

ИНФОРМАТИКА

УДК 621.371.39:681.322.01

ПЕРЦЕПТУАЛЬНЫЙ МЕТОД ПОВЫШЕНИЯ КАЧЕСТВА ЧАСТОТНО-ОГРАНИЧЕННОГО РЕЧЕВОГО СИГНАЛА: ОДНОКАНАЛЬНЫЙ ВАРИАНТ, КОМБИНИРОВАННАЯ СИСТЕМА, ЧАСТОТНЫЙ ПОДХОД

А.Е. АНОШЕНКО, А.А. ПЕТРОВСКИЙ

*Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, 220013, Беларусь**Поступила в редакцию 19 августа 2003*

В данной работе предлагается одноканальная комбинированная система редактирования шумов окружающей среды и подавления эхо-сигнала на основе обработки в частотной области и нового психоакустического мотивированного правила спектрального взвешивания. Перцептуальный подход повышения качества речевого сигнала данной системы характеризуется по сравнению с известными на порядок меньшей вычислительной сложностью, единым для редактирования шума и эхо-сигнала правилом спектрального взвешивания.

Ключевые слова: эхо-сигнал, спектральное взвешивание.

Введение

В настоящее время громкоговорящие устройства становятся все более и более распространенными в современных системах связи, а в некоторых случаях и просто необходимыми, в частности там, где является опасным использование традиционных устройств (например, телефоны установленные в автомобилях). В таких системах используются микрофоны, расположенные на достаточном удалении от говорящего, и свободно расположенные громкоговорители. Это приводит к тому, что сигнал, снимаемый с микрофона, содержит шумы окружающей среды, а также акустическое эхо от громкоговорителя. Эти две искажающие составляющие сигнала могут значительно ухудшать качество речи или даже делать ее невозможной для восприятия, поэтому система громкоговорящей связи должна иметь оборудование для их устранения. На рис. 1 показана схема построения комбинированной системы подавления акустического эха и редактирования шумов окружающей среды, которая нашла наибольшее применение на сегодняшний день [1]. Сигнал на входе микрофона в такой системе описывается следующим выражением:

$$y(n) = z(n) + d(n) + e(n), \quad (1)$$

где $z(n)$ — речевой сигнал ближнего диктора; $d(n)$ — шумы окружающей среды; $e(n)$ — акустическое эхо от громкоговорителя.

Целью работы системы подавления акустического эха и редактирования шумов является снижение уровня составляющих $b(n)$ и $e(n)$ до величины, не слышимой человеческим ухом.

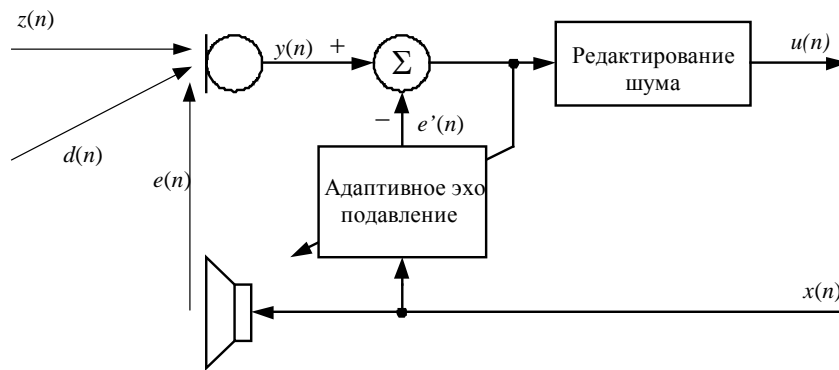


Рис. 1. Комбинированная система подавления акустического эха и редактирования шумов окружающей среды

Комбинированная система подавления акустического эха и редактирования шумов окружающей среды состоит из двух основных подсистем: подсистемы подавления эхо-сигнала и подсистемы редактирования шума. Для подавления эхо-сигнала на сегодняшний день наибольшее применение нашли системы на базе адаптивных фильтров, использующие нормализованный алгоритм метода наименьших квадратов (НМНК) [1]. НМНК фильтр работает как предсказатель эхо-сигнала. Для снижения уровня составляющей эхо-сигнала предсказанное эхо вычитается из входного сигнала.

В качестве подсистемы редактирования шумов наибольшее применение нашел метод спектрального вычитания [2]. Сущность его заключается в том, что для снижения уровня шума его спектр вычитается из спектра входного сигнала. У данного подхода есть существенные недостатки.

Во-первых, обработка сигнала с помощью НМНК адаптивного фильтра происходит во временной области по одному отсчету, в то время как снижение уровня шума осуществляется в частотной области блоками по несколько отсчетов. Таким образом, хотя система и называется комбинированной, ее можно разделить на две независимые подсистемы, требующие дополнительного согласования.

Во-вторых, НМНК фильтр показывает прекрасные результаты при невысоких порядках фильтра, но при увеличении порядка качество подавления эхо-сигнала значительно снижается [1].

