

УДК 621.3

Ф.В. Колокольцев, Г.Н. Розоринов, д-р техн. наук

Выбор типа порога для сжатия речевых сигналов

Рассмотрены типы порогов, используемых для сжатия речевых сигналов. Показано, что качество восстановленной речи напрямую зависит от выбора характеристики порогования. Дано математическое описание спектрального состава сигнала, подвергнутого порогованию, выявлена его связь с величиной порога и рассчитан коэффициент нелинейных искажений. Построены графики выявленных зависимостей, по которым можно судить о выборе типа и величины порога.

The threshold types for the compression of vocal signals are considered. It is shown that quality of the recovered speech directly depends on the choice of threshold description. Mathematical description of spectral composition of threshold signal is given. Relationship of spectral composition with the threshold level is established and a klirr factor is calculated. The graphs of the established dependences allowed to estimate a typeselection and threshold level are built.

Ключевые слова: сжатие, порог, спектральный состав, речевой сигнал, коэффициент нелинейных искажений.

Введение

Настоящая работа тесно связана с реальной проблемой, возникающей при обработке сигнала в сотовых телефонных системах, ввиду присутствия в них фоновых шумов. Когда речевой сигнал передается с сотового телефона на свою базовую станцию (или в другом направлении), он должен быть сжатым для того, чтобы оператор мог передавать столько сигналов, сколько позволяет реальная пропускная способность канала [1].

Для бесперебойной работы схемы сжатия необходимо обеспечить высокое значение отношения сигнал/шум на ее входе. Поэтому удаление фонового шума производится перед процессом сжатия. Это может выполнить идеальный фильтр, который пропускает только речевой сигнал и удаляет нежелательные шумы, например, шум автомобилей или разговор других людей. Очевидно, что практически эта задача является довольно сложной, поэтому предпринимаются попытки использовать различные методы обработки сигнала, наиболее успешны-

ми из которых являются пороговые [2, 3 – 7]. При этом, однако, отсутствуют сведения о влиянии типа порогования и уровня порога на качественные характеристики восстановленного сигнала.

Целью настоящей работы является исследование разных пороговых способов снижения уровня фонового шума при сжатии речевых сигналов и выбор наилучшего из них по критерию качества восстановленного сигнала.

Основная часть

Пусть входной сигнал x_l состоит из речевой s_l и шумовой n_l компонент:

$$x_l = s_l + n_l. \quad (1)$$

Необходимо применить такой алгоритм обработки сигнала x_l , чтобы уменьшить шумовую компоненту n_l , или даже получить только компоненту s_l на выходе системы обработки.

При этом очевидными являются следующие требования:

- устранить или уменьшить фоновый шум во время и вне речевого интервала;
- как можно меньше воздействовать на качество речи;
- не изменять характеристики фонового шума;
- обеспечить процедуру обработки сигнала в режиме реального времени.

В основе пороговых способов снижения шума лежит уменьшение значений коэффициентов преобразования сигнала при его переносе из одной области представления в другую (время–частота, время–пространство, и др.). Предполагается, что шумовая составляющая отображается малыми коэффициентами и пороговый способ используется для уменьшения или полного удаления малых коэффициентов. Затем сигнал подвергается обратному преобразованию.

Наиболее часто используются мягкий и жесткий пороги [2, 3 – 5]. Поскольку предполагается, что алгоритм обработки должен действовать в режиме реального времени, а длина N входного сигнала может быть большой, то входной сигнал делят на небольшие сегменты (кадры). Алгоритм обработки применяется на каждом сегменте, а выходной результат представляется в виде композиции отдельных обра-

ботанных сегментов. Обработка отдельного сегмента требует времени, что в итоге ведет к задержке сигнала в системе связи.

Задержка в системе связи не может быть слишком длительной, поэтому максимальный размер сегмента ограничен [1]. Таким образом задержка является функцией частоты дискретизации. В том случае, когда частота дискретизации равна 8 кГц, а сегмент содержит 128 отсчетов, задержка составляет 16 мс.

Обозначим величину порога θ . Тогда в случае жесткого порога функциональная характеристика (ФХ) порогования описывается уравнением

$$f(x) = [1(|x| - \theta)]x, \quad (2)$$

где $1(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0, \end{cases}$ – единичная функция.

