2 + \alpha_3 z^{-N_p}$.

в) Для каждого из перечисленных выше случаев определить и изобразить комплексный кепстр (указание: использовать разложение в ряд для $\log[P_n(z)]$). Опустить члены вида $\log[z^{\pm N_p}]$.

г) Для какого положения окна справедливы следующие утверждения:
последовательность $p_n(m)$ минимально-фазовая;
последовательность $p_n(m)$ максимально-фазовая;
первый кепстральный пик максимален;
первый кепстральный пик минимален.

д) Как изменятся ваши ответы на перечисленные выше вопросы, если окно удлинится? Укоротится?

7.11. Преобразование сигнала $x(n)$ определяется как

$$X(z) = \sum_{n=0}^{N-1} x(n)\, z^{-n}.$$

Вычислим $X(z)$ в последовательности точек $z_k = AW^{-k}k = 0, 1, \ldots, M-1$, где A и W — произвольные комплексные целые числа. Если сделать простую подстановку $nk = [n^2 + k^2 - (k-n)^2]/2$, то $X(z_k)$ можно записать в виде

$$X(z_k) = P(k) \sum_{n=0}^{N-1} y(n)\, g(k-n),$$

т. е. $X(z_k)$ — свертка $y(n)$ и $g(n)$.

а) Определить $P(k)$, $y(n)$ и $g(n)$ через $x(n)$, A и W.

б) Изобразить точки z_k на z-плоскости.

в) Можете ли Вы предложить способ применения БПФ для вычисления приведенного выше выражения?

8

Кодирование речевых сигналов на основе линейного предсказания

8.0. Введение

Линейное предсказание является одним из наиболее эффективных методов анализа речевого сигнала. Этот метод становится доминирующим при оценке основных параметров речевого сигнала, таких, как, например, период основного тона, форманты, спектр, функция площади речевого тракта, а также при сокращенном представлении речи с целью ее низкоскоростной передачи и экономного хранения. Важность метода обусловлена высокой точностью получаемых оценок и относительной простотой вычислений. В данной главе излагаются основные положения метода линейного предсказания и приводятся рекомендации по его практическому использованию.