В-третьих, НМНК фильтр обладает значительно более высокой вычислительной сложностью по сравнению с алгоритмами редактирования шума.

Эти недостатки можно устранить, используя вместо НМНК фильтра блочный МНК фильтр с обработкой сигнала в частотной области.

В данной работе предлагается одноканальная комбинированная система редактирования шумов окружающей среды и подавления эхо-сигнала на основе обработки в частотной области и нового психоакустического мотивированного правила спектрального взвешивания.

Комбинированная система подавление эха и редактирование шумов окружающей среды

Подавление эха и редактирование шумов окружающей среды в частотной области. Подсистема подавления эхо-сигнала работает как предсказатель эхо-сигнала, использующий функцию ошибки для корректировки параметров системы. Функция ошибки представляет собой разность эхо-сигнала и его предсказанного значения [1]. В качестве эхо-сигнала обычно используется входной сигнал во время речевых пауз. Однако при использовании схемы, показанной на рис. 1, во время речевых пауз входной сигнал представляет собой сумму эхо-сигнала и шумов окружающей среды. Поэтому после вычитания из входного сигнала предсказанного эха результат будет представлять собой сумму функции ошибки и шумов окружающей среды.

Шумовая составляющая данного сигнала несколько уменьшает скорость сходимости алгоритма. Но при условии, что уровень шума достаточно низок и шум не коррелирован с эхом, шумовой составляющей пренебрегают. Однако если уровень шумов окружающей среды достаточно высок, то уменьшение скорости сходимости алгоритма эхо-предсказания становится заметным. Для уменьшения влияния шумов окружающей среды на качество подавления акустического эха необходимо добавить в систему дополнительный модуль первичного подавления шумового сигнала. Так как система подавления шума вносит в сигнал некоторые искажения, то наиболее предпочтительной, является схема подключения модуля первичного редактирования шума, приведенная на рис. 2 [1].

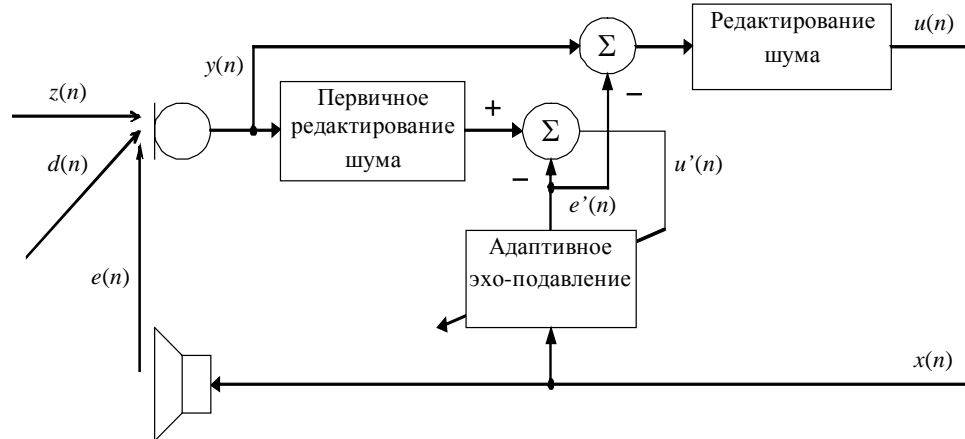


Рис. 2. Комбинированная система подавления акустического эха и редактирования шумов окружающей среды с модулем первичного редактирования шума

При такой схеме построения подсистема первичного подавления шумового сигнала не вносит дополнительных искажений в выходной сигнал и при этом позволяет избавиться от шумовой составляющей при оценке эхо-сигнала и, как следствие, улучшить качество подавления эхо-сигнала. При использовании схемы совместно с подсистемой эхо-подавления во временной области возникает проблема двойного преобразования данных из временной области в частотную и обратно. Необходимость выполнения преобразования дважды ведет к увеличению вычислительной сложности системы в целом. Эта причина и является основным сдерживающим фактором для использования данной схемы. Однако если комбинированная система подавления эха и редактирования шума реализуется полностью в частотной области, то проблема двойного преобразования данных исчезает.

При реализации адаптивного фильтра в частотной области наилучшим вариантом является использование для небольших порядков фильтра блочного адаптивного фильтра с обработкой сигнала в частотной области (БАФЧО) [3], а для высоких порядков – рассогласованного секционированного блочного адаптивного фильтра в частотной области (РСБАФЧО) [4]. Использование данных алгоритмов позволяет снизить вычислительную сложность системы, однако величина снижения не столь велика. Это обусловлено тем, что для выполнения линейной свертки в данных фильтрах требуется использовать дополнительные блоки преобразования Фурье во временную область и обратно.

Альтернативой блочным алгоритмам метода наименьших квадратов в частотной области является алгоритм подавления эхо-сигнала на основе метода спектрального вычитания. Основная идея алгоритма заключается в том, что эхо-сигнал рассматривается как дополнительный вид шумов и снижается одновременно с шумом по единой методике. Для этого сначала оценивается спектральная плотность мощности эхо-сигнала и находится соотношение сигнал/эхо. Зная данное соотношение, система взвешивает входной речевой сигнал.

