

УДК 621.391.83: 004.934.2

А.Н. Продеус, В.С. Дидковский

Национальный технический университет Украины «КПИ», Киев

ОЦЕНИВАНИЕ АПРИОРНОГО ОТНОШЕНИЯ СИГНАЛ-ШУМ В АЛГОРИТМАХ ШУМОПОДАВЛЕНИЯ

Сопоставлены три метода оценивания априорного отношения сигнал-шум (SNR) в алгоритмах шумоподавления. Показано, что оценка методом «управление решением» позволяет обеспечить максимально высокое качество звучания речевого сигнала. «Грубая» оценка оказалась наилучшей для применения в системах автоматического распознавания речи при интегральном SNR > 15 дБ. Оценка максимального правдоподобия занимает промежуточное место между оценками методом «управление решением» и «грубой» оценкой.

Ключевые слова: шумовая помеха; алгоритм шумоподавления; априорное отношение сигнал-шум; показатель качества; речевой сигнал.

Введение

Алгоритмы шумоподавления широко применяются в системах связи и в устройствах коррекции недостатков слуха, позволяя обеспечить приемлемое качество звучания и хорошую разборчивость речи. Кроме того, алгоритмы шумоподавления используются как препроцессоры в системах автоматического распознавания речи (APP), качество работы которых существенно зависит от уровня шумовых помех [1 – 4].

Эффективность алгоритмов шумоподавления существенно зависит от выбора способов и параметров оценки априорного SNR. Поэтому поиски новых, более эффективных, способов оценивания априорного SNR не прекращаются, о чем свидетельствуют результаты ряда недавних работ [5 – 7]. Предложенные сравнительно недавно алгоритмы шумоподавления Wiener-TSNR и Wiener-HRNR, реализующие один из таких способов, весьма впечатляют своей способностью к радикальному подавлению шумовой помехи [5, 6]. Однако более тщательная проверка качества этих алгоритмов показывает, что достигается это ценой внесения существенных искажений в речевой сигнал. В результате страдает не только качество речевого сигнала, но и существенно снижается качество работы систем APP. Поскольку алгоритмы Wiener-TSNR и Wiener-HRNR основаны на особой коррекции оценки априорного SNR, можно надеяться повысить эффективность этих алгоритмов, варьируя способом и параметрами оценки априорного SNR.

Целью данной работы является проверка справедливости данного предположения.

1. Постановка задачи

С целью анализа влияния метода и параметров оценки априорного отношения сигнал-шум (SNR) на качество алгоритма подавления шума, в данной работе сопоставляются три способа оценивания апри-

орного SNR. При этом рассматриваются три алгоритма шумоподавления: традиционный алгоритм Винеровской фильтрации (Wiener), а также два сравнительно недавно предложенных алгоритма Wiener-TSNR и Wiener-HRNR [5,6].

1.1. Алгоритмы шумоподавления

Все рассматриваемые в данной работе алгоритмы шумоподавления реализуют метод частотной коррекции, согласно которому спектр амплитуд восстановленного сигнала оценивают путем цифровой фильтрации:

$$\left[\hat{\lambda}_x(f, m) \right]^{1/2} = \hat{G}(f, m) \cdot \left[\hat{\lambda}_y(f, m) \right]^{1/2},$$

где $\hat{\lambda}_y(f, m)$ и $\hat{\lambda}_x(f, m)$ – оценки спектров мощности m -го фрейма сигнала $y(t)$ и восстановленного сигнала $\hat{x}(t)$, $\hat{G}(f, m)$ – оценка передаточной функции корректирующего фильтра, используемого при обработке m -го фрейма сигнала $y(t)$. При восстановлении сигнала $\hat{x}(t)$ обычно используют фазовый спектр сигнала $y(t)$.

Для алгоритма Wiener

$$\text{где } \hat{G}_{\text{wnr}}(f, m) = \frac{\hat{\xi}(f, m)}{1 + \hat{\xi}(f, m)}, \quad (1)$$

$$\hat{\xi}(f, m) = \frac{\hat{\lambda}_s(f, m)}{\hat{\lambda}_d(f, m)} - \quad (2)$$

оценка априорного SNR, связанная с оценкой апостериорного SNR $\hat{\gamma}(f, m) = \frac{\hat{\lambda}_y(f, m)}{\hat{\lambda}_d(f, m)}$ соотношением

$$\hat{\xi}(f, m) = \hat{\gamma}(f, m) - 1.$$

Передаточную функцию алгоритма Wiener-TSNR формируют в два этапа.

