

ВЫБОР ИНТЕРВАЛОВ АНАЛИЗА СИГНАЛА ПРИ РАСПОЗНАВАНИИ РЕЧИ

Рассматривается проблема выбора интервалов анализа сигнала при распознавании речи. Исследуются преимущества и недостатки прямоугольного временного окна при кратковременном спектральном анализе. Предлагаются методы выбора интервалов анализа для речевых сигналов, которые позволяют устранить недостатки прямоугольного окна.

При распознавании речи обрабатываемый сигнал сначала разбивается на небольшие отрезки (интервалы анализа), которые затем классифицируются. Полученная последовательность отрезков речевого сигнала может затем подвергаться дальнейшей обработке, например, с целью получения последовательности слов. Интервалы анализа обычно имеют одинаковую длину, а для их классификации переходят к спектральному представлению сигнала. Перед нахождением спектра сигнал подвергается преобразованию с помощью непрямоугольного в общем случае временного окна (как правило, Хэмминга или Ханна). Такое преобразование искажает сигнал и может затруднить его классификацию. Использование прямоугольного окна, не искажающего сигнал, также имеет свои недостатки. Однако возможно выбирать интервалы анализа таким образом, что недостатки прямоугольного окна устраняются и при этом сохраняются его полезные свойства.

ТРЕБОВАНИЯ К ДЛИНЕ ИНТЕРВАЛОВ АНАЛИЗА

Речевой сигнал можно представить как функцию времени $p(t)$, которая выражает изменение во времени какой-либо физической величины, например звукового давления или напряжения.

Согласно общепринятой акустической теории речеобразования [1], преобразование Лапласа от функции $p(t)$ звукового давления можно представить в виде произведения

$$P(s) = S(s) \cdot T(s),$$

где $S(s)$ есть преобразование Лапласа для функции источника звука, а $T(s)$ – передаточная функция речевого тракта. Передаточная функция $T(s)$ является мероморфной функцией и полностью определяется бесконечным числом комплексно-сопряженных полюсов и нулей.

Функция $s(t)$ источника звука представляет собой изменение во времени звукового давления, вызванное источником звука. Источник звука может быть либо голосовым, либо шумовым. Голосовой источник обусловлен колебанием голосовых связок, функция $s(t)$ в этом случае является квазипериодической. Среднее значение T_0 периода колебаний голосовых связок называется периодом основного тона речи, а обратная ей величина $F_0 = 1/T_0$ представляет собой частоту основного тона.

Отклонения периода основного тона от среднего значения можно классифицировать следующим образом:

1) случайные отклонения (джиттер-эффект [2]). Эти отклонения невелики и возможны как в сторону увеличения, так и в сторону уменьшения периода основного тона. Случайные отклонения не несут какой-либо смысловой нагрузки;

2) вибрато. Вибрато есть частотная модуляция колебаний голосовых связок и проявляется в периодическом или квазипериодическом изменении периода

основного тона. Вибрато характерно для пения, и целью его использования является придание голосу красоты. Как правило, вибрато не несет какой-либо смысловой нагрузки;

3) плавное увеличение или уменьшение периода основного тона. Этот тип отклонений является частью интонационного рисунка речи. Обычно увеличение частоты основного тона вызывается увеличением подсвязочного давления, вследствие чего увеличивается также и громкость речи, и наоборот, хотя увеличение частоты основного тона может происходить и без увеличения подсвязочного давления. Отслеживание изменения частоты основного тона совместно с интенсивностью звука является важным при определении границ слогов, фонетических слов и фонетических синтагм и необходимым для определения словесных и синтагматических ударений и интонации, которая указывает на общую коммуникативную цель говорящего.

Отслеживание изменения частоты основного тона требует анализа речевого сигнала на достаточно небольших временных интервалах.

Небольшие интервалы анализа (порядка одного периода основного тона) требуются и при отслеживании сдвига формант, имеющего место во время переходных процессов в речевом тракте. Например, слова [mama] и [nana] различаются человеком только за счет сдвига формант в начале произнесения первого и второго [a], в то время как сами по себе [m] и [n] практически не различаются [1].

ДИСКРЕТИЗАЦИЯ РЕЧЕВЫХ СИГНАЛОВ

Если в процессе произнесения данного речевого сигнала ему предшествовали и за ним следовали достаточно большие паузы, функцию $s(t)$ источника звука можно считать финитной. В этом случае изображение $S(s)$ будет целой функцией, а $P(s)$ – мероморфной, из чего следует нефинитность функции $p(t)$.

В реальности наблюдение функции $p(t)$ ограничивают каким-либо интервалом T_1 , превращая ее в финитную. Преобразование Лапласа $P_\phi(s)$ такой финитной функции $p_\phi(t)$ будет целой функцией, т.е. информация о точном нахождении полюсов функции $P(s)$, а следовательно, и передаточной функции $T(s)$, будет потеряна.

Пусть отсчет времени для функции $p_\phi(t)$ ведется от момента времени $t=0$. Тогда функцию $p_\phi(t)$ можно разложить в ряд Фурье на интервале $[0, T_1]$

$$p_\phi(t) = \sum_{n=-\infty}^{\infty} c_n^\phi e^{i \frac{2\pi}{T_1} nt}.$$

Здесь предполагается, что это разложение существует.

Перед обработкой на компьютере речевой сигнал $p_\phi(t)$ дискретизируется с шагом Δt . Соответствующее

сигналу $p_\phi(t)$ дискретное колебание $p_o(t)$ можно представить выражением [3]

$$p_o(t) = \Delta t \sum_{k=0}^{N-1} p_\phi(k\Delta t) \delta(t - k\Delta t),$$

где $N=T_1/\Delta t$ – количество отсчетов сигнала $p_\phi(t)$, $\delta(t)$ – дельта-функция Дирака.

Коэффициенты c_n^ϕ ряда Фурье дискретного сигнала $p_o(t)$ выражаются через коэффициенты c_n^ϕ ряда Фурье сигнала $p_\phi(t)$ следующим образом:

$$c_n^\phi = \sum_{k=-\infty}^{\infty} c_{n-Nk}^\phi, \quad (1)$$

а само значение коэффициентов c_n^ϕ дается дискретным преобразованием Фурье (ДПФ)

$$c_n^\phi = \frac{1}{N} \sum_{k=0}^{N-1} p_\phi(k\Delta t) e^{-i\frac{2\pi}{N}nk}. \quad (2)$$

Интервал разложения для $p_o(t)$ равен $[0, T_1]$.

Предполагается, что в дискретном колебании $p_o(t)$ на выходе цифроаналогового преобразователя будут отфильтрованы гармоники с круговыми частотами выше $\pi/\Delta t$, и на выходе синтезирующего фильтра будет восстановленный сигнал $p_o(t)$ [4]. В случае идеального синтезирующего фильтра спектр сигнала $p_o(t)$ будет состоять из коэффициентов c_n^ϕ , $n = 0, \pm 1, \dots, \pm[N/2]$, где квадратные скобки означают взятие целой части числа.

Если для коэффициентов c_n^ϕ ряда Фурье сигнала $p_\phi(t)$ выполняется условие

$$c_n^\phi = 0 \text{ для } |n| \geq N/2, \quad (3)$$

то выражение (1) преобразуется к виду

$$c_n^\phi = c_n^\phi, \quad n = 0, \pm 1, \dots, \pm[N/2], \quad (4)$$

т.е. сигналы $p_\phi(t)$ и $p_o(t)$ будут совпадать.

Если условие (3) не выполняется, то происходит эффект наложения спектров [5]. Чтобы избежать этого эффекта, перед дискретизацией сигнала $p_\phi(t)$ производится фильтрация гармоник с круговыми частотами, большими $\pi/\Delta t$. Однако, поскольку аналоговые фильтры нижних частот не являются идеальными, то эффект наложения спектров может иметь место и в этом случае.

