ШУМОПОНИЖАЮЩЕЕ УСТРОЙСТВО ДЛЯ ВОКОДЕРА

Бабкин В.В.
Центр ЦОС СПб ГУТ, Санкт-Петербург, пр. Большевиков д.22 корп.1.
vb@dsp-sut.spb.ru, (812)589-51-85, www.dsp.sut.ru

Окружающий акустический шум является мощным мешающим фактором, снижающим качество работы систем цифровой речевой связи. При этом падает как разборчивость речи, так и ее качество, выражаемое в терминах естественности звучания, узнаваемости голоса и т. д. Помимо эффекта маскирования речи шумом, одной из основных причин такого снижения является сильный рост искажений при прохождении зашумленной речью преобразования в устройстве низкоскоростной компрессии речи (вокодере). В частности, для низкоскоростных (0.6–4 кбит/с) параметрических вокодеров, снижение качества становится заметным для отношений сигнал/шум (ОСШ) на входе менее +15…+20 дБ, вследствие возрастания числа ошибок в оценке параметров модели синтеза речевого сигнала на основе анализа текущей речи. Нижний предел для ОСШ, при котором речь становиться неразборчивой, лежит около 0…+3дБ, в зависимости от типа вокодера, вида шума и способа измерения ОСШ.

Так как частотные спектры акустического шума и речи перекрываются, набор линейных методов фильтрации, приводящих к повышению ОСШ для речи на входе вокодера, ограничен. К ним можно отнести пространственную фильтрацию с помощью направленных микрофонных систем и предварительную частотную фильтрацию входного сигнала, формирующую АЧХ для узкополосных (0.3-3.4 кГц) систем связи, например по рек. ITU-T P.48. Первые методы имеют самостоятельное значение, они лишь отодвигают наступление порога понижения ОСШ на входе вокодера в конкретной ситуации, но на сам порог не влияют. Вторые методы не эффективны, так как имеют постоянные среднестатистические параметры фильтрации, не учитывающие динамики спектральных характеристик конкретных шумовых и речевых сигналов.

В условиях квазистационарных аддитивных статистически независимых от речи акустических шумов для улучшения ОСШ речевых сигналов широко применяются шумопонижающие устройства (ШПУ), построенные на основе нелинейных адаптивных алгоритмов очистки речи от шумов [1,2]. Чаще всего они строятся с использованием различных кратковременных преобразований сигнала, оценки компонент шума и речи в преобразованной области, подавления компонент шума с последующим обратным преобразованием очищенного сигнала во временную область. Обработка ведется по кадрам длительностью 10-30 мс. Используются следующие преобразования: дискретное преобразование Фурье (ДПФ), дискретное косинусное преобразование, вейвлет преобразование, преобразование Карунена-Лоева (Karhunen-Loeve) и др.

Наиболее распространены алгоритмы на основе ДПФ с обработкой кратковременного спектра амплитуд сигнала в частотной области следующими методами: спектрального вычитания [3]; Винеровской фильтрации [1]; статистических оценок амплитуд речевого сигнала по критерию максимального правдоподобия [4] или минимума среднеквадратической ошибки [5,6]. В настоящее время, наиболее развит метод статистических оценок значений логарифма спектра амплитуд речевого сигнала по критерию минимума среднеквадратической ошибки (MMSE-LSA) [6], дополненный статистической оценкой вероятности присутствия речи в шуме [8,9].

Считается, что для людей с нормальным слухом адаптивные алгоритмы очистки речи от шумов на основе методов спектрального вычитания увеличивают субъективное качество очищенной речи, повышают ее ОСШ, снижают утомляемость при длительном прослушивании, но не способны повысить ее разборчивость по сравнению с разборчивостью исходной речи в шумах.

Однако, применение ШПУ совместно с низкоскоростными речепреобразующими устройствами (РПУ) с параметрическим способом преобразования речи, способно повысить и качество и разборчивость синтетической речи на выходе РПУ при работе в шумах по сравнению с использованием РПУ без ШПУ, так как при этом улучшается точность оценки речевых параметров в анализаторе вокодера [3, 9].