Метод спектрального вычитания обладает меньшей вычислительной сложностью по сравнению с алгоритмами БАФЧО и РСБАФЧО. Это обусловлено следующими причинами:

1) метод спектрального вычитания работает с действительными числами (спектральная плотность мощности), а не с комплексными, как алгоритмы БАФЧО и РСБАФЧО;

2) при использовании метода спектрального вычитания не используются дополнительные блоки преобразования Фурье во временную область и обратно, которые необходимы алгоритмам БАФЧО и РСБАФЧО для обеспечения линейности свертки.

Главный недостаток метода спектрального вычитания заключается в появлении в выходном сигнале "музыкального" тона. Этот эффект воспринимается слушателем как присутствие в сигнале неестественных "металлических" ноток. Данный эффект проявляется так же и при редактировании шумов окружающей среды.

От эффекта "музыкального" тона позволяют избавиться алгоритмы спектрального вычитания, учитывающие психоакустические особенности восприятия человека [5,6]. Использование данных алгоритмов увеличивает вычислительную сложность системы подавления эхо-сигнала. Однако при комбинированном подавлении эхо сигнала и редактировании шума используется только один модуль спектрального вычитания. Это позволяет снизить вычислительную сложность системы в целом за счет совместного использования модуля расчета психоакустических характеристик сигнала.

Таким образом, использование алгоритма спектрального вычитания является предпочтительным по сравнению с алгоритмами БАФЧО и РСБАФЧО.

Психоакустически мотивированное правило спектрального взвешивания. При использовании блочных алгоритмов дискретное преобразование Фурье (ДПФ) блока входных данных, полученных с микрофона, может быть описано на основе (1) следующим уравнением:

$$\mathbf{Y}(k, f) = \mathbf{Z}(k, f) + \mathbf{D}(k, f) + \mathbf{E}(k, f), \quad (2)$$

где k — номер блока данных; f — частота; $\mathbf{Z}(k, f)$ — ДПФ полезного речевого сигнала $z(n)$; $\mathbf{D}(k, f)$ — ДПФ шумов окружающей среды $d(n)$; $\mathbf{E}(k, f)$ — ДПФ эхо-сигнала $e(n)$.

На основе формулы (2) может быть получено выражение для нахождения чистой речи во входном сигнале:

$$\mathbf{Z}(k, f) = \mathbf{Y}(k, f) - \mathbf{E}(k, f) - \mathbf{D}(k, f). \quad (3)$$

Однако использование формулы (3) сопряжено с некоторыми трудностями. Главной из них является то, что оценка $\mathbf{E}(k, f)$ и $\mathbf{D}(k, f)$ является достаточно сложной задачей.

Так как человек практически не различает искажения фазы сигнала, то вычитание можно выполнять только с амплитудами соответствующих спектров, оставляя фазу без изменения. В этом случае вычитание амплитуд спектров может быть заменено умножением на действительную взвешивающую функцию (см. рис. 3):

$$\mathbf{Z}(k, f) = \mathbf{Y}(k, f) \cdot \mathbf{G}(k, f), \quad (4)$$

где $\mathbf{G}(k, f)$ — взвешивающая функция.

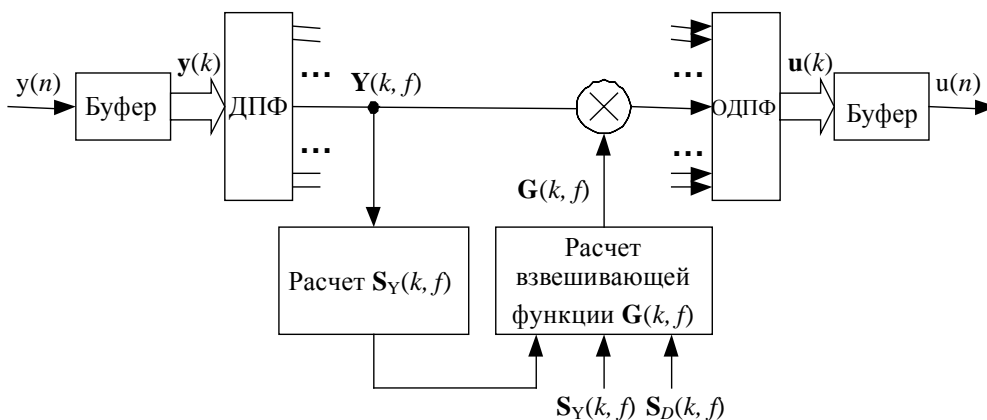


Рис. 3. Блок-схема спектрального взвешивания

Взвешивающая функция $\mathbf{G}(k)$ может быть найдена из следующих двух выражений для расчета спектральной плотности мощности (СПМ) сигнала $\mathbf{S}(k)$, полученных из (3) и (4) соответственно:

$$\mathbf{S}_z(k, f) = \mathbf{S}_y(k, f) - \mathbf{S}_E(k, f) - \mathbf{S}_D(k, f), \quad (5)$$

$$\mathbf{S}_z(k, f) = \mathbf{S}_y(k, f) \cdot \mathbf{G}^2(k, f), \quad (6)$$

где $\mathbf{S}_z(k, f)$ — СПМ сигнала $z(n)$; $\mathbf{S}_y(k, f)$ — СПМ сигнала $y(n)$; $\mathbf{S}_E(k, f)$ — СПМ эхо-сигнала удаленного диктора; $\mathbf{S}_D(k, f)$ — СПМ шумов окружающей среды.