График ФХ при жестком пороговании показан на рис. 1,а.

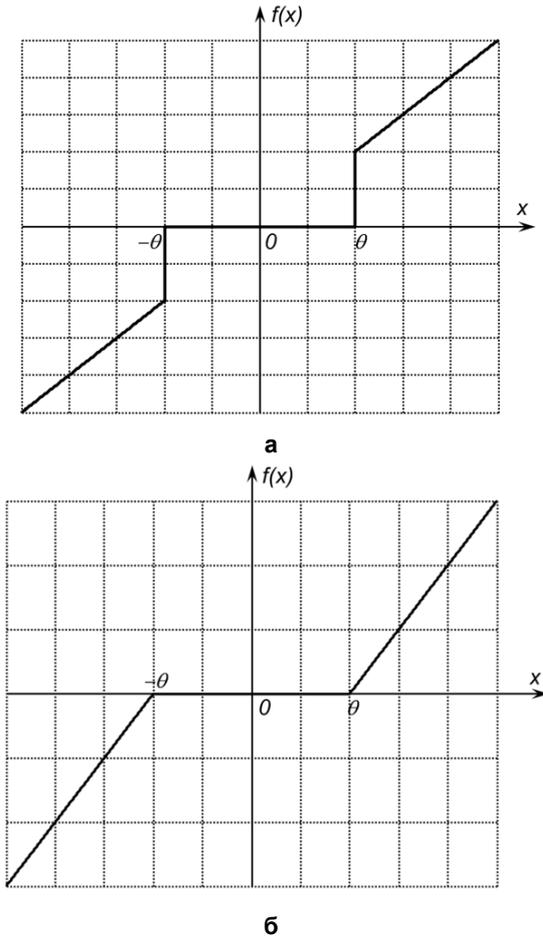


Рис.1. Функциональные характеристики порогования: а – жесткого, б – мягкого

В случае использования мягкого порога ФХ имеет вид:

$$f(x) = [1(|x| - \theta)][x - \theta \text{sign}(x)], \quad (3)$$

а график такой ФХ показан на рис. 1,б.

Анализ графиков, представленных на рис.1 свидетельствует о том, что наибольшее влияние на качество передачи речи, особенно при наличии фоновых шумов, оказывают начальные участки ФХ. Поэтому попытаемся видоизменить эти участки так, чтобы обеспечить наименьшие искажения восстановленного сигнала. Наиболее простыми решениями этой задачи являются реализация ФХ порогования типов, показанных на рис. 2. Назовем их ФХ линейного супермягкого порогования (рис. 2,а) и ФХ квадратичного супермягкого порогования (рис. 2,б).

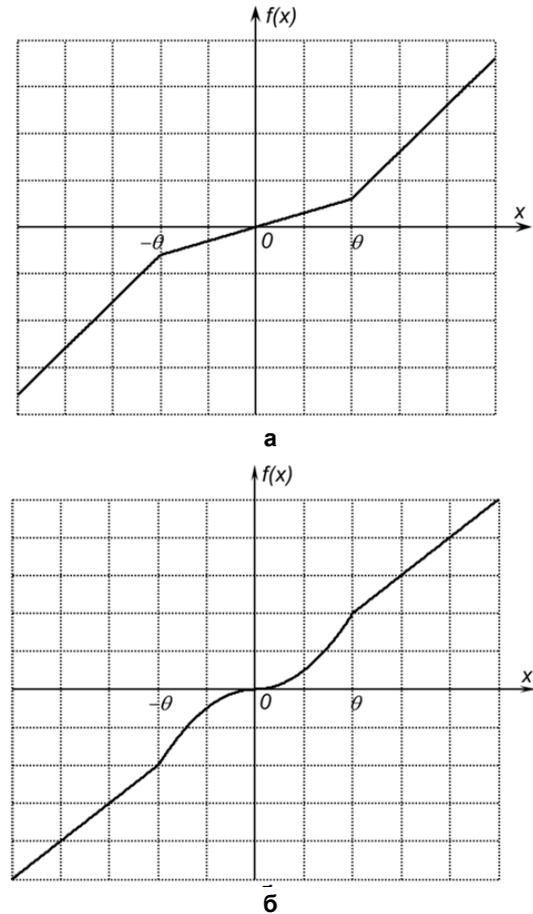


Рис.2. Функциональные характеристики супермягкого порогования: а – линейного, б – квадратичного