В случае четного N из условия (3) следует $c_{-N/2}^\phi = c_{N/2}^\phi = 0$. Однако на практике при выполнении ДПФ часто оказывается, что $c_{-N/2}^\phi \neq 0$ и $c_{N/2}^\phi \neq 0$. Поскольку в этом случае условие (3) уже не выполняется, введем более слабое условие

$$c_n^\phi = 0 \text{ для } |n| > N/2. \quad (3')$$

Из (1) и (3') в случае четного N следует

$$\begin{cases} c_n^\phi = c_n^\phi, \quad n = 0, \pm 1, \dots, \pm\left(\frac{N}{2}-1\right), \\ c_{-N/2}^\phi = c_{N/2}^\phi = c_{-N/2}^\phi + c_{N/2}^\phi. \end{cases}$$

Пусть $c_n^\phi = \text{Re}\{c_n^\phi\} + i \text{Im}\{c_n^\phi\}$, тогда, поскольку $c_{-n}^\phi = \text{Re}\{c_n^\phi\} - i \text{Im}\{c_n^\phi\}$,

$$c_{-N/2}^\phi = c_{N/2}^\phi = 2 \text{Re}\{c_{N/2}^\phi\}. \quad (5)$$

Из равенства (5) можно сделать вывод, что синусная составляющая гармоники с номером $N/2$ сигнала $p_\phi(t)$ при дискретизации теряется, что естественно, так как функцию $\sin(\pi t/\Delta t)$ нельзя корректно дискретизовать с шагом Δt . С другой стороны, косинусная составляющая корректно дискретизируется, но из-за эффекта наложения спектров происходит ее удвоение.

Из (5) следует, что в случае четного N идеальный синтезирующий фильтр на выходе цифроаналогового преобразователя должен уменьшать в 2 раза амплитуды гармоник с номерами $\pm N/2$.

В связи с вышеизложенным удобно ввести коэффициенты $\bar{c}_n^\phi$, связанные с коэффициентами c_n^ϕ ряда Фурье дискретного сигнала $p_o(t)$ следующим образом:

$$\bar{c}_n^\phi = \begin{cases} c_n^\phi & \text{для } n = 0, \pm 1, \dots, \pm\left[\frac{N-1}{2}\right], \\ \frac{c_n^\phi}{2} & \text{для } n = \pm\frac{N}{2} \text{ при четном } N. \end{cases} \quad (6)$$

Таким образом, из выполнения условия (3') будет следовать

$$\bar{c}_n^\phi = \begin{cases} c_n^\phi, \quad n = 0, \pm 1, \dots, \pm\left[\frac{N-1}{2}\right], \\ \text{Re}\{c_n^\phi\} & \text{для } n = \pm\frac{N}{2} \text{ при четном } N. \end{cases} \quad (7)$$

Поскольку синусная составляющая гармоники с номером $N/2$ сигнала $p_\phi(t)$ потеряна, то можно считать, что она была равна нулю. Тогда из (3') и (7) получаем

$$\bar{c}_n^\phi = c_n^\phi, \quad n = 0, \pm 1, \dots, \pm[N/2], \quad (4')$$

и с учетом вышесказанного сигналы $p_\phi(t)$ и $p_o(t)$ будут совпадать и в этом случае.

Если абсолютные значения коэффициентов c_n^ϕ для $|n| > N/2$ достаточно малы, так что выполняется приближенное равенство

$$\bar{c}_n^\phi \approx c_n^\phi, \quad n = 0, \pm 1, \dots, \pm[N/2],$$

то удобно исходным считать сигнал

$$p_\phi(t) = \sum_{n=-[N/2]}^{[N/2]} \bar{c}_n^\phi e^{i\frac{2\pi}{T_1}nt}. \quad (8)$$

Для сигнала (8) будут автоматически выполняться условия (3') и (4'). В дальнейших рассуждениях будем считать, что для сигнала $p_\phi(t)$ условия (3') и (4') выполняются.

ПРОБЛЕМА ВЫБОРА ИНТЕРВАЛОВ АНАЛИЗА ДЛЯ РЕЧЕВЫХ СИГНАЛОВ

При распознавании речи большой интерес представляют спектральные свойства речевого сигнала $p_\phi(t)$. Для нахождения спектра дискретного сигнала $p_o(t)$ используется ДПФ. В этом случае мы в соответствии с равенством (4') одновременно получаем и спектр сигнала $p_\phi(t)$.

Но поскольку сигнал $p_\phi(t)$ в общем случае имеет достаточно большую длительность и представляет собой реализацию некоторого количества фонов, то

интересны, прежде всего, спектральные свойства не всего сигнала $p_\phi(t)$, а его локальных участков.

В связи с этим возникает проблема сегментации речи: перед тем как определять, к какому классу относится тот или иной участок речевого сигнала, речевой сигнал необходимо сегментировать на эти участки.

Сегментация речевого сигнала неоднозначна и зависит от используемых критериев. Общепринятым подходом является разбиение сигнала на пересекающиеся или непересекающиеся интервалы времени фиксированной длины в диапазоне от 2 до 50 мс, исходя из предположения, что на интервале данной длины изменение характеристик речевого тракта достаточно незначительно.

Перед выполнением спектрального анализа из сигнала $p_\phi(t)$ с помощью временного окна $w(m)$ выделяется кратковременный сигнал

$$p_{\phi m}(t) = \Delta t \sum_{k=0}^{N-1} w(k-m) p_\phi(k\Delta t) \delta(t-k\Delta t), \quad (9)$$

равный нулю вне интервала $[t_0, t_0+T_2]$, $t_0=m\Delta t$, $T_2=M\Delta t$. ДПФ этого сигнала на интервале разложения $[t_0, t_0+T_2]$ есть

$$c_{mn}^{\phi} = e^{-i\frac{2\pi}{M}nm} \frac{1}{M} \sum_{k=0}^{M-1} w(k) p_\phi((m+k)\Delta t) e^{-i\frac{2\pi}{M}nk}, \quad (10)$$

где c_{mn}^{ϕ} – n -й коэффициент ряда Фурье сигнала $p_{\phi m}(t)$.

Множитель $e^{-i\frac{2\pi}{M}nm}$ в (10) представляет собой линейное изменение фазы, связанное с задержкой сигнала на m отсчетов. Обычно этот множитель не учитывается, поскольку положение сигнала $p_{\phi m}(t)$ на временной оси, как правило, неважно.

Наиболее распространенными типами окон являются прямоугольное и обобщенное окна Хэмминга [6, 7].

Окна, отличные от прямоугольного, дополнительно искажают сигнал $p_{\phi m}(t)$, поэтому искажается и ряд Фурье (10) в сравнении с рядом для прямоугольного окна.

Рассмотрим для примера искажение, вносимое M -точечным обобщенным окном Хэмминга

$$w_H(m) = \begin{cases} \alpha - (1-\alpha) \cos\left(\frac{2\pi m}{M}\right), & 0 \leq m \leq M-1, \\ 0 & \text{при других } m. \end{cases}$$

Пусть на интервал длиной $T_2=M\Delta t$, выделяемый окном $w_H(m)$, приходится ровно один период колебания голосовых связок, взятый из некоторого речевого сигнала. Для выделенного таким образом исходного сигнала $p_\phi(t)$ длительностью T_2 ряд Фурье (2) совпадает с рядом Фурье (10) в случае использования M -точечного прямоугольного окна.

В случае же использования обобщенного окна Хэмминга ряд Фурье (10) будет совпадать с рядом Фурье сигнала

$$p_{\phi u}(t) = p_\phi(t) \left[\alpha - (1-\alpha) \cos\left(\frac{2\pi}{T_2}t\right) \right]. \quad (11)$$

Чтобы отследить отличия спектров у сигналов $p_\phi(t)$ и $p_{\phi u}(t)$, удобно представить сигнал $p_\phi(t)$ рядом

Фурье в тригонометрической форме:

$$p_\phi(t) = \frac{a_0}{2} + \sum_{n=1}^{[N/2]} A_n \cos\left(\frac{2\pi n}{T_2}t + \theta_n\right). \quad (12)$$

Из (11) и (12) видно, что преобразование сигнала $p_\phi(t)$ с помощью обобщенного окна Хэмминга эквивалентно амплитудной модуляции с круговой частотой $2\pi/T_2$ каждой гармоники сигнала $p_\phi(t)$, т.е. у каждой гармоники с круговой частотой $2\pi n/T_2$ появляются копии с круговыми частотами $2\pi(n-1)/T_2$ и $2\pi(n+1)/T_2$ и амплитудой $(1-\alpha)A_n/2$. Эти копии накладываются на соседние гармоники, искажая их.