Из выражений (5) и (6) следует что

$$\mathbf{G}(k, f) = \sqrt{1 - \frac{\mathbf{S}_E(k, f)}{\mathbf{S}_y(k, f)} - \frac{\mathbf{S}_D(k, f)}{\mathbf{S}_y(k, f)}}. \quad (7)$$

Так как $\mathbf{S}_y(k, f)/\mathbf{S}_D(k, f)$ равен квадрату отношения сигнал/шум, а $\mathbf{S}_y(k, f)/\mathbf{S}_E(k, f)$ — квадрату отношения сигнал/эхо, то выражение (7) может быть переписано в следующем виде:

$$\mathbf{G}(k, f) = \sqrt{1 - \frac{1}{\mathbf{SNR}^2(k, f)} - \frac{1}{\mathbf{SER}^2(k, f)}}, \quad (8)$$

где $\mathbf{SNR}(k, f)$ — отношение сигнал/шум; $\mathbf{SER}(k, f)$ — отношение сигнал/эхо.

При использовании функции (8) в выходном сигнале появляется эффект "музыкального" тона, который проявляется в виде неприятных на слух, неестественных, "металлических" ноток. Увеличение значения взвешивающей функции $\mathbf{G}(k, f)$ позволяет уменьшить эффект "музыкального" тона, однако при этом снижается и качество подавления шума. Наилучшим выходом в этой ситуации является избирательное увеличение $\mathbf{G}(k, f)$. Для этого может быть использован психоакустический эффект частотного маскирования [7], показанный на рис. 4.

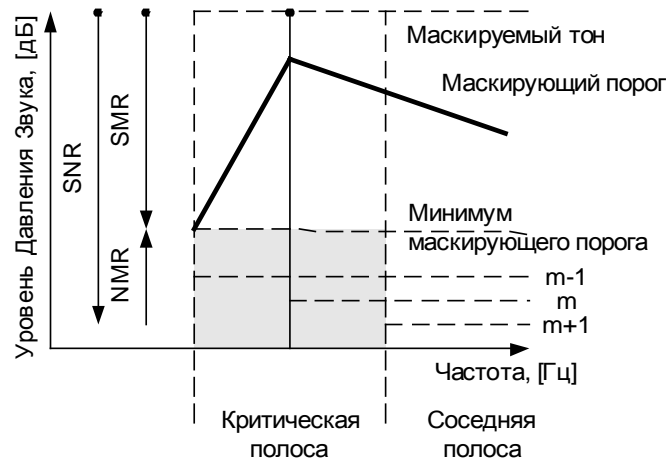


Рис. 4. Эффект маскирования

Существует два варианта частотного маскирования: тональное и шумовое. Эффект тонального маскирования заключается в следующем. Если в частотной характеристике сигнала присутствует некоторая составляющая с высоким значением мощности, то она делает для человека неслышимыми соседние с ней составляющие с меньшим значением мощности. Для того чтобы некоторый тон был замаскирован, необходимо, чтобы его уровень был ниже определенного порога, который зависит от уровня маскирующего тона и разности частот между маскирующим и маскируемым тоном. Эффект шумового маскирования похож на тональное

маскирование за исключением того, что в качестве маскирующего тона используется сумма всех частотных составляющих сигнала в пределах критической полосы. Данный маскирующий тон располагается по середине критической полосы. Совокупность всех тональных и шумовых порогов маскирования образует абсолютный порог маскирования.

Если некоторый тон по уровню ниже абсолютного порога маскирования, то человеческое ухо не слышит данную частотную составляющую, а значит, для данной частоты значение $\mathbf{G}(k, f)$ можно принять равным 1. Для остальных частот $\mathbf{G}(k, f)$ надо выбирать так, чтобы уровень шума оказался ниже абсолютного порога маскирования. В этом случае выходной сигнал будет равен

$$\mathbf{S}_U(k, f) = \mathbf{S}_Z(k, f) + \mathbf{T}_M(k, f). \quad (9)$$

С учетом формулы (4) выражение для расчета СПМ выходного сигнала будет иметь следующий вид:

$$\mathbf{S}_U(k, f) = \mathbf{S}_Y(k, f) - \mathbf{S}_D(k, f) - \mathbf{S}_E(k, f) + \mathbf{T}_M(k, f). \quad (10)$$