Шаг 1. Оценивают априорное SNR:

$$\hat{\xi}_{\text{TSNR}}(f, m) = \hat{\xi}_{\text{DD}}(f, m+1) \approx \hat{\lambda}_s(f, m) / \hat{\lambda}_d(f, m), \quad (3)$$

где $\hat{\xi}_{DD}(f, m)$ – оценка $\hat{\xi}(f, m)$, вычисленная методом «управление решением» (decision-directed - DD) [2].

Шаг 2. Формируют передаточную функцию алгоритма Wiener-TSNR:

$$\hat{G}_{TSNR}(f, m) = \frac{\hat{\xi}_{TSNR}(f, m)}{1 + \hat{\xi}_{TSNR}(f, m)}. \quad (4)$$

Алгоритм Wiener-HRNR предназначен для регенерации гармонических компонентов, утраченных при подавлении шума. Этапы формирования соответствующей передаточной функции при этом таковы.

Этап 1. Восстановленный с помощью алгоритма TSNR сигнал $\hat{s}(t)$ подвергают однополупериодному линейному детектированию:

$$s_{\text{harm}}(t) = \hat{s}(t) \cdot P[\hat{s}(t)], \quad (5)$$

$$P[x] = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Этап 2. Оценивают априорное отношение сигнал-шум:

$$\hat{\xi}_{HRNR}(f, m) = \rho(f, m) \cdot \hat{\lambda}_s(f, m) / \hat{\lambda}_d(f) + [1 - \rho(f, m)] \cdot \hat{\lambda}_{\text{harm}}(f, m) / \hat{\lambda}_d(f), \quad (6)$$

где $\hat{\lambda}_{\text{harm}}(f, m)$ – оценка спектра мощности сигнала (5); $\rho(f, m)$, $0 \leq \rho(f, m) \leq 1$ – весовой коэффициент. Хотя при задании $\rho(f, m)$ имеется определенная свобода выбора, в [5] предложено принимать $\rho(f, m) = \hat{H}_{TSNR}(f, m)$.

Этап 3. Формируют коэффициент передачи алгоритма HRNR:

$$\hat{G}_{HRNR}(f, m) = \frac{\hat{\xi}_{HRNR}(f, m)}{1 + \hat{\xi}_{HRNR}(f, m)}. \quad (7)$$

1.2. Способы оценивания априорного SNR

В соотношениях (2) – (7) важную роль играет оценка априорного отношения сигнал-шум $\hat{\xi}(f, m)$. В данной работе рассмотрим три разновидности такой оценки:

- «грубая» оценка;
- оценка, сформированная методом «управление решением»;
- оценка, сформированная методом максимального правдоподобия (МП).

«Грубой» будем называть оценку вида:

$$\hat{\xi}_{RO}(f, m) = \hat{\gamma}(f, m) - 1. \quad (8)$$

Оценка, сформированная методом «управление решением», имеет вид:

$$\hat{\xi}_{DD}(f, m) = \alpha \cdot \hat{\lambda}_s(f, m - 1) / \hat{\lambda}_d(f) + (1 - \alpha) \cdot P[\hat{\gamma}(f, m) - 1], \quad 0 \leq \alpha \leq 1, \quad (9)$$

где $\hat{\lambda}_s(f, m) = \hat{G}^2(f, m) \hat{\lambda}_y(f, m)$, α – параметр усреднения. В [2] отмечено, что оптимальным значением является $\alpha = 0,98$ при частоте дискретизации $F_s = 8$ кГц и сдвиге фреймов $N_{\text{inc}} = 64$ выборки. С параметром α тесно связан параметр, именуемый «постоянной времени усреднения»:

$$\tau_{\text{avr}} = -\frac{N_{\text{inc}}/F_s}{\ln \alpha}.$$

На наш взгляд, параметр τ_{avr} более удобен как для теоретического анализа, так и для использования в приложениях, поскольку указывает размеры участка речевого сигнала, подвергаемого статистической обработке. При частоте дискретизации $F_s = 8$ кГц и сдвиге фреймов $N_{\text{inc}} = 64$ оптимальное значение $\tau_{\text{avr}} = 0,396$ с. В данной работе использовали сигналы, дискретизированные с частотой $F_s = 22050$ Гц, со сдвигом фреймов $N_{\text{inc}} = 64$, поэтому значения постоянных времени τ_{avr} , соответствующих различным α , будут иными (табл. 1). Как следует из табл. 1, значение $\tau_{\text{avr}} = 0,312$ ближе всего к предложенному в [2] оптимальному значению $\tau_{\text{avr}} = 0,396$, поэтому в рамках данной работы близким к оптимальному является значение $\alpha = 0,95$.

Таблица 1

Значения параметра усреднения и постоянной времени

α	0,02	0,5	0,8	0,9	0,95	0,98
$\tau_{\text{avr}}, \text{с}$	0,004	0,023	0,072	0,152	0,312	0,792

Оценка, сформированная методом МП, имеет вид:

$$\hat{\xi}_{ML}(f, m) = P(\bar{\gamma}(f, m) - 1), \quad (10)$$

$$\bar{\gamma}(f, m) = \alpha' \cdot \bar{\gamma}(f, m - 1) + (1 - \alpha') \cdot \hat{\gamma}(f, m) / \beta,$$

$$0 \leq \alpha' \leq 1, \quad \beta \geq 1,$$

где β – дополнительный корректирующий коэффициент. В [2] отмечено, что выбор α' целесообразно производить, соизмеряя постоянную времени экспоненциального усредняющего окна с интервалом корреляции $\gamma(m)$. По существу, данная рекомендация подтверждает высказанное выше соображение об удобстве использования постоянной времени усреднения как более информативного параметра.

2. Методика исследований

Эффективность способов оценивания априорного отношения SNR оценивалась с использованием нескольких показателей качества алгоритмов шумоподавления. Поскольку аналитическими методами такое оценивание осуществить невозможно, применялись методы экспериментального оценивания, с использованием реализаций речевых сигналов, искусственно зашумленных дискретным белым шумом различной интенсивности.

2.1. Показатели качества алгоритмов шумоподавления

При оценке эффективности алгоритмов шумоподавления использованы три показателя качества. Во-первых, это показатель, именуемый «точностью распознавания» системы APP [8]:

$$\text{Acc\%} = \frac{N - D - S - I}{N} \times 100\% ,$$

где N – общее количество распознаваемых слов; D – количество ошибочных удалений; S – количество замен; I – количество ошибочных вставок.

Кроме того, в роли показателей качества речи использованы перцептуальная оценка качества речи PESQ, являющаяся стандартом для линий связи [9], и логарифмическое отношение правдоподобия LLR, оказавшееся весьма эффективной мерой качества и разборчивости речи [10, 11].

Показатель LLR вычисляют для каждого из сегментов речевого сигнала, а затем полученные результаты усредняют. Вычисления для сегментов производят в соответствии с соотношением

$$\text{LLR}_{\text{seg}} = \ln \left(\frac{\bar{a}_p \mathbf{R}_c \bar{a}_p^T}{\bar{a}_c \mathbf{R}_c \bar{a}_c^T} \right) ,$$

где \bar{a}_c и \bar{a}_p – коэффициенты линейного предсказания чистого и откорректированного сигналов, соответственно; \mathbf{R}_c – матрица коэффициентов автокорреляции чистого сигнала.

Алгоритм вычислений PESQ намного более громоздок, его описание можно найти в [12].

2.2. Организация экспериментальных исследований

Чистые речевые сигналы оцифровывались и записывались с параметрами: частота дискретизации 22100 Гц, линейное квантование глубиной 16 бит. Запись происходила в заглушенном помещении со временем реверберации 0,1 с, отношение сигнал-шум для записанных сигналов было близким 35 дБ. Обучение системы APP происходило именно на таких образцах чистых речевых сигналов. Тестирование системы APP осуществлялось на образцах зашумленной речи. Добавление дискретного белого шума $n(t)$ к чистому сигналу $x(t)$ выполнялось в среде Matlab, с обеспечением широкого диапазона значений отношения сигнал-шум входного сигнала, от минус 10 дБ до плюс 30 дБ. Следующие этапы обработки – коррекция искаженного сигнала $y(t)$, вычисление показателей качества восстановленного сигнала $\hat{x}(t)$, автоматическое распознавание речи и вычисление показателя Acc% – также выполнялись в среде Matlab. Перцептуальный показатель качества речи PESQ оценивался для широкой полосы частот 7 кГц с использованием специального программного обеспечения [13].