ФХ при линейном и квадратичном супермягком пороговании имеют вид, соответственно:

$$f(x) = \begin{cases} x - \text{sign}(x)(1 - \beta)\theta, & |x| \geq \theta, \\ \beta x, & |x| < \theta. \end{cases} \quad (4)$$

$$f(x) = \begin{cases} x, & |x| \geq \theta, \\ \frac{x^2}{\theta} \text{sign}(x), & |x| < \theta. \end{cases} \quad (5)$$

В формуле (4) β – угловой коэффициент, определяющий наклон начального линейного участка ФХ.

При сжатии речевого сигнала с потерями, с ростом коэффициента сжатия (величины порога

θ) возрастают и искажения сигнала [2, 3]. Это проявляется в появлении паразитных гармонических составляющих в восстановленном речевом сигнале. Для примера на рис. 3 показано порогование синусоидального сигнала.

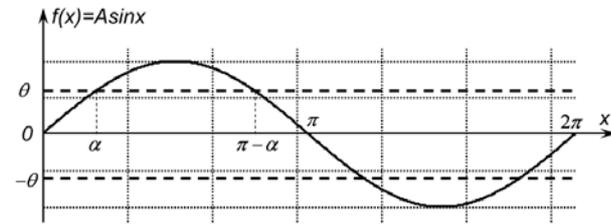


Рис. 3. Порогование синусоидального сигнала

Очевидно, что величина порога $\theta = |A \sin x|$, и если обозначить моменты достижения сигналом порога через α , то получим $\alpha = \arcsin\left(\frac{\theta}{A}\right)$.

На рис. 4 показаны диаграммы синусоидального сигнала, полученные при использовании жесткого, мягкого и супермягких типах порогования.

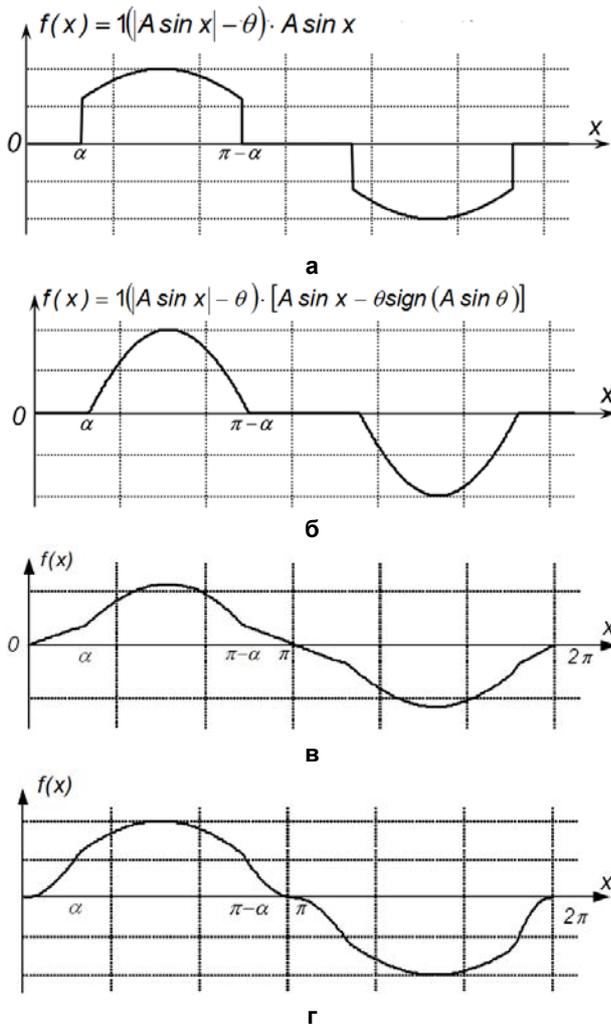


Рис. 4. Диаграммы синусоидального сигнала при использовании различных типов порогования: а – жесткого, б – мягкого, в – линейного супермягкого, г – квадратичного супермягкого