Основными проблемами построения ШПУ являются: оценка сглаженных спектральных характеристик шумового сигнала на основе анализа смеси речи и шума, оценка мгновенных значений спектра амплитуд речевого сигнала, поиск алгоритмов адаптации характеристик взвешивающего фильтра, не приводящих к возникновению артефактов звучания («музыкальных» шумов), поиск компромисса между степенью подавления шума и степенью искажения речи.

Существующими методиками субъективного и объективного тестирования характеристик ШПУ раздельно от РПУ являются рекомендации 3GPP TS 26.077, ITU-T P.835 и разрабатываемая в настоящее время рекомендация ITU-T G.VED.

Если рассматривать ШПУ совместно с РПУ как единую систему, то можно опираться на хорошо зарекомендовавшие себя методы тестирования качества работы низкоскоростных РПУ такие как оценка слоговой разборчивости, оценка качества речи по PESQ-MOS ITU-T P.862 [10] и др. Субъективные методы оценки слоговой разборчивости речи точны, но очень сложны в применении, требуют привлечения большого числа экспертов-аудиторов и длительного времени испытаний. Объективные методы, напротив, легко применимы, позволяют количественно оценивать параметры качества речи в автоматическом режиме, облегчают процесс поиска путей улучшения алгоритма в процессе его разработки, но их применение целесообразно только тогда, когда они дают оценки, близкие к субъективным оценкам. При разработке ШПУ использовались объективные оценки качества речи по критерию PESQ-MOS и субъективные оценки методом предпочтений при сравнительном прослушивании.

Разработанное ШПУ предназначено для совместного использования с помехоустойчивым РПУ RMELP 4400 бит/с, построенном на принципах совместной оптимизации схем речевого и канального кодирования [11,12]. Целью объединения ШПУ и РПУ являлось сохранение работоспособности системы цифровой речевой связи в экстремальных условиях эксплуатации – как в условиях сильных акустических шумов, так и при наличии большого числа битовых ошибок в цифровом канале связи, достигающих 7%, что характерно при связи между машинами в условиях городской застройки по радиоканалу.

Разработанное ШПУ построено на основе метода MMSE-LSA с модификацией усиления на основе оценки вероятности наличия речи. Блок схема алгоритма представлена на рис.1.

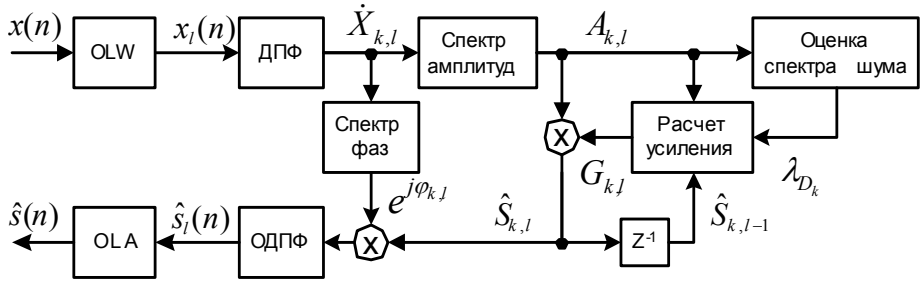

Рис.1. Блок схема алгоритма ШПУ.

Входной сигнал $x(n) = s(n) + d(n)$, состоящий из смеси речи $s(n)$ и шума $d(n)$, разбивается на кадры $x_l(n)$, где l – номер кадра, длительностью 32 мс в блоке взвешивания с перекрытием (OLW). Для каждого кадра, используя ДПФ, находится комплексный частотный спектр входного сигнала $\dot{X}_{k,l}$, где k – номер частотной полосы, а также спектр фаз $e^{j\varphi_{k,l}}$ и спектр амплитуд $A_{k,l}$.