Для моделирования системы APP и оценки показателя Acc% применялся программный инструментарий НТК [8]. Обучение системы APP выполнялось с использованием 269 образцов 27 слов украинского языка, записанных для 2-х дикторов-женщин. При этом фонемный словарь состоял из 27 фонем украинского языка. Использовались 39-мерные классификационные признаки типа MFCC_0_D_A. Тестовый сигнал представлял собой зашумленный звуковой файл дискретной речи с записью всех 27 слов, использованных при обучении, с паузами между словами 0,2...0,5 с.

3. Результаты исследований

3.1. Оптимизация параметра усреднения в методе «управление решением»

Результаты оценки влияния варьирования параметром α на точность автоматического распознавания представлены на рис. 1.

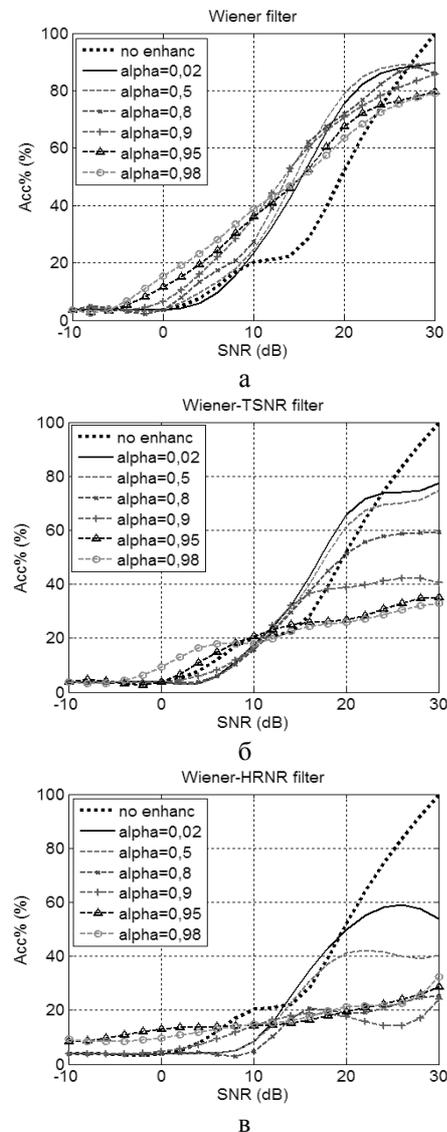


Рис. 1. Варьирование параметром α для винеровского фильтра (а), TSNR-фильтра (б) и HRNR-фильтра (в)

Как следует из рис. 1, рекомендованное в [2] и считающееся сегодня общепринятым оптимальное, в смысле качества речи, значение параметра $\alpha = 0,98$ является оптимальным и в смысле максимума показателя Acc%, однако это справедливо лишь для $SNR < 15$ дБ. Для $SNR > 15$ дБ оптимальным, в смысле максимума показателя Acc%, является значение $\alpha = 0,02 \dots 0,5$ для винеровского фильтра или $\alpha = 0,02$ для оценок Wiener-TSNR и Wiener-HRNR.

Результаты влияния варьирования параметром α на качество речи (показатели WB-PESQ и LLS) представлены на рис. 2, 3.