Разложим показанные на рис. 4 сигналы в ряды Фурье. Тогда получим:

для жесткого порога

$$(b_n)_{Ж} = \frac{4A\sqrt{\cos^2 \alpha + n^2 \sin^2 \alpha}}{\pi(1-n^2)} \times \sin \frac{\pi n}{2} \sin \left[\arctg \frac{\text{ctg} \alpha}{n} - n \arccos(\sin \alpha) \right], \quad (6)$$

для мягкого порога

$$(b_n)_M = \frac{4A}{\pi(n+1)} \sin \frac{\pi n}{2} \left\{ \frac{\sin[(n-1)(\pi/2-\alpha)]}{n-1} - \frac{\sin \alpha \sin[\pi(\pi/2-\alpha)]}{n} \right\}, \quad (7)$$

для линейного супермягкого порога

$$(b_n)_{ЛСМ} = \frac{2A(1-\beta)}{\pi} \sin \frac{\pi n}{2} \left\{ \frac{\sin[(n-1)(\pi/2-\alpha)]}{n-1} - \frac{2 \sin \alpha \sin[n(\pi/2-\alpha)]}{n} + \frac{\sin[(n+1)(\pi/2-\alpha)]}{n+1} \right\} + \frac{\beta}{\pi(n-1)} \sin[\pi(n-1)], \quad (8)$$

для квадратичного супермягкого порога

$$(b_n)_{КСМ} = \frac{2A}{\pi} \sin \frac{\pi n}{2} \left\{ \frac{\sin[(n-1)(\pi/2-\alpha)]}{n-1} + \frac{\sin[(n+1)(\pi/2-\alpha)]}{n+1} + \frac{1}{\sin \alpha} \left[\frac{2}{n} \sin \frac{n\alpha}{2} \cos\left[\frac{n}{2}(\pi-\alpha)\right] - \frac{\cos[(\pi-\alpha)n/2-\alpha] \sin[(n/2+1)\alpha]}{n+2} - \frac{\cos[(\pi-\alpha)n/2+\alpha] \sin[(n/2-1)\alpha]}{n-2} \right] \right\}. \quad (9)$$

Построенные по выражениям (6 – 9) графики амплитудных спектров этих сигналов для значений $\sin \alpha = 0,6$ показаны на рис. 5.

Из рис. 5 видно, что амплитуда первой гармоники после порогования уменьшилась по сравнению с исходным сигналом, а в спектрах появились третья, пятая и другие нечетные гармоники, искажающие исходный сигнал. Для различных типов порогования отношение первой гармоники к амплитуде исходного синусоидального сигнала определяется по формулам:

$$\left(\frac{b_1}{A}\right)_{Ж} = \frac{1}{\pi}(\sin 2\alpha + \pi - 2\alpha), \quad (10)$$

$$\left(\frac{b_1}{A}\right)_M = \frac{1}{\pi}(\pi - 2\alpha - \sin 2\alpha), \quad (11)$$

$$\left(\frac{b_1}{A}\right)_{ЛСМ} = 1 - \frac{1-\beta}{\pi}(2\alpha + \sin 2\alpha), \quad (12)$$

$$\left(\frac{b_1}{A}\right)_{КСМ} = \frac{1}{\pi} \left\{ \frac{1}{\sin \alpha} \left[3(1 - \cos \alpha) - \frac{1 - \cos 3\alpha}{3} \right] + \pi - 2\alpha + \sin 2\alpha \right\}. \quad (13)$$