Это явление можно проинтерпретировать и другим образом, а именно: при преобразовании сигнала с помощью обобщенного окна Хэмминга происходит размытие спектра по частотной шкале. Из приведенных рассуждений видно, что при использовании временных окон, отличных от прямоугольного, будет происходить то или иное искажение спектра, которое можно представлять себе в виде его некоторого размытия.

Размытие спектра является недостатком, которого желательно избегать, так как при этом могут потеряться или притупиться формантные максимумы, что делает задачу классификации данного участка речевого сигнала более сложной.

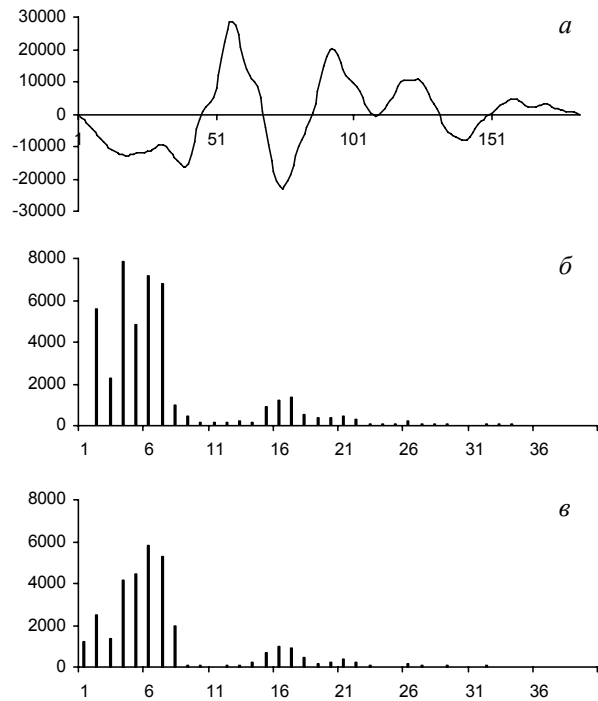


Рис. 1. Сравнение ДПФ с использованием прямоугольного окна и окна Ханна: а) дискретный сигнал, для которого выполняется ДПФ, по оси абсцисс отложены номера отсчетов; б) амплитудный спектр для прямоугольного окна; в) амплитудный спектр для окна Ханна. Спектральные коэффициенты пронумерованы начиная с единицы

На рис. 1, б и в представлены для сравнения спектры, полученные с использованием прямоугольного окна и окна Ханна (обобщенного окна Хэмминга с параметром $\alpha = 0,5$) на примере дискретного сигнала, представляющего собой один период речевого сигнала, взятый из реализации фонемы /а/, произнесенной

женским голосом (показаны только первые 40 спектральных коэффициентов, модули остальных коэффициентов пренебрежимо малы). На рис. 1,б отчетливо видны формантные максимумы, соответствующие первым двум формантам: F_1 находится около гармоники $n=3$ (на рисунке она имеет номер 4) с частотой 723 Гц, F_2 находится около гармоники $n=5$ с частотой 1205 Гц. На рис. 1,в видно, что из-за размытия спектра формантный максимум, соответствующий первой форманте, исчез.

В [6] отмечается, что частотное разрешение у прямоугольного окна больше, чем у окна Хэмминга, однако прямоугольное окно редко используется при спектральном анализе речи из-за того, что преобразование Фурье

$$X_m(e^{i\omega}) = \sum_{k=-\infty}^{\infty} w(k-m)p_{\phi}(k\Delta t)e^{-i\omega k}$$

дискретного сигнала $p_{\phi m}(t)$ в случае прямоугольного окна имеет более изрезанный характер между гармониками, чем, например, преобразование Фурье с окном Хэмминга [6].

Но если заданы значения $X_m(e^{i\omega})$ в точках $\omega=2\pi n/M$, $n=0, \dots, [M/2]$, т.е. гармоники, то значения $X_m(e^{i\omega})$ для других значений ω не несут информации, так как могут быть однозначно вычислены. Поэтому данное свойство прямоугольного окна не должно препятствовать его использованию при спектральном анализе речи.

Рассмотрим еще один недостаток прямоугольного окна, который не упоминается в литературе.

В случае прямоугольного окна выражение (9) приводится к виду

$$p_{\phi m}(t) = \Delta t \sum_{k=m}^{m+M-1} p_{\phi}(k\Delta t)\delta(t-k\Delta t),$$

а выражение (10) – к виду

$$c_{mn}^{\phi} = e^{-i\frac{2\pi}{M}nm} \frac{1}{M} \sum_{k=0}^{M-1} p_{\phi}((m+k)\Delta t) e^{-i\frac{2\pi}{M}nk}.$$

Для нас интересен случай, когда $M < N$, где N – количество отсчетов сигнала $p_{\phi}(t)$. Найдем соответствие между спектрами дискретного сигнала $p_{\phi m}(t)$ и сигнала

$$p_{\phi m}(t) = \begin{cases} p_{\phi}(t), & t_0 \leq t < t_0 + T_2, \\ 0 & \text{в противном случае,} \end{cases}$$

которому соответствует дискретное колебание $p_{\phi m}(t)$.

Для этого разложим сигнал $p_{\phi m}(t)$ в ряд Фурье на интервале $[t_0, t_0+T_2]$:

$$p_{\phi m}(t) = \sum_{n=-\infty}^{\infty} c_{mn}^{\phi} e^{i\frac{2\pi n}{T_2}t}.$$

Искомое соответствие между спектрами $p_{\phi m}(t)$ и $p_{\phi}(t)$ определяется выражением (1), которое можно переписать в следующем виде:

$$c_{mn}^{\phi} = \sum_{k=-\infty}^{\infty} c_{m,n-Nk}^{\phi}. \quad (13)$$

В выражении (13) для нас важны только коэффициенты с номерами $n=0, \pm 1, \dots, \pm[M/2]$, так как остальные коэффициенты получаются периодическим

повторением с шагом M коэффициентов из данного диапазона.

В дальнейшем при анализе скорости убывания коэффициентов c_{mn}^{ϕ} на производную функции $p_{\phi m}(t)$ накладывается достаточно слабое ограничение абсолютной интегрируемости, которое следует хотя бы из условия (3').

Если периодически продолженная функция $p_{\phi m}(t)$ терпит разрывы первого рода в точках t_0+nT_2 , $n \in \mathbb{Z}$, т.е. $p_{\phi m}(t_0) \neq p_{\phi m}(t_0+T_2-0)$, то ряд Фурье сигнала $p_{\phi m}(t)$ будет иметь бесконечное число ненулевых коэффициентов c_{mn}^{ϕ} , которые при $n \rightarrow \infty$ будут бесконечно малыми первого порядка [8], т.е. будут сравнительно медленно убывать.

Анализируя выражение (13), приходим к выводу, что в случае разрыва функции $p_{\phi m}(t)$ на краях интервала $[t_0, t_0+T_2]$ спектр дискретного сигнала $p_{\phi m}(t)$ будет искажен или зашумлен медленно убывающими хвостами спектра непрерывного сигнала $p_{\phi m}(t)$.

Абсолютная величина разрыва

$$c = |p_{\phi m}(t_0 + T_2 - 0) - p_{\phi m}(t_0)|$$

также непосредственно влияет на величину коэффициентов c_{mn}^{ϕ} . Если представить c_{mn}^{ϕ} в виде

$$c_{mn}^{\phi} = \frac{1}{2}(a_n - ib_n),$$

то абсолютные значения коэффициентов b_n будут асимптотически эквивалентны $c/\pi n$ при $n \rightarrow \infty$ [8], из чего следует, что чем больший разрыв терпит функция $p_{\phi m}(t)$ на краях интервала $[t_0, t_0+T_2]$, тем более зашумленным окажется спектр сигнала $p_{\phi m}(t)$.