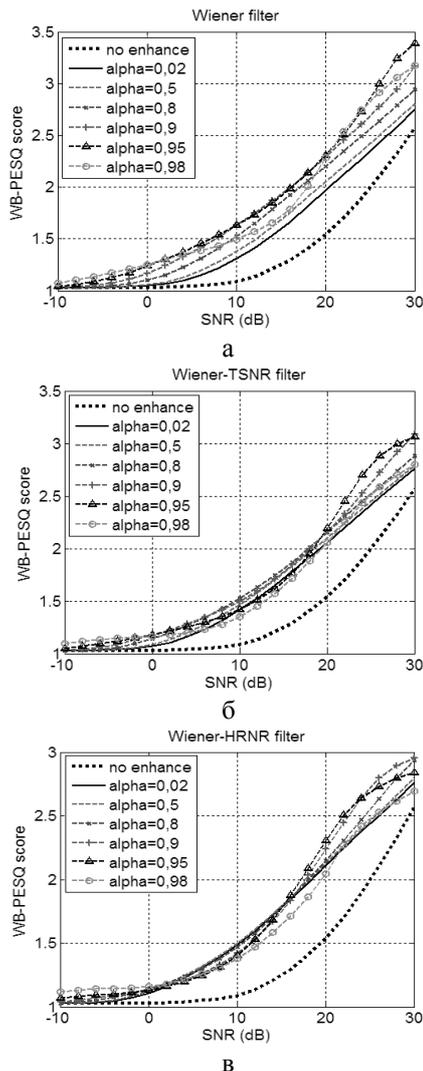


Рис. 2. Варьирование α для винеровского фильтра (а), TSNR-фильтра (б) и HRNR-фильтра (в) для показателя WB-PESQ

Как следует из приведенных на рис. 2, а графиков, для винеровского алгоритма значение $\alpha \approx 0,95$ (постоянная времени 0,312 с) является наилучшим в смысле WB-PESQ, а значения $\alpha \approx 0,02 \dots 0,5$ являются наихудшими для всего рассмотренного диапазона SNR. Алгоритмы Wiener-TSNR и Wiener-HRNR (рис. 3, б, в) менее чувствительны к варьированию значениями параметра α .

Несколько сложнее поведение графиков для показателя LLR. В случае винеровского алгоритма (рис. 3, а) значение $\alpha \approx 0,98$ может считаться оптимальным для $SNR < 10$ дБ, однако для $SNR > 10$ дБ предпочтительными оказываются значения $\alpha \leq 0,95$. В случае алгоритмов Wiener-TSNR и Wiener-HRNR (рис. 3, б, в) зависимость LLR от α и SNR выражена значительно сильнее: для $SNR > 10$ дБ оптимальными являются значения $\alpha \approx 0,02 \dots 0,5$.

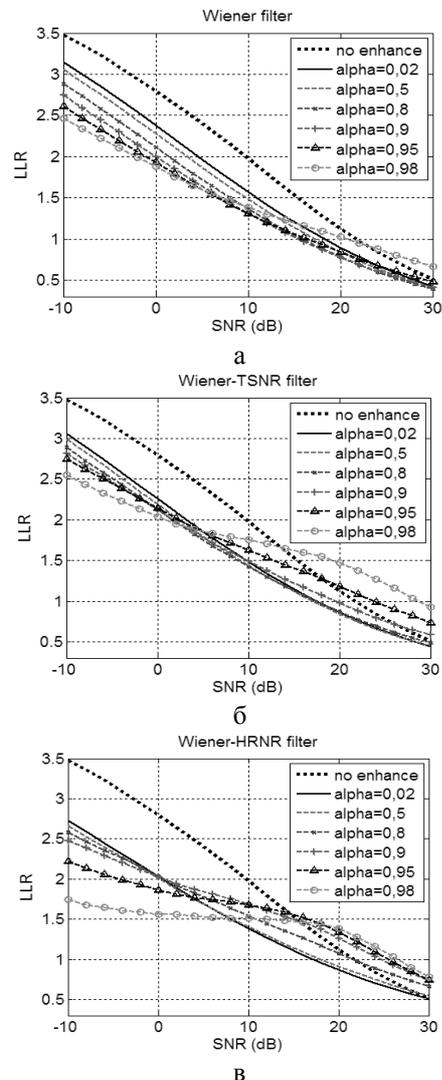


Рис. 3. Варьирование α для винеровского фильтра (а), TSNR-фильтра (б) и HRNR-фильтра (в) для показателя LLR

Нетрудно заметить, что поведение показателя LLR как функции параметров α и SNR напоминает поведение показателя Acc% (рис. 1). Данное обстоятельство позволяет рекомендовать показатель LLR как приближенную меру точности распознавания речи.

3.2. Оптимизация параметра усреднения в методе МП

Рассмотрим результаты исследований влияния значений параметра усреднения α' соотношения (10) на поведение показателей мер Acc%, PESQ и

LRR. Учитывая невысокое качество алгоритмов Wiener-TSNR и Wiener-HRNR, ограничимся рассмотрением винеровского алгоритма.