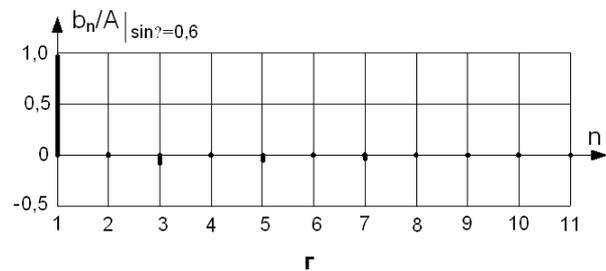
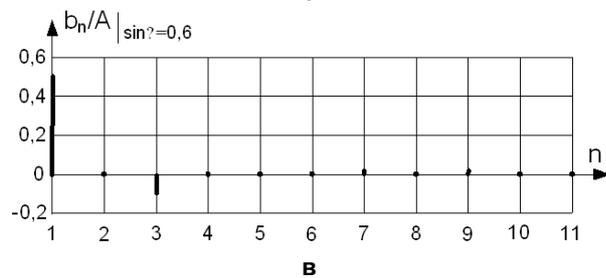
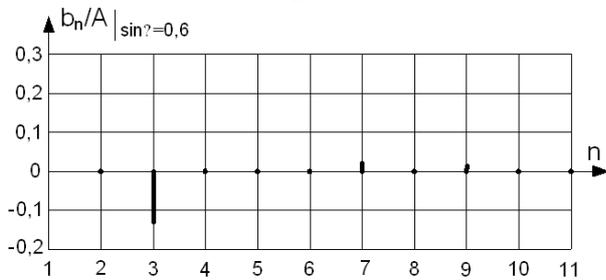
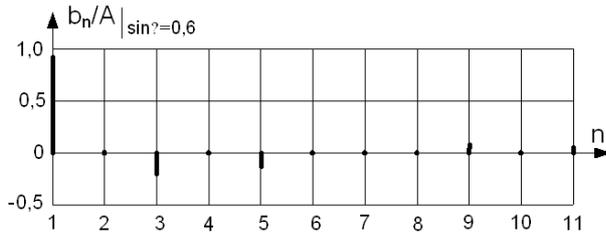


Рис. 5. Амплитудные спектры синусоидального сигнала при использовании различных типов порогования: а – жесткого, б – мягкого, в – линейно-супермягкого, г – квадратично-супермягкого

Для определения отношения сигнал/шум при различных типах порогования воспользуемся энергетическими соотношениями. Энергию сигнала $P_{эф}$ с периодом T можно определить как действующее его значение

$$P_{эф} = \sqrt{\frac{1}{T} \int_0^T f^2(x) dx}. \quad (14)$$

В этом случае энергия исходного синусоидального сигнала $f(x) = A \sin x$ равна $A/\sqrt{2}$. Отношение энергии синусоидального сигнала подвергнутого порогованию к энергии исходного сигнала нетрудно найти по формулам:

$$\left(\frac{P_{эф}}{A/\sqrt{2}}\right)_{Ж} = \sqrt{\frac{1}{\pi}(\sin 2\alpha + \pi - 2\alpha)}, \quad (15)$$

$$\left(\frac{P_{эф}}{A/\sqrt{2}}\right)_M = \sqrt{\frac{1}{\pi} \left[\pi - 2\alpha - 3 \sin 2\alpha + 2(\pi - 2\alpha) \sin^2 \alpha \right]}, \quad (16)$$

$$\left(\frac{P_{эф}}{A/\sqrt{2}}\right)_{ЛСМ} = \sqrt{\frac{1}{\pi} (1-\beta)(\beta-3) \sin 2\alpha - 2\alpha(1-\beta)^2 + \pi + 2(\pi-2\alpha)(1-\beta)^2 \sin^2 \alpha}, \quad (17)$$

$$\left(\frac{P_{эф}}{A/\sqrt{2}}\right)_{КСМ} = \sqrt{\frac{1}{\pi} \left[\frac{1}{\sin^2 \alpha} \left(\frac{3}{2} \alpha - \sin 2\alpha + \frac{1}{8} \sin 4\alpha \right) + \pi - 2\alpha + \sin 2\alpha \right]}, \quad (18)$$

На рис. 6 представлены графики, построенные по формулам (10–13) и (15–18).

Из рис. 6 следует, что по сравнению с жестким порогом (рис. 6, а) при использовании мягкого порога резко снижается уровень первой гармоники выходного сигнала, а, следовательно, и его энергия, и уже при $\sin \alpha = 0,5 - 0,7$ обе эти характеристики асимптотически стремятся к нулю (рис. 6, б). В случае применения линейного супермягкого порогования эти характеристики стремятся к значению β (рис. 6, в). Отметим, что при использовании жесткого порога значения амплитудной и энергетической характеристик остаются высокими вплоть до значения $\sin \alpha \approx 0,5$, а затем они резко уменьшаются до нуля. И только при использовании квадратично-супермягкого порога значения исследуемых характеристик остаются высокими для всех значений порога (рис. 6 г).