На основе формул (6) и (10) может быть получено выражение для расчета взвешивающей функции

$$\mathbf{G}(k, f) = \sqrt{1 - \frac{\mathbf{S}_D(k, f)}{\mathbf{S}_Y(k, f)} - \frac{\mathbf{S}_E(k, f)}{\mathbf{S}_Y(k, f)} + \frac{\mathbf{T}_M(k, f)}{\mathbf{S}_Y(k, f)}}. \quad (11)$$

или

$$\mathbf{G}(k, f) = \sqrt{1 - \frac{1}{\text{SNR}^2(k, f)} - \frac{1}{\text{SER}^2(k, f)} + \frac{1}{\text{SMR}^2(k, f)}}. \quad (12)$$

где $\text{SMR}(k, f)$ — отношение сигнал / абсолютный порог маскирования.

Блок-схема комбинированной системы подавления эхо-сигнала и редактирования шумов, учитывающей психоакустические особенности восприятия человека, показана на рис. 5.

Метод оценки спектральной плотности мощности эхо-сигнала. Для оценки отношения сигнал / эхо необходимо знать оценку СПМ эхо-сигнала.

Эхо сигнал представляет сумму сигналов от громкоговорителя, пришедших с различной временной задержкой и ослаблением. Таким образом, в общем случае эхо сигнал может быть представлен в виде следующей формулы:

$$e(n) = \sum_{i=0}^n h(i) \cdot x(n-i), \quad (13)$$

где \mathbf{h} — импульсная характеристика помещения, i -й элемент которой показывает уровень затухания сигнала от источника эхо до микрофона, пришедшего с задержкой $n-i$.

Формула (13) учитывает все возможные составляющие эхо-сигнала с момента начала измерений. Однако на практике через определенный промежуток времени N за счет поглощения звука $i-N$ составляющие эхо-сигнала становятся настолько малы, что ими можно пренебречь. Таким образом, можно считать, что импульсная характеристика помещения конечна и формула для оценки эхо-сигнала будет иметь следующий вид:

$$e(n) = \sum_{i=0}^N h(i) \cdot x(n-i), \quad (14)$$

где N — длительность эхо-сигнала.

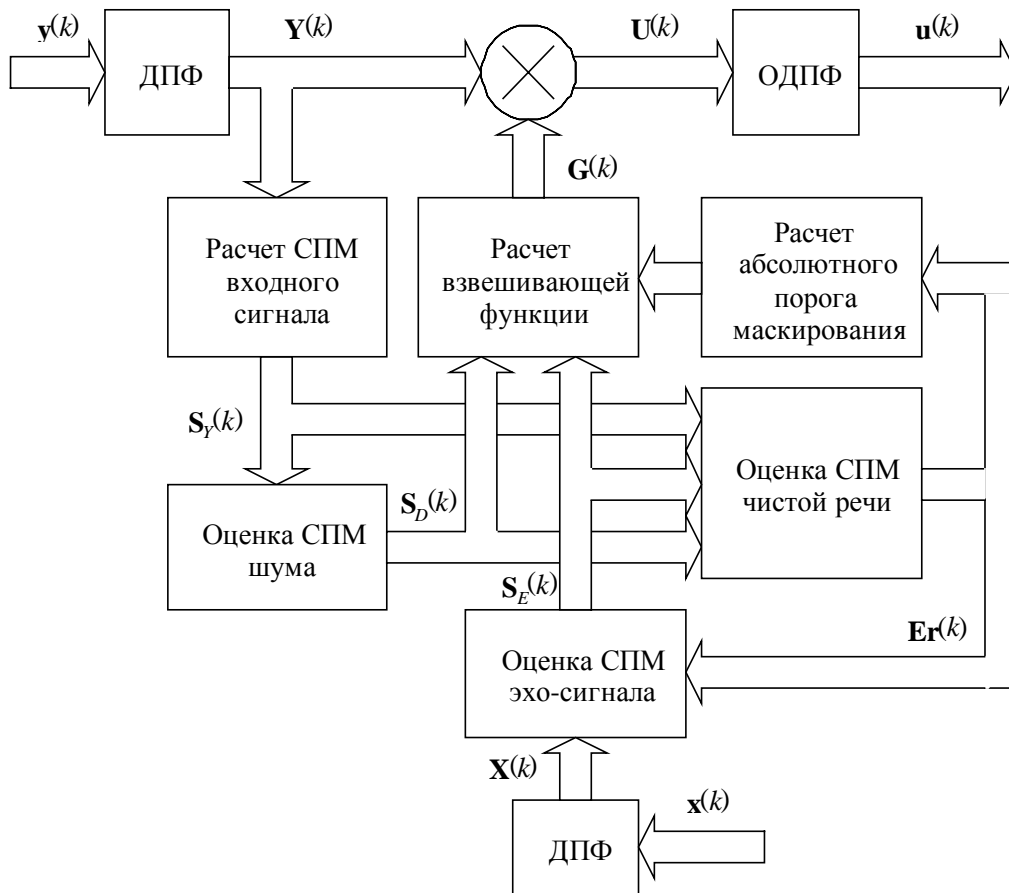


Рис. 5. Комбинированная система подавления эхо-сигнала и редактирования шумов, учитывающая психоакустические особенности восприятия человека