Описанный эффект зашумления спектра дискретного сигнала является недостатком, которого необходимо избегать. Из вышеизложенного виден и способ уменьшения зашумления спектра сигнала $p_{\phi m}(t)$: необходимо уменьшить величину разрыва функции $p_{\phi m}(t)$ на краях интервала анализа, доведя ее в идеале до нуля. Этого можно достичь путем выбора подходящего интервала анализа.

На скорость убывания хвостов спектра функции $p_{\phi m}(t)$ влияют также и разрывы ее производных, а именно: если у функции $p_{\phi m}(t)$ $k-1$ производных непрерывны, а k -я производная абсолютно интегрируема, то для модулей коэффициентов c_{mn}^{ϕ} ряда Фурье функции $p_{\phi m}(t)$ имеет место соотношение

$$\lim_{n \rightarrow \infty} n^k |c_{mn}^{\phi}| = 0.$$

Выбирая интервал анализа таким образом, чтобы выполнялись условия

$$p_{\phi m}^{(k)}(t_0 + 0) = p_{\phi m}^{(k)}(t_0 + T_2 - 0) \quad (14)$$

для $k < k_0$, и предполагая, что k_0-1 производных у функции $p_{\phi m}(t)$ непрерывны на интервале $[t_0, t_0+T_2]$, а k_0 -я ее производная абсолютно интегрируема на этом же интервале, что будет выполняться, например, при выполнении условия (3'), мы получаем, что последовательность $|c_{mn}^{\phi}|$ при $n \rightarrow \infty$ будет бесконечно малой более высокого порядка малости, чем $1/n^{k_0}$.

Таким образом, обобщая, можно сказать, что для минимизации эффекта зашумления спектра дискретного сигнала $p_{om}(t)$ интервал анализа надо выбирать таким образом, чтобы выполнялись условия (14) для как можно большего k . В пределе, если условия (14) выполняются для любого k , мы получаем скорость убывания хвостов спектра функции $p_{\phi m}(t)$ быстрее любой полиномиальной функции.

Чтобы определить абсолютную величину $c(M)$ разрыва функции $p_{\phi m}(t)$ на краях интервала $[t_0, t_0+T_2]$, имея в наличии только дискретную функцию $p_{\phi}(t)$, предположим непрерывность функции $p_{\phi}(t)$. Отсюда следует

$$p_{\phi m}(t_0 + T_2 - 0) = p_{\phi}(t_0 + T_2) = p_{\phi}(t_0 + T_2)$$

$$\text{и} \quad c(M) = |p_{\phi}((m+M)\Delta t) - p_{\phi}(m\Delta t)|. \quad (15)$$

МЕТОД ВЫБОРА ИНТЕРВАЛА АНАЛИЗА, ДАЮЩИЙ ТОЧНОЕ РЕШЕНИЕ

Будем считать, что начало $t_0 = m\Delta t$ интервала анализа задано. Требуется найти его длину $T_2 = M\Delta t$, исходя из критерия минимума зашумления спектра. На длину накладываются ограничения $T_{\min} \leq T_2 \leq T_{\max}$.

Ограничение $T_{\min} \leq T_2$ предполагает, что отрезок речевого сигнала с длительностью, меньшей чем $T_{\min}$, недостаточно хорошо отражает свойства кратковременно установившегося процесса в речевом тракте. В свою очередь, ограничение $T_2 \leq T_{\max}$ предполагает, что на отрезке речевого сигнала с длительностью, большей чем $T_{\max}$, процесс, происходящий в речевом тракте, успевает достаточно сильно измениться, а для распознавания речи наибольший интерес представляют именно кратковременно установившиеся процессы.

Минимизация зашумления спектра заключается в минимизации модулей $|c_{mn}^{\phi}|$ спектральных коэффициентов сигнала $p_{\phi m}(t)$ для $n > M/2$. В качестве минимизируемой величины естественно выбрать среднюю на интервале T_2 мощность

$$P_t(M) = 2 \sum_{n=[M/2]+1}^{\infty} |c_{mn}^{\phi}|^2$$

сигнала $p_{\phi m}(t)$ вне основной полосы с граничной круговой частотой $\pi/\Delta t$. Средняя мощность $P_t(M)$ зависит и от начала t_0 интервала анализа, который мы полагаем фиксированным.

При выполнении условия (3') можно найти сколь угодно точное решение этой задачи. Действительно, поскольку функция $p_{\phi}(t)$ представима конечной суммой известных гармоник, мы можем восстановить ее значения в любой точке, а следовательно, разложить ее в ряд Фурье на интервале $[t_0, t_0+T_2]$, после чего вычислить $P_t(M)$ на основании равенства

$$P_t(M) = \frac{1}{T_2} \int_{t_0}^{t_0+T_2} p_{\phi m}^2(t) dt - \sum_{n=-[M/2]}^{[M/2]} |c_{mn}^{\phi}|^2, \quad (16)$$

которое выполняется ввиду полноты тригонометрической системы.

Для вычисления $P_t(M)$ по формуле (16) нам понадобится вычислить значения $M+2$ интегралов. При использовании численного интегрирования потребу-

ется вычислять значения подынтегральных функций в различных точках. Вычисление значений подынтегральных функций в каждой точке потребует суммирования N синусоидальных и косинусоидальных составляющих ряда Фурье функции $p_{\phi}(t)$. Количество элементарных операций, которое нужно выполнить для вычисления $P_t(M)$, зависит от метода и точности численного интегрирования. Пусть для приближенного вычисления интеграла на интервале Δt необходимо вычислить значения $p_{\phi}(t)$ в C точках. Тогда на интервале T_2 необходимо вычислить значения $p_{\phi}(t)$ в CM точках, и общее число элементарных операций, нужное для вычисления $P_t(M)$, составит $CN(M^2+2M)$.

Для нахождения наилучшего интервала анализа понадобится вычислить значения $P_t(M)$ для $M=M_{\min}, \dots, M_{\max}$, где $M_{\min}=T_{\min}/\Delta t$, $M_{\max}=T_{\max}/\Delta t$.

Пусть $Q = M_{\max} - M_{\min}$. Тогда трудоемкость нахождения наилучшего интервала анализа при численном вычислении интегралов составит

$$T(Q) = \sum_{M=M_{\min}}^{M_{\max}} CN(M^2 + 2M) = CN(aQ^3 + bQ^2 + cQ + d), \quad (17)$$

где

$$a = 1/3,$$

$$b = M_{\min} + 3/2,$$

$$c = M_{\min}^2 + 3M_{\min} + 7/6,$$

$$d = M_{\min}^2 + 2M_{\min}.$$

Трудоемкость $T(Q)$, даваемая выражением (17), является кубической относительно Q , квадратичной относительно $M_{\min}$ и линейной относительно N . Также необходимо учитывать, что для разложения функции $p_{\phi}(t)$ в ряд Фурье на интервале $[0, T_1)$ потребуется выполнить ДПФ, которое имеет в общем случае квадратичную трудоемкость относительно N .

Эксперимент показывает, что скорость вычисления значений $P_t(M)$ с помощью численного интегрирования является крайне низкой, что делает данный способ нахождения наилучшего интервала анализа непригодным для практического применения. Кроме того, из-за вычислительных погрешностей, возникающих при численном интегрировании, полученный результат может отличаться от точного решения.