В экспериментах было принято $\beta = 1$, а значения параметра α' изменялись от 0,02 до 0,98. Результаты вычислений Acc% приведены на рис. 4, а, а результаты вычислений WB-PESQ и LLR - на рис. 4, б, в.

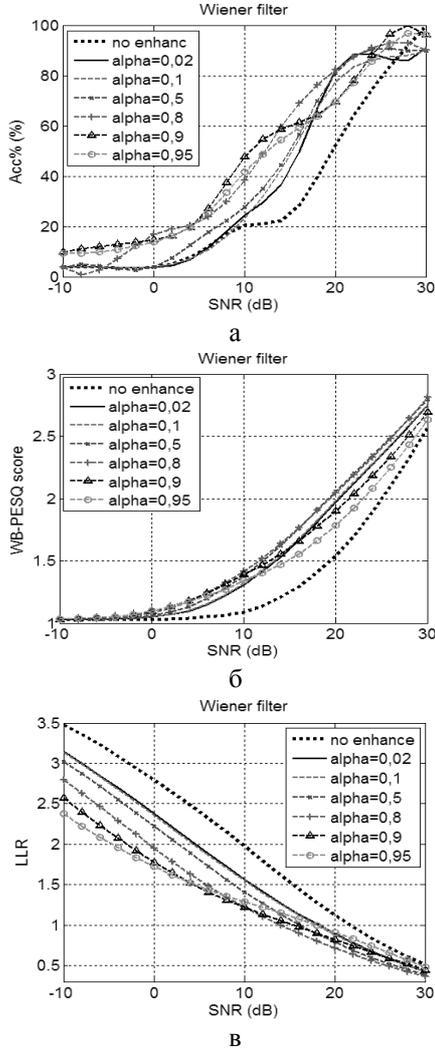


Рис. 4. Вычисление априорного SNR методом МП

При $\text{SNR} > 10$ дБ для всех показателей оптимальным можно считать $\alpha' = 0,8$. Интересно, что достигаемые при этом значения Acc% весьма близки к полученным методом DD при $\alpha = 0,02 \dots 0,5$ (рис. 1, а).

3.3. «Грубая» оценка априорного SNR

Отмеченная выше оптимальность алгоритма шумоподавления, в смысле максимума Acc%, при $\alpha = 0,02 \dots 0,5$ и $\text{SNR} > 15$ дБ, позволяет предположить, что «грубая» оценка (8) для данных значений SNR предпочтительнее оценки (9) с типичным для нее оптимальным значением $\tau_{\text{avr}} = 0,396$. Действительно, полагая в (9) $\alpha \rightarrow 0$, получаем

$$\lim_{\alpha \rightarrow 0} \hat{\xi}_{\text{DD}}(f, m) \approx \hat{\gamma}(f, m) - 1 = \hat{\xi}_{\text{RO}}(f, m),$$

т.е. при малых значениях параметра усреднения α оценка DD эквивалентна «грубой» оценке.

Естественно также предположить, что использование оценки (10) приведет к некоторому промежуточному результату. Представленные на рис. 5, а результаты подтверждают справедливость этих предположений.