Условие квазистационарности шума предполагает, что его спектральные характеристики изменяются значительно медленнее спектральных характеристик речи. Для оценки среднего значения амплитудного спектра шума $D_{k,l}$ используется рекуррентное экспоненциальное усреднение спектра амплитуд входного сигнала $A_{k,l}$ по предыдущим кадрам с постоянной времени 1–2 с. Данная оценка хорошо усредняет быстрые изменения спектра речевого сигнала и следит за медленными изменениями среднего спектра шума, однако, она является смещенной в сторону завышения, из-за влияния речи, поэтому процесс усреднения управляется с помощью детектора речевой активности с мягким решением. Оценка энергии (дисперсии) спектральных компонент шума строится как $\lambda_{D_k} = {D_{k,l}}^2$.

Для получения оценки $\ln \hat{S}_{k,l}$ мгновенного значения логарифма спектра амплитуд речевого сигнала $\ln S_{k,l}$, являющегося неизвестной случайной величиной (с.в.), используется минимизация среднеквадратичной ошибки (СКО) $E\left\{\left(\ln S_{k,l} - \ln \hat{S}_{k,l}\right)^2\right\}$. Спектр амплитуд $A_{k,l}$ входного сигнала, состоящего из смеси речи и шума, является наблюдаемой с.в. Статистическая оценка для $\ln \hat{S}_{k,l}$, дающая минимум СКО равна ее условному математическому ожиданию: $\ln \hat{S}_{k,l} = E\left\{(\ln S_{k,l}) \mid A_{k,l}\right\}$.

Удобно построить взвешивающий фильтр с коэффициентами усиления $G_{k,l}$, так чтобы оценка $\hat{S}_{k,l}$ выражалась непосредственно как $\hat{S}_{k,l} = A_{k,l} \cdot G_{k,l}$. В [6] показано, что если комплексные спектры речи и шума моделировать в виде двумерных с.в. с нормальным распределением, то фильтр, решающий задачу минимизации СКО для оценки $\ln \hat{S}_{k,l}$ должен иметь вид:

$$G_{k,l}(\gamma_{k,l}, \xi_{k,l}) = \frac{\xi_{k,l}}{1+\xi_{k,l}} \exp\left\{ \frac{1}{2} \int_{\upsilon_{k,l}}^{\infty} \frac{e^{-t}}{t} dt \right\},\ \upsilon_{k,l} \triangleq \frac{\xi_{k,l}\gamma_{k,l}}{1+\xi_{k,l}},\ \gamma_{k,l} \triangleq \frac{A_{k,l}^2}{\lambda_{D_{k,l}}} - \text{«апостериорное» ОСШ},\ \xi_{k,l} \triangleq \frac{\lambda_{S_{k,l}}}{\lambda_{D_{k,l}}} -$$

«априорное» ОСШ, $\lambda_{S_{k,l}}$ – дисперсия компонент речи, $\lambda_{D_{k,l}}$ – дисперсия компонент шума (для которой существует оценка). Так как дисперсия компонент речи не известна, оценка для $\hat{\xi}_{k,l}$ строится исходя из совмещения двух подходов: оценки, основанной на решении для прошлого кадра, и оценки, полученной на основе метода вычитания спектра мощности для текущего кадра:

$$\hat{\xi}_{k,l} = \alpha \frac{\hat{S}_{k,l-1}^2}{\lambda_{D_{k,l-1}}} + (1-\alpha) Max\left\{ \gamma_{k,l} - 1,\ 0 \right\} = \alpha G_{k,l-1}^2 \gamma_{k,l-1} + (1-\alpha) Max\left\{ \gamma_{k,l} - 1,\ 0 \right\},\ \ \alpha = 0.98$$

Основным преимуществом метода MMSE-LSA [6] по сравнению с классическим подходом спектрального вычитания [3] является почти полное отсутствие явления “музыкального” шума. При детальном анализе работы алгоритма, выполненном в [7], отмечается, что это достигнуто преимущественно за счет использования указанного правила оценки «априорного» ОСШ для управления АЧХ фильтра $G_{k,l}$. В методе спектрального вычитания усиление $G_{k,l}(\gamma_{k,l})$ является функцией только «апостериорного» ОСШ $\gamma_{k,l}$, поэтому АЧХ фильтра резко изменяется от кадра к кадру, порождая «музыкальный» шум. Для MMSE-LSA подхода усиление $G_{k,l}(\gamma_{k,l}, \xi_{k,l})$ является двупараметрическим и зависит, в основном, от $\hat{\xi}_{k,l}$, которое более плавно отслеживает изменения амплитудного спектра речи, поэтому «музыкальный» шум отсутствует или мало заметен.