Количество искажения сигнала после порогования можно определить с помощью коэффициента нелинейных искажений χ как отношение действующего значения выходного сигнала без его первой гармоники к действующему значению входного сигнала (в рассматриваемом случае синусоидального сигнала):

$$\chi = \frac{\sqrt{\frac{1}{2} \sum_{n=2}^{\infty} b_n^2}}{A/\sqrt{2}} = \frac{\sqrt{P_{эф}^2 - \frac{1}{2} b_1^2}}{A/\sqrt{2}}. \quad (19)$$

Подставив в (19) выражения (10 – 13) и (15 – 18), получим формулы для расчета коэффициентов нелинейных искажений сигнала при использовании различных типов порогов:

$$(\chi)_{Ж} = \frac{1}{\pi} \sqrt{(\pi - 2\alpha + \sin 2\alpha)(2\alpha - \sin 2\alpha)}, \quad (20)$$

$$(\chi)_M = \frac{1}{\pi} \sqrt{\pi \left[2\alpha - \sin 2\alpha + 2(\pi - 2\alpha) \sin^2 \alpha \right] - (2\alpha + \sin 2\alpha)^2}, \quad (21)$$

$$(\chi)_{ЛСМ} = \left\{ \frac{1}{\pi} \left[(1-\beta)(\beta-3) \sin 2\alpha - 2\alpha(1-\beta^2) + \pi + 2(\pi-2\alpha)(1-\beta)^2 \sin^2 \alpha \right] - \left[1 - \frac{1-\beta}{\pi} (2\alpha + \sin 2\alpha)^2 \right]^{\frac{1}{2}} \right\} \quad (22)$$

$$(\chi)_{КСМ} = \frac{1}{\pi} \left\{ \pi \left[\frac{1}{\sin^2 \alpha} \left(\frac{3}{2} \alpha - \sin 2\alpha + \frac{1}{8} \sin 4\alpha \right) + \pi - 2\alpha + \sin 2\alpha \right] - \left[\frac{1}{\sin \alpha} \left(3 - 3 \cos \alpha - \frac{1 - \cos 3\alpha}{3} \right) + \pi - 2\alpha + \sin 2\alpha \right]^2 \right\}^{\frac{1}{2}}. \quad (23)$$

Рассчитанные по этим формулам коэффициенты нелинейных искажений показаны в виде графиков на рис. 7.

Из рис.7 видно, что наибольшие искажения сигнал претерпевает при использовании жесткого порогования. Кроме того, в зависимости от типа порога максимум значения коэффициента нелинейных искажений соответствует разным значениям $\sin \alpha$. Только в случае мягкого и линейного супермягкого порогов эти значения совпадают и равны 0,3.