Интегралы в выражении (16) можно вычислить и аналитически. Для этого представим функцию $p_{\phi}(t)$ в виде ряда Фурье в тригонометрической форме:

$$p_{\phi}(t) = \frac{a_0}{2} + \sum_{n=1}^{[N/2]} A_n \sin\left(\frac{2\pi n}{T_1}t + \theta_n\right).$$

Тогда средняя на интервале $[t_0, t_0+T_2]$ мощность сигнала $p_{\phi m}(t)$ будет равна

$$\begin{aligned} \frac{1}{T_2} \int_{t_0}^{t_0+T_2} p_{\phi m}^2(t) dt &= \left(\frac{a_0}{2}\right)^2 + \frac{1}{2} \sum_{n=1}^{[N/2]} A_n^2 + \frac{N}{\pi M} \times \\ &\times \left\{ \frac{a_0}{2} \sum_{n=1}^{[N/2]} A_n \frac{1}{n} \left[\cos\left(2\pi \frac{m}{N}n + \theta_n\right) - \right. \right. \\ &\left. \left. - \cos\left(2\pi \frac{m+M}{N}n + \theta_n\right) \right] + \frac{1}{8} \sum_{n=1}^{[N/2]} A_n^2 \frac{1}{n} \times \right. \end{aligned}$$

$$\begin{aligned}
& \times \left[\sin 2 \left(2\pi \frac{m}{N} n + \theta_n \right) - \sin 2 \left(2\pi \frac{m+M}{N} n + \theta_n \right) \right] + \\
& + \frac{1}{4} \sum_{n=1}^{[N/2]} A_n \sum_{\substack{k=1 \\ k \neq n}}^{[N/2]} A_k \left[\frac{1}{n-k} \times \right. \\
& \times \left\{ \sin \left(2\pi \frac{m+M}{N} (n-k) + \theta_n - \theta_k \right) - \right. \\
& \left. \left. - \sin \left(2\pi \frac{m}{N} (n-k) + \theta_n - \theta_k \right) \right\} - \right. \\
& \left. - \frac{1}{n+k} \left\{ \sin \left(2\pi \frac{m+M}{N} (n+k) + \theta_n + \theta_k \right) - \right. \right. \\
& \left. \left. - \sin \left(2\pi \frac{m}{N} (n+k) + \theta_n + \theta_k \right) \right\} \right] \Bigg\}. \quad (18)
\end{aligned}$$

Для вычисления средней мощности сигнала $p_{\phi m}(t)$ в основной полосе с граничной круговой частотой $\pi/\Delta t$ необходимо вычислить квадраты модулей коэффициентов c_{mn}^{ϕ} , которые равны

$$|c_{mn}^{\phi}|^2 = \text{Re}^2 \{c_{mn}^{\phi}\} + \text{Im}^2 \{c_{mn}^{\phi}\},$$

$$\text{Re} \{c_{mn}^{\phi}\} = \frac{1}{T_2} \int_{t_0}^{t_0+T_2} p_{\phi m}(t) \cos \frac{2\pi}{T_2} n t dt, \quad (19)$$

$$\text{Im} \{c_{mn}^{\phi}\} = \frac{1}{T_2} \int_{t_0}^{t_0+T_2} p_{\phi m}(t) \sin \frac{2\pi}{T_2} n t dt. \quad (20)$$

Интеграл (19) для $n \neq 0$ можно представить в виде

$$\begin{aligned}
\frac{1}{T_2} \int_{t_0}^{t_0+T_2} p_{\phi m}(t) \cos \frac{2\pi}{T_2} n t dt &= \frac{1}{2} \left\{ \sum_{\substack{k=1 \\ kM \neq nN}}^{[N/2]} A_k \sin \theta_k + \frac{N}{2\pi} \times \right. \\
& \times \left[\sum_{\substack{k=1 \\ kM \neq nN}}^{[N/2]} A_k \frac{1}{kM - nN} \left\{ \cos \left(2\pi \frac{m}{N} k - 2\pi \frac{mn}{M} + \theta_k \right) - \right. \right. \\
& \left. \left. - \cos \left(2\pi \frac{m+M}{N} k - 2\pi \frac{mn}{M} + \theta_k \right) \right\} + \sum_{k=1}^{[N/2]} A_k \times \right. \\
& \times \frac{1}{kM + nN} \left\{ \cos \left(2\pi \frac{m}{N} k + 2\pi \frac{mn}{M} + \theta_k \right) - \right. \\
& \left. \left. - \cos \left(2\pi \frac{m+M}{N} k + 2\pi \frac{mn}{M} + \theta_k \right) \right\} \right] \Bigg\}. \quad (21)
\end{aligned}$$

Интеграл (20) для $n \neq 0$ можно представить в виде

$$\begin{aligned}
\frac{1}{T_2} \int_{t_0}^{t_0+T_2} p_{\phi m}(t) \sin \frac{2\pi}{T_2} n t dt &= \frac{1}{2} \left\{ \sum_{\substack{k=1 \\ kM \neq nN}}^{[N/2]} A_k \cos \theta_k + \frac{N}{2\pi} \times \right. \\
& \times \left[\sum_{\substack{k=1 \\ kM \neq nN}}^{[N/2]} A_k \frac{1}{kM - nN} \times \right.
\end{aligned}$$

$$\begin{aligned}
& \times \left\{ \sin \left(2\pi \frac{m+M}{N} k - 2\pi \frac{mn}{M} + \theta_k \right) - \right. \\
& \left. - \sin \left(2\pi \frac{m}{N} k - 2\pi \frac{mn}{M} + \theta_k \right) \right\} - \sum_{k=1}^{[N/2]} A_k \times \\
& \times \frac{1}{kM + nN} \left\{ \sin \left(2\pi \frac{m+M}{N} k + 2\pi \frac{mn}{M} + \theta_k \right) - \right. \\
& \left. - \sin \left(2\pi \frac{m}{N} k + 2\pi \frac{mn}{M} + \theta_k \right) \right\} \Bigg\}. \quad (22)
\end{aligned}$$

В случае $n=0$ коэффициент $c_{m,0}^{\phi}$ равен

$$\begin{aligned}
c_{m,0}^{\phi} &= \frac{1}{T_2} \int_{t_0}^{t_0+T_2} p_{\phi m}(t) dt = \frac{a_0}{2} + \frac{N}{2\pi M} \sum_{k=1}^{[N/2]} A_k \frac{1}{k} \times \\
& \times \left\{ \cos \left(2\pi \frac{m}{N} k + \theta_k \right) - \cos \left(2\pi \frac{m+M}{N} k + \theta_k \right) \right\}. \quad (23)
\end{aligned}$$

Трудоёмкость вычисления $P_t(M)$ по формулам (18), (21) – (23) равна $N^2/4 + 2N + NM$, и трудоёмкость нахождения наилучшего интервала анализа при аналитическом вычислении интегралов составит

$$\begin{aligned}
T(Q) &= \sum_{M=M_{\min}}^{M_{\max}} (N^2/4 + 2N + NM) = \\
&= aQ^2 + bQ + c, \quad (24)
\end{aligned}$$

где
и

$$\begin{aligned}
Q &= M_{\max} - M_{\min} \\
a &= N/2, \\
b &= N^2/4 + N(M_{\min} + 5/4), \\
c &= N^2/4 + N(M_{\min} + 2).
\end{aligned}$$

Трудоёмкость (24) является квадратичной относительно Q и линейной относительно $M_{\min}$, т.е. по этим параметрам лучше, чем трудоёмкость (17). Относительно N трудоёмкость (24) является квадратичной, т.е. теоретически хуже, чем трудоёмкость (17), однако необходимо учитывать, что и при численном, и при аналитическом вычислении интегралов потребуется нахождение ДПФ функции $p_{\phi}(t)$, а оно имеет квадратичную относительно N трудоёмкость.

Экспериментально было установлено, что скорость вычисления значений $P_t(M)$ с помощью формул (18), (21) – (23) существенно выше, чем с помощью численного интегрирования. Тем не менее, эта скорость остается достаточно низкой, и данный способ нахождения наилучшего интервала анализа также является малоприменимым для практического применения.