Модификатор усиления, учитывающий вероятность присутствия речи, имеет вид: $G_{k,l}^{mm} = \frac{\Lambda_{k,l}\left(A_{k,l}\right)}{\Lambda_{k,l}\left(A_{k,l}\right)+1}$, где $\hat{\Lambda}_{k,l} = \frac{1-q_k}{q_k} \frac{\exp\left(\nu_{k,l}\right)}{1+\xi_{k,l}}$ – обобщенное отношение правдоподобия двух взаимоисключающих гипотез о наличии и об отсутствии речи во входном сигнале $A_{k,l}$, $q_k = 0.2$ – априорная вероятность гипотезы отсутствия речи. Таким образом, оценка спектра амплитуд речевого сигнала строится как $\hat{S}_{k,l} = A_{k,l} \cdot G_{k,l} \cdot G_{k,l}^{mm}$. Далее, используя спектр фаз входного сигнала, строится оценка комплексного спектра очищенного от шума речевого сигнала для текущего кадра, которая с помощью обратного ДПФ преобразуется во временную область $\hat{s}_l(n)$ и путем наложения с перекрытием (OLA) формируется выходной сигнал $\hat{s}(n)$.

Для целей тестирования алгоритм ШПУ реализован на языке Си для ПЭВМ в виде программной модели, использующей арифметику с плавающей точкой, с файловым вводом-выводом сигналов. При реализации ШПУ на ЦПОС TMS320VC5510 в арифметике с фиксированной точкой потребуется производительность не более 20 MIPS и память данных и программ не более 10 К 16-ти разрядных слов.

Схема испытаний ШПУ совместно с РПУ представлена на рис. 2. Результаты испытаний в акустических шумах различного типа в виде графиков зависимости величины PESQ-MOS от ОСШ входного сигнала, представлены на рис.3. Демонстрационные файлы можно послушать на web странице Центра ЦОС СПб ГУТ.

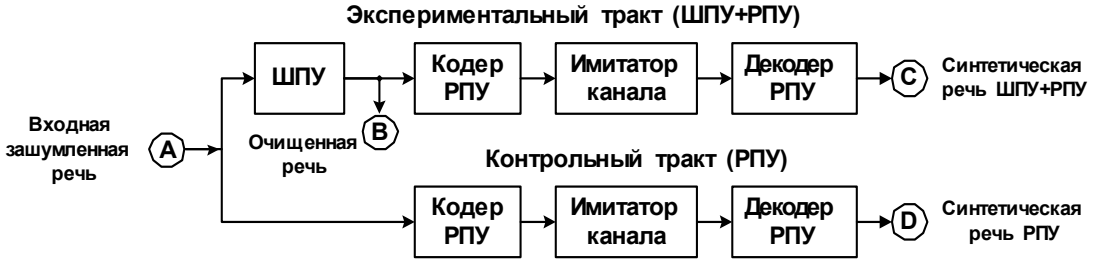

Рис. 2. Схема испытаний ШПУ совместно с РПУ.