При обработке входного сигнала блоками по B отсчетов длина импульсной характеристики помещения устанавливается кратной размеру входного блока (если это условие не выполняется, то необходимо увеличить длину импульсной характеристики). В результате группировки слагаемых по блокам формула (14) примет следующий вид

$$e(n) = \sum_{j=0}^{Q-1} \sum_{i=0}^{B-1} h(jB+i) x(n-jB-i), \quad (15)$$

где Q — количество блоков данных, притом $N = B \cdot Q$.

Выражение $\sum_{i=0}^{B-1} h(jB+i) x(n-jB-i)$ в формуле (15) представляет собой свертку, которая может быть вычислена с помощью дискретного преобразования Фурье (ДПФ):

$$\sum_{i=0}^{B-1} h(jB+i) x(n-jB-i) = \mathbf{H}(jB) \mathbf{X}(n-jB), \quad (16)$$

где $\mathbf{H}(jB)$ — ДПФ вектора $[h(jB) \dots h((j+1)B-1)]$, $\mathbf{X}(n-jB)$ — ДПФ вектора $[x(n-jB) \dots x(n-(j+1)B-1)]$.

Используя свойство, что ДПФ последовательности $x_1(n) + x_2(n)$ равно $\mathbf{X}_1(n) + \mathbf{X}_2(n)$, можно записать:

$$\mathbf{E}(n, f) = \sum_{j=0}^{Q-1} \mathbf{H}(jB, f) \mathbf{X}(n-jB, f), \quad (17)$$

где $\mathbf{E}(n, f)$ — ДПФ сигнала $e(n)$.

Так как обработка сигнала проводится блоками по B отсчетов, то в выражении (17) можно заменить номер отсчета n на номер блока k :

$$\mathbf{E}(k, f) = \sum_{j=0}^{Q-1} \mathbf{H}(j, f) \mathbf{X}(k-j, f). \quad (18)$$

Обозначив \mathbf{S}_H через \mathbf{W} на основе формулы (18) можно получить выражение для расчета спектральной плотности мощности эхо-сигнала:

$$\mathbf{S}_E(k, f) = \sum_{j=0}^{Q-1} \mathbf{W}(j, f) \mathbf{X}(k-j, f).$$

При использовании формулы (18) для оценки СПМ эхо-сигнала необходимо знать характеристику помещения \mathbf{W} . Данный параметр можно оценить, если будет известна спектральная плотность мощности эха. В моменты времени, когда ближний диктор молчит и говорит удаленный диктор, сигнал на выходе модуля первичного подавления шума будет представлять собой сумму акустического эха и остаточного шума. Так как уровень остаточного шума значительно ниже уровня эха и, как правило, не коррелирован с ним, то данной составляющей можно пренебречь и рассматривать сигнал на выходе данного модуля как оценку чистого эха.

Для оценки величины \mathbf{W} на основе оценки чистого эха могут быть использованы различные адаптивные алгоритмы, большинство из которых являются модификациями стандартных итерационных процедур регулирования на основе решения проблемы минимизации в режиме реального времени. В качестве критерия оптимизации в данных алгоритмах наибольшее использование нашла Средняя Квадратичная Ошибка. Так как Нормализованный алгоритм Метода Наименьших Квадратов (НМНК) обладает наибольшей скоростью сходимости, то при построении системы подавления эхо сигнала использован именно этот метод.

Для корректировки \mathbf{W} необходимо сначала рассчитать величину ошибки оценки спектральной плотности мощности $\mathbf{Er}(k)$:

$$\mathbf{Er}(k, f) = \mathbf{S}_{D'}(k, f) - \sum_{j=0}^{Q-1} \mathbf{S}_X(k-j, f) \mathbf{W}(j, f),$$

где $\mathbf{S}_{D'}(n, f)$ — СПМ сигнал на выходе модуля первичного подавления шума.

В начальный момент времени матрица \mathbf{W} содержит нулевые значения. При обработке каждого блока входного сигнала, если говорит только удаленный диктор, матрица \mathbf{W} корректируется по формуле

$$\mathbf{W}(i, f) = \max \left(\mathbf{W}(i, f) + \frac{\mu \mathbf{Er}(k, f) \mathbf{S}_X(k-i, f)}{\|\mathbf{S}_X(k-i)\|^2}, 0 \right),$$

где μ — коэффициент сходимости алгоритма в диапазоне $[0;1]$; $\|\mathbf{S}_X\|^2$ — квадрат евклидовой нормы.