Кроме этого, вычисления по формулам (18), (21) – (23) даже при использовании чисел расширенной точности стандарта IEEE-854 дают неточный результат. Например, были получены отрицательные значения $P_t(M)$, составляющие около 2,8% от вычисленной средней мощности сигнала $p_{\phi m}(t)$ на интервале $[t_0, t_0+T_2]$. В таких случаях вычисление $P_t(M)$ с помощью численного интегрирования при достаточно малом шаге интегрирования дает более точный резуль-

тат. Это не позволяет говорить, что вычисления по формулам (18), (21) – (23) в общем случае дают более точный результат, чем результат, полученный с помощью численного интегрирования.

Таким образом, несмотря на то, что метод минимизации $P_t(M)$ теоретически дает точное решение задачи выбора интервала анализа по данному критерию, на практике получить точное решение может оказаться затруднительным.

На рис. 2, б для иллюстрации эффекта зашумления спектра приведена функция $P_t(M)$ зашумления спектра дискретного сигнала, изображенного на рис. 2, а. Как видно из примера на рис. 2, зашумление спектра дискретного сигнала увеличивается по мере увеличения разрыва функции на краях интервала анализа. Также видно, что чем больше длина интервала анализа, тем меньше зашумление спектра при той же величине разрыва функции на краях интервала. Следствием этого является то, что точки максимумов разрыва функции и зашумления спектра не совпадают. Однако точки минимумов разрыва функции и зашумления спектра совпадают, функция $P_t(M)$ на области определения $M=1, \dots, 100$ достигает минимума в трех точках: $M=1$, $M=50$ и $M=100$ (флуктуации при больших значениях $P_t(M)$ не учитываются). Функция $P_t(M)$ достигает наименьшего значения в точке $M=100$, $P_t(100) \approx 0$, поскольку спектр исходного непрерывного сигнала для $M=100$ состоит из одной гармоники.

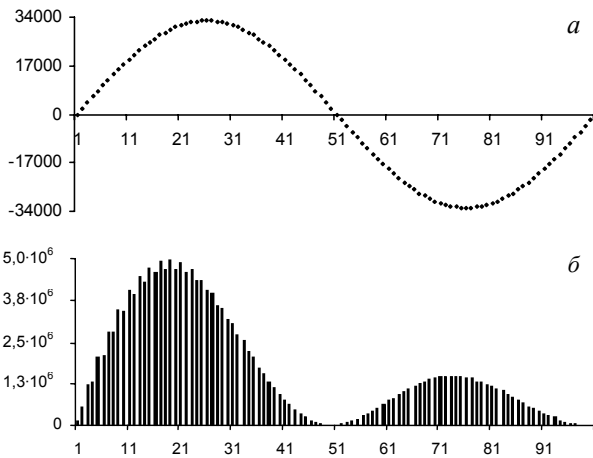


Рис. 2. Пример зашумления спектра дискретного сигнала: а) дискретный сигнал – один период дискретизованной синусоиды, $N=100$; б) функция $P_t(M)$ для $m=0$ и $M=1, \dots, 100$

Минимизация $P_t(M)$ есть минимизация абсолютно-го зашумления спектра, но можно минимизировать и относительное зашумление спектра, т.е. минимизировать величину

$$R_t(M) = \frac{P_t(M)}{P_\phi(M)},$$

где $P_\phi(M)$ есть средняя мощность сигнала $p_{\phi m}(t)$ на интервале $[t_0, t_0+T_2]$

$$P_\phi(M) = \frac{1}{T_2} \int_{t_0}^{t_0+T_2} p_{\phi m}^2(t) dt.$$

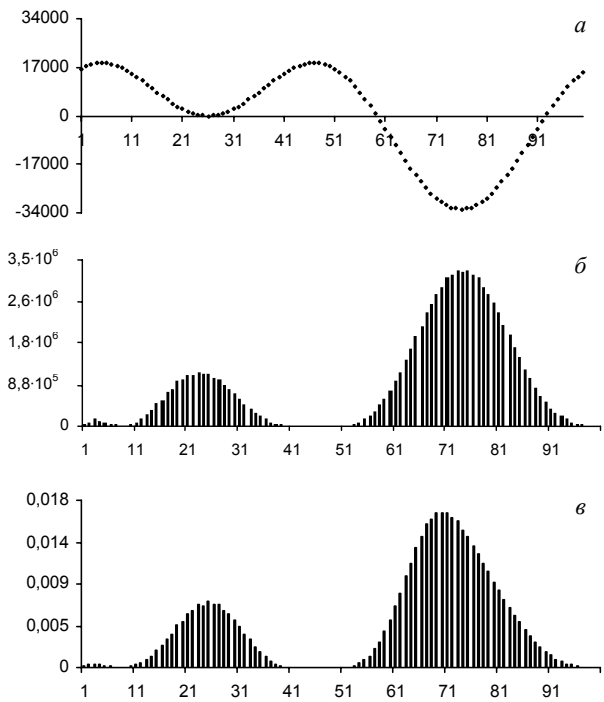


Рис. 3. Сравнение абсолютного и относительного зашумления спектра дискретного сигнала: а) дискретный сигнал – один период дискретизованной функции (25), $N=100$; б) функция $P_t(M)$ для $m=0$ и $M=1, \dots, 100$; в) функция $R_t(M)$ для $m=0$ и $M=1, \dots, 100$

Сравнение функций $P_t(M)$ и $R_t(M)$ приведено на рис. 3, б и в на примере дискретной функции (рис. 3, а), полученной при дискретизации одного периода непрерывной функции

$$p_\phi(t) = 16383(\sin \omega_1 t + \cos 2\omega_1 t), \quad (25)$$

где $\omega_1 = 2\pi/T_1$. В примере на рис. 3 функции $P_t(M)$ и $R_t(M)$ очень похожи с точностью до масштабного множителя. Точки минимумов у функций $P_t(M)$ и $R_t(M)$ на области определения $M=1, \dots, 100$ совпадают. Этими точками являются точки $M=1$, $M=8$, $M=42$, $M=50$ и $M=100$. Эти же точки являются точками минимумов разрыва функции на краях интервала анализа.

Сходство функций $P_t(M)$ и $R_t(M)$ будет наблюдаться в тех случаях, когда функция $P_\phi(M)$ является слабо осциллирующей. Для речевых сигналов это обычно выполняется. Исключением является случай, когда функция $p_{\phi m}(t)$ близка к нулю в точке t_0 . В этом случае функция $P_\phi(M)$ будет близка к нулю при небольших M с последующей стабилизацией на некотором среднем уровне при увеличении M . При таком поведении функции $P_\phi(M)$ функции $P_t(M)$ и $R_t(M)$ будут похожи с точностью до фиксированного масштабного множителя, за исключением небольших значений M , при которых значения функции $R_t(M)$ будут резко завышены. Однако, учитывая ограничение $M_{\min} \leq M$ и предполагая, что на интервале длительностью $M_{\min}$ функция $P_\phi(M)$ успевает стабилизироваться, можно в первом приближении считать, что минимизация абсолютного и относительного зашумления спектра будет давать одинаковый результат.

МЕТОД МИНИМИЗАЦИИ ВЫСОКОЧАСТОТНЫХ ГАРМОНИК ДИСКРЕТНОГО СИГНАЛА

При спектральном анализе речи рассматриваемый диапазон частот обычно ограничивают сверху значением 8 – 10 кГц [2], поскольку в этом диапазоне заключается основная доля энергии речевого сигнала, а энергия более высокочастотных гармоник пренебрежимо мала. При выполнении ДПФ на высокочастотные гармоники дискретного сигнала $p_{dm}(t)$ будут накладываться хвосты спектра исходного непрерывного сигнала $p_{fm}(t)$. Можно предположить, что чем больше средняя мощность $P_h(M)$ хвостов спектра сигнала $p_{fm}(t)$, тем большей окажется средняя мощность $P_h(M)$ высокочастотных гармоник соответствующего ему дискретного сигнала $p_{dm}(t)$. В общем случае это предположение выполняться не будет, так как при наложении хвостов спектра непрерывного сигнала на высокочастотные гармоники дискретного сигнала будет происходить как их увеличение, так и их уменьшение, в зависимости от фаз суммируемых гармоник. Однако в среднем поведение функций $P_i(M)$ и $P_h(M)$ будет схожим, и минимизацию функции $P_h(M)$ можно использовать для приближенного поиска интервалов анализа, на которых зашумление спектра дискретного сигнала минимально.