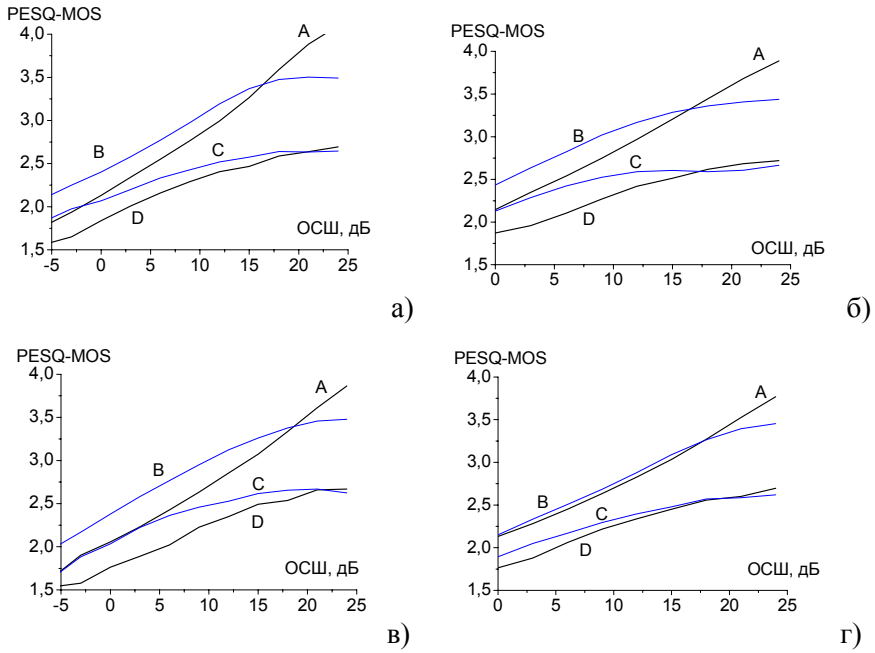

Рис. 3. Результаты испытаний ШПУ в шумах: а) белый, б) автомобиля, в) улицы, г) толпы.

Результаты показывают, что для входных сигналов с ОСШ в диапазоне 0…+20 дБ использование ШПУ дает хорошо заметный на слух положительный эффект. В зависимости от вида шума и величины входного ОСШ, выигрыш в качестве речи по шкале PESQ-MOS достигает 6–8 дБ. Максимальный выигрыш наблюдается для белого шума, шума внутри автомобиля и шума улицы. Минимальный – для шума толпы и шума офиса, для которых нарушено основное условие квазистационарности частотных спектров.

Таким образом, совместная оптимизация алгоритмов шумоподавления, речевого и канального кодирования, объединенных в один помехоустойчивый вокодер позволяет одновременно эффективно бороться с помехами двух видов – с акустическими шумами и с ошибками, возникающими в цифровых каналах связи. Это значительно повышает надежность работы систем цифровой речевой связи, организуемых по КВ и УКВ радиоканалам, в реальных условиях эксплуатации.

Литература

1. Speech Enhancement (Signals and Communication Technology). Editors: Benesty J., Makino S., Chen J. Springer, 2005, 406 pages.
2. Vary P., Martin R. Digital Speech Transmission: enhancement, coding and error concealment. Wiley & Sons, 2006.
3. Boll, S. Suppression of acoustic noise in speech using spectral subtraction. IEEE Transactions on Acoustics, Speech, and Signal Processing, Volume 27, Issue 2, Apr. 1979, pp. 113 – 120.
4. McAulay, R., Malpass, M. Speech enhancement using a soft-decision noise suppression filter. IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. 28, Issue 2, Apr. 1980, pp. 137–145.
5. Ephraim Y., and Malah D. Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator. IEEE Trans. ASSP-32, No. 6, pp. 1109–1121, December 1984.
6. Ephraim Y. and Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator, IEEE Trans. ASSP-33, No. 2, pp. 443-445, 1985.
7. Cappe O. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor, IEEE Trans. Speech and Audio Processing, Vol. 2, No. 1, pp. 345-349, April 1994.
8. Cohen I. On speech enhancement under signal presence uncertainty. ICASSP-2001, pp.167-170.
9. Martin R., D. Malah, R.V. Cox, A. J. Accardi. A Noise Reduction Preprocessor for Mobile Voice Communication. EURASIP Journal of Applied Signal Processing, 2004, № 8, pp. 1046-1058.
10. ITU-T Recommendation P.862. Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and codecs. 2001.
11. Бабкин В. В., Ланнэ А.А., Шаптала В.С. Оптимизационная задача выбора речевого и канального кодирования. Отчеты DSPA-2005, стр. 123–127, Москва 16-18 марта 2005 г.
12. Бабкин В. В., Ланнэ А. А., Шаптала В. С. Помехоустойчивые вокодеры для систем цифровой радиосвязи в КВ и УКВ диапазонах. Отчеты 1-ой межд. НПК "Исследование, разработка и применение высоких технологий в промышленности", стр. 21-22, СПб, 2005 г.