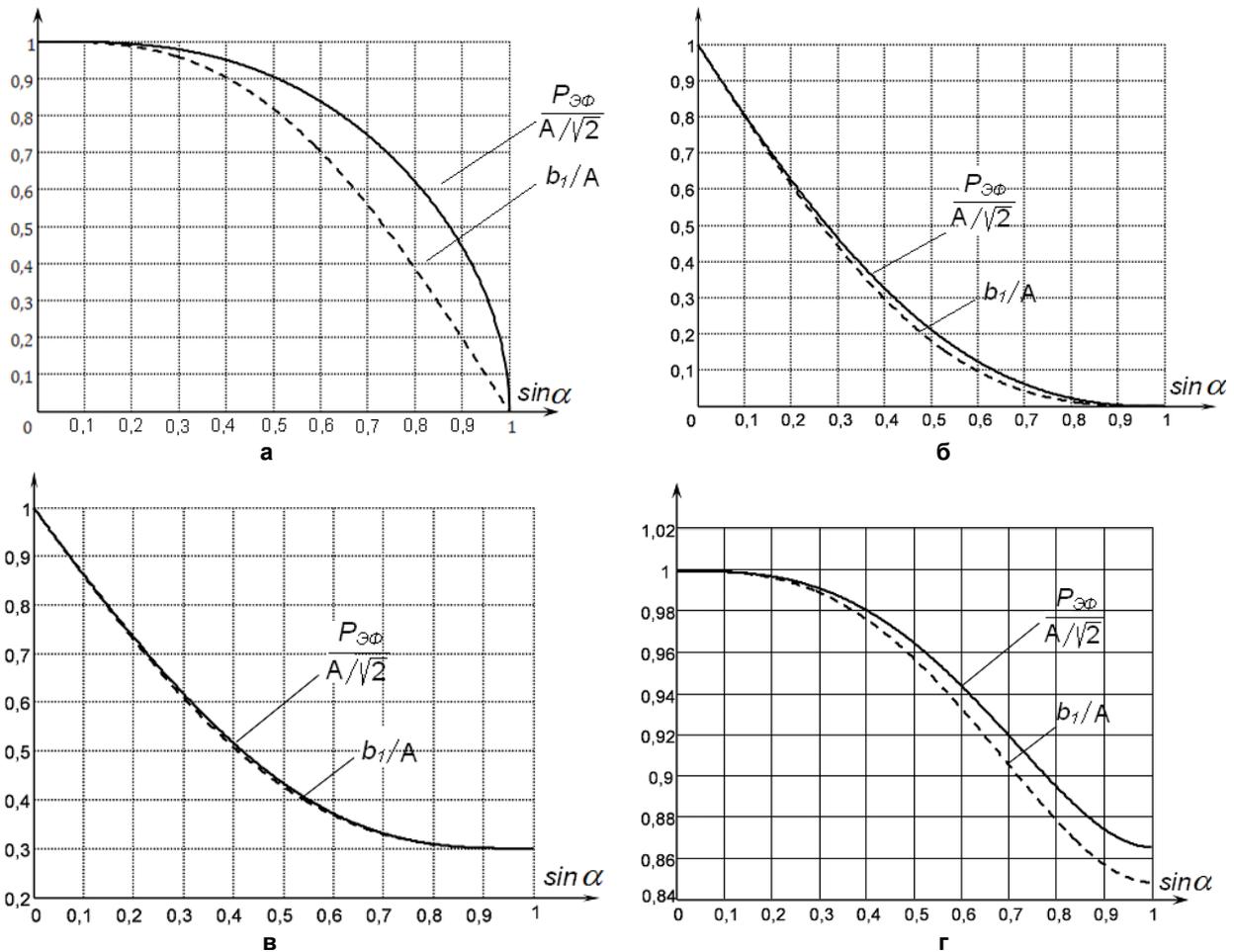


Рис. 6. Зависимости функций b_1/A и $\frac{P_{эф}}{A/\sqrt{2}}$ от величины и типа порога: а – жесткого, б – мягкого, в – линейного супермягкого, г – квадратичного супермягкого