Пусть f_h – частота, ниже которой заключена основная доля энергии речевого сигнала, тогда средняя мощность высокочастотных гармоник дискретного сигнала $p_{dm}(t)$ будет равна

$$P_h(M) = 2 \sum_{n=n_{\min}}^{n_{\max}} |\bar{c}_{mn}^{\partial}|^2, \quad (26)$$

где $n_{\min} = \lceil f_h M \Delta t \rceil$, $n_{\max} = \lfloor M/2 \rfloor$, а коэффициенты $\bar{c}_{mn}^{\partial}$ задаются аналогично (6).

Пример функции $P_h(M)$ для $f_h = 15000$ Гц, $\Delta t = 1/44100$ с и дискретного сигнала, изображенного на рис. 2,а, приведен на рис. 4.

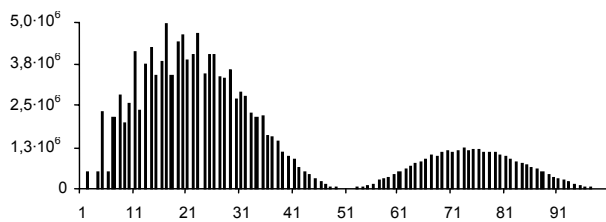


Рис. 4. Пример функции $P_h(M)$ для дискретного сигнала, изображенного на рис. 2,а, $f_h=15000$ Гц, $\Delta t=1/44100$ с, $m=0$, $M=1, \dots, 100$

Сравнивая функции на рис. 2,б и 4, можно обнаружить их сходство, за возможным исключением небольших значений M ($M=1, \dots, 30$), при которых функция $P_h(M)$ ведет себя достаточно неустойчиво. Как и следовало ожидать, значения функции $P_h(M)$ оказываются несколько заниженными в сравнении со значениями функции $P_i(M)$. На области определения $M=40, \dots, 100$ функция $P_h(M)$ достигает минимума в двух точках: $M=51$ и $M=100$. Второе решение ($M=100$) совпадает с решением, полученным с помощью точного метода выбора интервала анализа. В этом случае

функция $P_h(M)$ достигает наименьшего значения $P_h(100) \approx 0,053$.

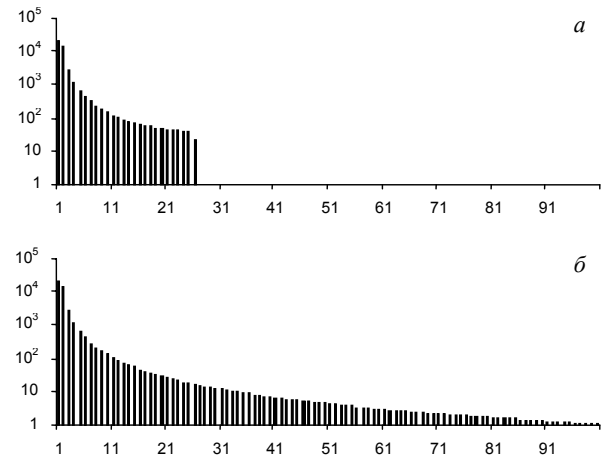


Рис. 5. Сравнение спектров дискретного, изображенного на рис. 2,а, и исходного непрерывного сигналов для интервала анализа длиной $M=50$ отсчетов: а) амплитудный спектр дискретного сигнала; б) амплитудный спектр исходного непрерывного сигнала, показаны только первые 100 коэффициентов

Решение $M=51$ очень близко к точному решению $M=50$, но все же отличается от него. Несмотря на то, что средняя мощность хвостов спектра сигнала $p_{fm}(t)$ в случае $M=51$ больше (рис. 6,б), чем в случае $M=50$ (рис. 5,б), их суммирование с высокочастотными гармониками сигнала $p_{dm}(t)$ приводит к тому, что в случае $M=51$ высокочастотные гармоники оказываются заниженными (рис. 6,а) в сравнении со случаем $M=50$ (рис. 5,а). Для $f_h=15000$ Гц, $\Delta t=1/44100$ с и $M=50, 51$ номер нижней высокочастотной гармоники равен $n_{\min}=18$.

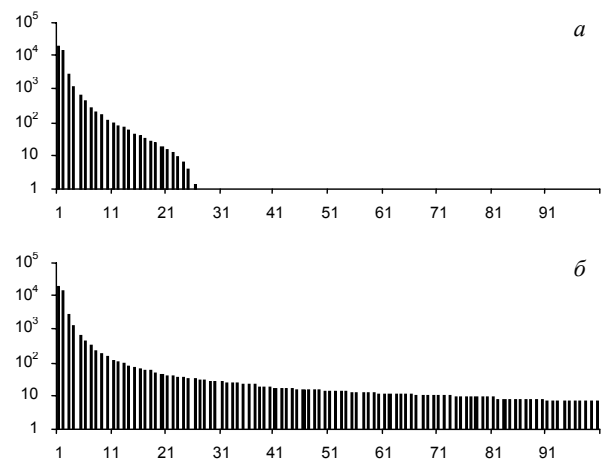


Рис. 6. Сравнение спектров дискретного, изображенного на рис. 2,а, и исходного непрерывного сигналов для интервала анализа длиной $M=51$ отсчетов: а) амплитудный спектр дискретного сигнала; б) амплитудный спектр исходного непрерывного сигнала, показаны только первые 100 коэффициентов

Трудоемкость вычисления $P_h(M)$ по формуле (26) составляет $M(n_{\max} - n_{\min} + 1)$, ее можно оценить сверху величиной

$$T_M = (1/2 - f_h \Delta t) M^2 + 2M,$$

отсюда трудоемкость нахождения наилучшего интервала анализа методом минимизации высокочастотных гармоник дискретного сигнала составит

$$T(Q) = \sum_{M=M_{\min}}^{M_{\max}} T_M = aQ^3 + bQ^2 + cQ + d, \quad (27)$$

где

$$Q = M_{\max} - M_{\min};$$

$$a = z/3;$$

$$b = zM_{\min} + z/2 + 1;$$

$$c = zM_{\min}^2 + (z+2)M_{\min} + z/6 + 1;$$

$$d = zM_{\min}^2 + 2M_{\min};$$

$$z = 1/2 - f_h \Delta t.$$

Трудоемкость (27) похожа на трудоемкость (17) за исключением того, что в формуле (27) отсутствует множитель CN . На первый взгляд, трудоемкость (24) лучше, чем (27), но если учесть, что $N > Q$ и $N > M_{\min}$, то трудоемкость (27) оказывается лучше трудоемкости (24). Кроме того, для вычисления $P_h(M)$ не надо выполнять ДПФ для всего сигнала $p_\phi(t)$.

В эксперименте скорость вычисления $P_h(M)$ оказалась весьма высокой, что делает метод минимизации высокочастотных гармоник дискретного сигнала пригодным для практического применения.

Вместо минимизации функции $P_h(M)$ можно минимизировать величину

$$R_h(M) = \frac{P_h(M)}{P_o(M)}$$

— аналог относительного зашумления спектра, где $P_o(M)$ есть средняя мощность дискретного сигнала $p_{om}(t)$ на интервале $[t_0, t_0+T_2]$

$$P_o(M) = \sum_{n=-[M/2]}^{[M/2]} |\bar{c}_{mn}^o|^2.$$

МЕТОД МИНИМИЗАЦИИ РАЗРЫВА ФУНКЦИИ НА КРАЯХ ИНТЕРВАЛА АНАЛИЗА

Для приближенного поиска интервалов анализа, на которых зашумление спектра дискретного сигнала минимально, очевидным является использование метода минимизации разрыва функции на краях интервала анализа. Минимизируемой величиной в этом случае будет функция $c(M)$, даваемая выражением (15).