Оценка спектральной плотности мощности шумового сигнала

Для оценки спектральной плотности мощности шумов окружающей среды используется метод определения СПМ шума во время речевых пауз. Метод заключается в том, что в качестве СПМ шума берут среднее значение входного сигнала, измеренного во время речевых пауз. Среднее значение мощности входного сигнала во время речевых пауз \mathbf{S}_{av} находится с помощью экспоненциального усреднения

$$S_{av}(k, f) = \begin{cases} \alpha \cdot S_Y(k, f) + (1 - \alpha) \cdot S_{av}(k - 1, f), & sl = 1, \\ S_{av}(k - 1, f), & sl = 0, \end{cases}$$

где $S_{av}(k, f)$ — новое значение среднего СПМ входного сигнала; $S_{av}(k-1, f)$ — предыдущее значение S_{av} ; sl — управляющий сигнал от детектора тишины, который принимает значение 1 если отсутствует речевая активность и 0 – в противоположном случае.

Так как СПМ шума S_D не может быть больше, чем СПМ входного сигнала, то S_D находится на основе S_{av} по формуле

$$S_D(k, f) = \min(S_{av}(k, f), S_Y(k, f)).$$

Детектор тишины и двойного разговора

Для обеспечения качественного подавления эха процесс адаптации подсистемы подавления эха в моменты, когда отсутствует речевой сигнал удаленного диктора, и в периоды, когда одновременно говорят и удаленный и ближний дикторы, необходимо останавливать. Для остановки процесса адаптации в данные временные отрезки в систему подавления эхо-сигнала необходимо включать детекторы одиночного и двойного разговора.

Для определения соответствующих периодов мощность сигнала сравнивается с некоторой пороговой величиной. В случае, если сумма спектральных мощностей сигнала по всем частотам не превышает некоторую пороговую величину, то принимается решение об отсутствии речевой активности. Порог берется пропорционально мощности шума.

Для определения периодов двойного разговора анализируется спектральная мощность сигнала удаленного диктора и сигнала на выходе системы подавления эха и шума. Для этого мощности входных сигналов сравниваются с соответствующими пороговыми величинами. В случае, если мощность сигнала удаленного диктора и мощность сигнала на выходе системы превышает пороговые величины, то принимается решение о наличии двойного разговора. Пороговая величина для сигнала удаленного диктора берется пропорциональной мощности шума данного сигнала, а порог для выходного сигнала берется пропорциональной мощности шума и остаточного эха выходного сигнала.

Определение периодов двойного разговора осуществляется в два этапа.

На первом этапе определяется присутствие речевой активности во входном сигнале и в сигнале удаленного диктора. Для этого мощности входного сигнала и мощности сигнала удаленного диктора сравниваются с порогами. Исследование показало, что порог лучше всего принимать пропорциональным уровню шумов окружающей среды (для определения ближнего диктора) и уровню шумов канала связи (для удаленного диктора). Моделирование показало, что оптимальным является значение коэффициента пропорциональности, равное 1,5–3.

На втором этапе сравниваются между собой мощности входного сигнала и мощности сигнала удаленного диктора. Если мощность входного сигнала больше мощности сигнала удаленного диктора с учетом шумов окружающей среды, то принимается решение о наличии двойного разговора. Моделирование показало, что для исключения ложных решений мощность входного сигнала должна быть не менее чем в 1,5 больше мощности сигнала удаленного диктора.

Результаты моделирования

Результаты моделирования работы комбинированной системы подавления эхо-сигнала и редактирования шумов окружающей среды на основе метода спектрального вычитания учитывающей психоакустические особенности слуха человека, показаны на рис. 6. При моделировании использовался диалог двух человек. Длительность акустического эха составляла 0,1 с, уровень шумов окружающей среды — 10 дБ.