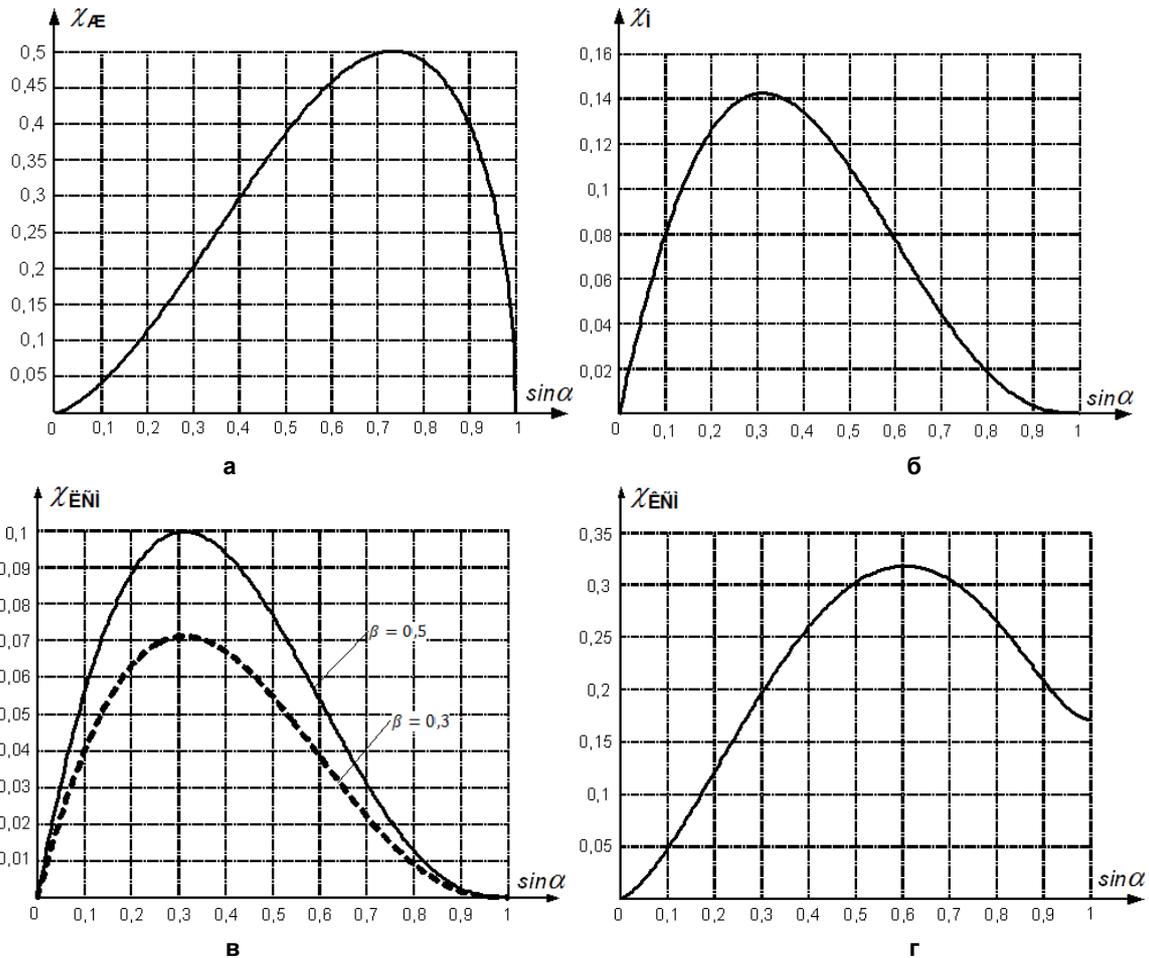


Рис. 7. Зависимости коэффициента нелинейных искажений от величины и типа порога: а – жесткого, б – мягкого, в – линейного супермягкого, г – квадратичного супермягкого

Выводы

1. Показано, что наибольшее влияние на качество передачи речевых сигналов, особенно при наличии фоновых шумов, оказывает вид начального участка функциональной характеристики порогования.

2. Изучены 4 типа функциональных характеристик порогования, отличающихся видами начальных участков. Дано математическое описание спектрального состава синусоидального сигнала, подвергнутого порогованию, выявлена его связь с величиной порога и рассчитан коэффициент нелинейных искажений. Построены графики выявленных зависимостей, которые помогают выбрать тип и величину порога.

Литература

1. Шелухин О.И., Лукьянцев Н.Ф. Цифровая обработка и передача речи / Под. ред. О.И. Шелухина. – М.: Радио и связь, 2000. – 456 с.

2. Розоринов Г.Н., Колокольцев Ф.В. Компандирование речевых сигналов для цифровой передачи по полосноограниченным каналам // Вісник Державного університету інформаційно-комунікаційних технологій. – 2006. – Т.4, №2. – С. 73 – 79.
3. Фёдоров Е.Е. Модели и методы преобразования речевых сигналов. – Донецк: Норд-Пресс, 2006. – 260 с.
4. Reccione M.C. The enhanced variable rate coder: Toll quality speech for CDMA // International Journal of Speech Technology. – 1999. – № 2. – P. 305–315.
5. Thiemann J. Acoustic Noise Suppression for Speech Signals using Auditory Masking Effects. – Montreal: McGill University. – 2001. – 83 p.
6. Дьяконов В., Абраменкова И. MATLAB. Обработка сигналов и изображений: Специальный справочник. – СПб.: Питер. – 2002. – 608 с.
7. Малла С. Вейвлеты в обработке сигналов: Пер. с англ. – М.: Мир, 2005. – 671 с.