В примерах на рис. 2 и 3 функция $c(M)$ достигает минимума в тех же точках, что и функция $P_f(M)$, т.е. в этих случаях решение, полученное с помощью приближенного метода минимизации разрыва функции на краях интервала анализа, совпадает с точным решением.

Трудоемкость приближенного поиска наилучшего интервала анализа методом минимизации разрыва функции на краях интервала является линейной и равна

$$T(Q) = Q.$$

СРАВНЕНИЕ ТОЧНОГО И ПРИБЛИЖЕННЫХ МЕТОДОВ ВЫБОРА ИНТЕРВАЛА АНАЛИЗА

На рис. 7, б–г приведены функции $P_f(M)$, $P_h(M)$ и $c(M)$, вычисленные для речевого сигнала, представляющего собой реализацию фонемы /и/ длительно-стью 298 мс. Вычисления проведены для $m=4400$ и $M=71, \dots, 491$. В точке $m=4400$ значение сигнала равно нулю, и легко видно, что чем больше значение сигнала на правой границе интервала анализа отклоняется от нуля, тем больше зашумление спектра дискретного сигнала. В целом функции $P_f(M)$, $P_h(M)$ и $c(M)$ очень похожи, но наибольший интерес представляют точки минимумов этих функций.

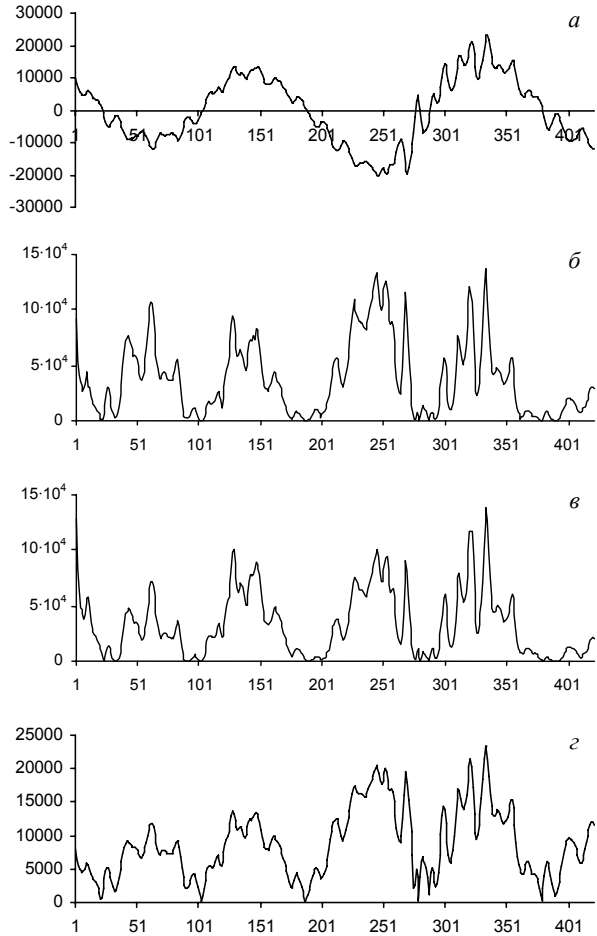


Рис. 7. Сравнение функций $P_f(M)$, $P_h(M)$ и $c(M)$ на примере реального речевого сигнала: а) фрагмент речевого сигнала, номер левого граничного отсчета равен 4470, ему соответствует номер 1 на оси абсцисс, длина фрагмента — 421 отсчет, $N=13149$; б) функция $P_f(M)$; в) функция $P_h(M)$, $f_h=15000$ Гц, $\Delta t=1/44100$ с; г) функция $c(M)$. Для всех трех функций левая граница интервалов анализа равна $m=4400$, длина интервалов меняется от $M_{\min}=71$ до $M_{\max}=491$, по оси абсцисс отложены номера интервалов

Первые десять минимумов с наименьшими значениями функции $P_f(M)$ приведены в табл. 1. Точки первых двух минимумов очень близки друг к другу, так что в практических приложениях их можно объединить в одну, поскольку за время двух отсчетов речевой сигнал не может существенно измениться.

Таблица 1

Длина интервала, M	Номер интервала	$P_d(M)$
173	103	73,4
171	101	134,0
257	187	167,8
449	379	176,5
460	390	266,2
357	287	478,0
349	279	506,8
91	21	920,3
432	362	1183,1
345	275	1444,2

Первые десять минимумов с наименьшими значениями функции $P_h(M)$ приведены в табл. 2. Здесь точки первых двух минимумов также очень близки друг к другу, и их можно объединить в одну. В первую десятку минимумов функции $P_h(M)$ вошли две точки $M=103$ и $M=161$, для которых зашумление спектра достаточно велико ($P_d(103)=2704,3$; $P_d(161)=2697,5$), что объясняется приближенным характером метода минимизации высокочастотных гармоник дискретного сигнала. Точки оставшихся восьми минимумов функции $P_h(M)$ очень близки к соответствующим точкам минимумов функции $P_d(M)$ из табл. 1, максимальное расхождение между точками минимумов составляет 3 отсчета. Точное совпадение минимумов функций $P_d(M)$ и $P_h(M)$ происходит в единственной точке $M=357$. Длина интервала $M=357$ отсчетов замечательна еще тем, что она равна периоду основного тона данного фрагмента речевого сигнала. Также необходимо обратить внимание на то, что точки минимумов из табл. 2 расположены в совершенно ином порядке, чем точки из табл. 1.

Первые десять минимумов с наименьшими значениями функции $c(M)$ приведены в табл. 3. Как и в случае функции $P_h(M)$, в первую десятку минимумов функции $c(M)$ вошли точки $M=103$ и $M=161$, для которых зашумление спектра достаточно велико. Точки оставшихся восьми минимумов функции $c(M)$ полно-

стью совпадают с соответствующими точками минимумов функции $P_d(M)$ из табл. 1, т.е. в данном случае метод минимизации разрыва функции на краях интервала анализа показывает лучший результат, чем метод минимизации высокочастотных гармоник дискретного сигнала. Этот факт вместе с линейной трудоемкостью метода минимизации разрыва функции делает его предпочтительным в сравнении с методом минимизации высокочастотных гармоник. Порядок расположения точек в табл. 3 отличается от того же порядка в табл. 1. Это говорит о том, что по величине $c(M)$ разрыва функции на краях интервала анализа можно лишь очень грубо судить о величине $P_d(M)$ зашумления спектра дискретного сигнала.

Таблица 2

Длина интервала, M	Номер интервала	$P_h(M)$
459	389	35,3
461	391	90,0
260	190	117,8
450	380	123,8
172	102	129,3
103	33	152,6
161	91	252,7
93	23	475,9
350	280	491,8
357	287	580,9

Таблица 3

Длина интервала, M	Номер интервала	$c(M)$
173	103	3
257	187	47
349	279	164
449	379	347
91	21	555
460	390	995
357	287	1274
103	33	1588
161	91	2013
345	275	2172

ЛИТЕРАТУРА

1. Фант Г. Акустическая теория речеобразования: Пер. с англ. М.: Наука, 1964. 284 с.
2. Кодзасов С.В., Кривнова О.Ф. Общая фонетика. М.: Рос. гос. гуманит. ун-т, 2001. 592 с.
3. Голд Б., Рэйдер Ч. Цифровая обработка сигналов: Пер. с англ. М.: Сов. радио, 1973. 368 с.
4. Гоноровский И.С. Радиотехнические цепи и сигналы. М.: Радио и связь, 1986. 512 с.
5. Оппенгейм А.В., Шафер Р.В. Цифровая обработка сигналов: Пер. с англ. М.: Связь, 1979. 416 с.
6. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов: Пер. с англ. М.: Радио и связь, 1981. 496 с.
7. Рабинер Л., Гоулд Б. Теория и применение цифровой обработки сигналов: Пер. с англ. М.: Мир, 1978. 848 с.
8. Толстов Г.П. Ряды Фурье. М.: Наука, 1980. 384 с.

Статья представлена кафедрой прикладной информатики факультета информатики Томского государственного университета, поступила в научную редакцию 23 июля 2003 г.