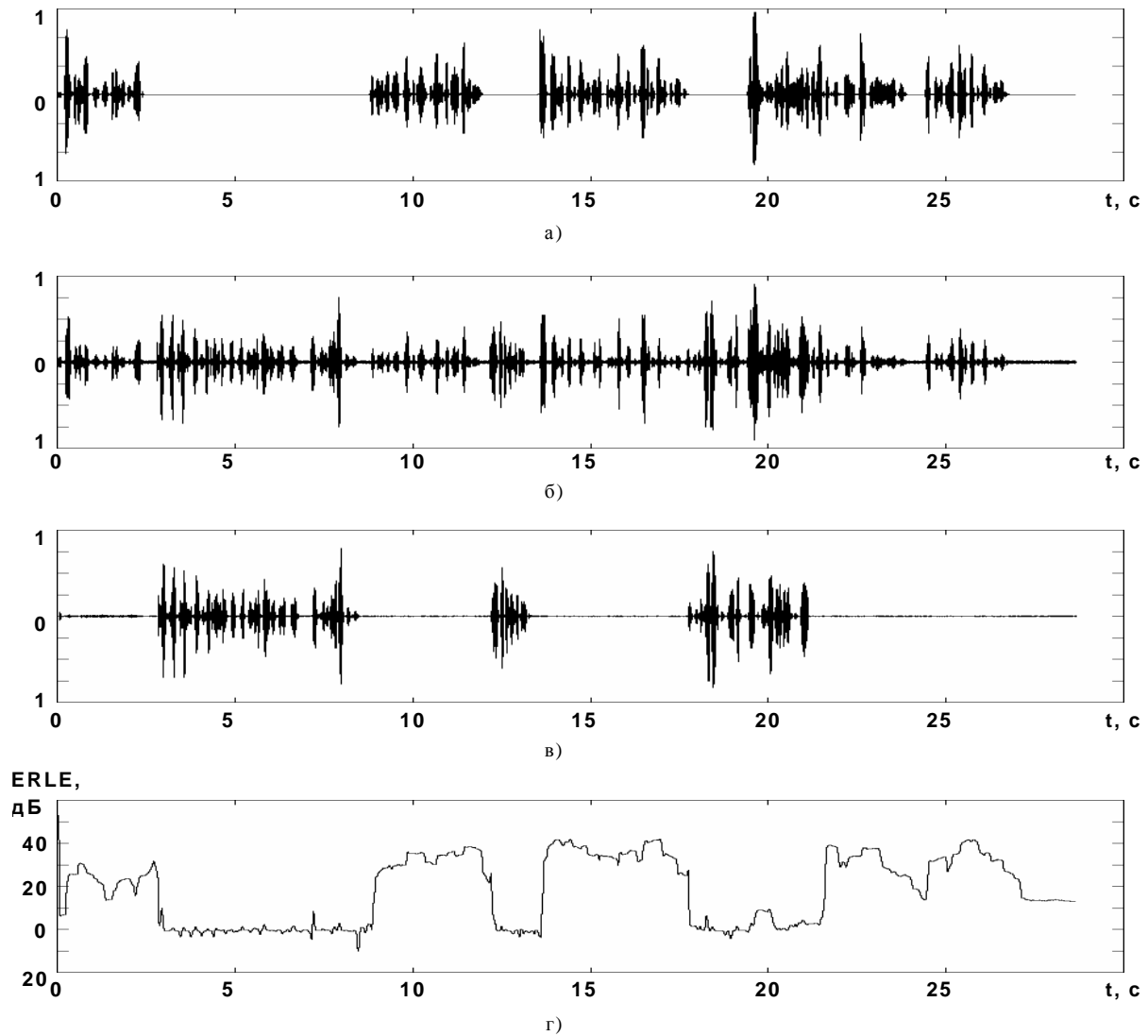


Рис. 6. Результаты моделирования: а — сигнал удаленного диктора; б — входной сигнал; в — выходной сигнал; г — уровень ослабления эхо-сигнала

Для оценки качества подавления эхо-сигнала использовался уровень ослабления данного сигнала, вычисляемый по формуле

$$ERLE = -10 \log_{10} \frac{E\{e^2(k)\}}{E\{s^2(k)\}}.$$

Моделирование показало, что предлагаемая методика подавления эха и редактирования шума позволяет получить ослабление эха $ERLE = 35\text{--}40$ дБ и ослабление шума — $30\text{--}35$ дБ. Эффект "музыкального" тона не наблюдался. Данные результаты находятся на уровне лучших мировых разработок, при этом выигрыш в вычислительной сложности составляет $10\text{--}20$ раз в зависимости от порядка фильтра.

THE PERCEPTUAL METHOD OF QUALITY IMPROVING OF A FREQUENCY LIMITED SPEECH SIGNAL: THE COMBINED SINGLE-CHANNEL SYSTEM, THE FREQUENCY APPROACH

A.E. ANOSHENKO, A.A. PETROVSKY

Abstract

In this paper the combined single-channel system of environment noise reducing and an echo-signal cancellation on the basis of processing in frequency domain and new psychoacoustical motivated rule of spectral weighing is offered. The perceptual approach of speech signal quality improving of the given system is characterized in comparison with known on the order by smaller computing complexity and uniform editing rule of spectral weighing for noise and echo-signal.

Литература

1. *Haykin S.* Adaptive filter theory — Third edition. Prentice-Hall, 1996.
2. *Berouti M., Schwartz R., Makhoul J.* // Proc. ICASSP, Washington, DC., Apr. 1979. P. 208–211.
3. *Ferrara E.R.* // IEEE transaction. 1980. Vol. ASSP-28, No. 4. P. 474–475.
4. *Egelmeers G.P.M.* Decoupling of partition factors in Partitioned Block FDAF. ECCTD, Davos, Switzerland, August, 1993. P. 323–329.
5. *Virag N.* // IEEE transaction on speech and audio processing. 1999. Vol. 5, No. 2. P. 126–137.
6. *Gustafsson S., Martin R., Jax P., Vary P.* // IEEE transaction on speech and audio processing. 2002. Vol. 10, No. 5. P 245–256.
7. *Zwicker E., Fastl H.* Psychoacoustics: Facts and Models. New York: Springer, 1990.