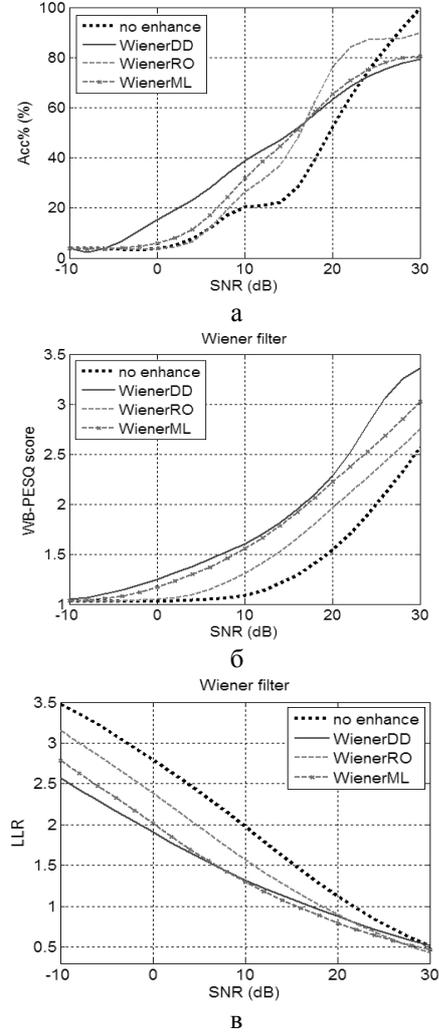


Рис. 5. Сопоставление методов оценивания априорного SNR

Здесь кривая WienerDD соответствует оцениванию ξ методом «управление решением» для $\tau_{\text{avr}} = 0,396$. Кривая WienerRO соответствует «грубому» оцениванию ξ в соответствии с (8). Кривая WienerML соответствует способу оценивания ξ в соответствии с (10), где принято $\alpha' = 0,5273$, $\beta = 2$.

Вместе с тем, метод DD обеспечивает наилучшее качество речи в смысле WB-PESQ для всех значений SNR при постоянной времени 0,396 с, а грубая оценка (RO) априорного SNR приводит к наихудшему качеству речи. Показатель LLR ведет себя аналогично, но лишь для $\text{SNR} < 10$ дБ. Для $\text{SNR} > 10$ дБ поведение LLR становится подобным поведению Acc%, хотя и в значительно менее выраженной степени.

Как видим, «грубая» оценка априорного SNR привела к неудовлетворительным результатам в плане качества речи, однако для ASR «грубая» оценка оказалась наилучшей при $SNR > 15$...17 дБ.

Выводы

Сопоставлены три метода оценивания априорного SNR: «управление решением», максимального правдоподобия и «грубая» оценка. Выработаны рекомендации по выбору параметров усреднения при оценивании априорного отношения сигнал-шум методом «управление решением» и методом максимального правдоподобия.

Показано, что при шумоподавлении в системе автоматического распознавания речи метод «управление решением» позволяет обеспечить наилучшие результаты при $SNR < 15$ дБ, однако при $SNR > 15$ дБ этот метод существенно проигрывает «грубой» оценке. При сопоставлении с использованием показателей качества речи ситуация обратная. Метод максимального правдоподобия занимает промежуточное положение между методом «управление решением» и «грубой» оценкой.

Полученные результаты позволяют осуществлять аргументированный выбор метода и параметров оценки априорного отношения сигнал-шум в алгоритмах шумоподавления, используемых при решении задач повышения качества речевых сигналов, а также задач повышения точности автоматического распознавания речи.

Список литературы

1. Springer Handbook of Speech Processing / Ed. J. Benesty, M. Sondhi, Y. Huang. – Berlin Heidelberg: Springer-Verlag, 2008. – 1176 p.
2. Ephraim Y. Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator / Y. Ephraim, D. Malah // IEEE Trans. on Acoustic, Speech, and Signal Processing. – Dec. 1984. – Vol. ASSP-32, No. 6. – P. 1109-1121.
3. Ephraim Y. Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator /

Y. Ephraim, D. Malah // IEEE Transactions on Acoustic, Speech, and Signal Processing. – Apr. 1985. – Vol. ASSP-33, No. 2. – P. 443-445.

4. Hu Y. Evaluation of objective quality measures for speech enhancement / Y. Hu, P. Loizou // IEEE Transactions on Speech and Audio Processing. – 2008. – 16(1). – P. 229-238.
5. Plapous C. Improved Signal-to-Noise Ratio Estimation for Speech Enhancement / C. Plapous, C. Marro, P. Scalart // IEEE Transactions on Audio, Speech, and Language Processing. – Nov. 2006. – Vol. 14, Issue 6. – P. 2098-2108.
6. Plapous C. A Two-Step Noise Reduction Technique / C. Plapous, C. Marro, P. Scalart, L. Mauuary // IEEE Int. Conf. on Acoustics, Speech and Signal Proc., Vol. 1, 17–21 May, 2004. – P. 289-292.
7. Djendi M. Reducing over- and under-estimation of the a priori SNR in speech enhancement techniques / M. Djendi, P. Scalart // Digital Signal Processing. – May 2014. – Vol. 32. – P. 124-136.
8. The HTK Book / Ed. S. Young, G. Evermann, M. Gales. – Cambridge: University Engineering Department, 2009. – 375 p.
9. Recommendation P.862 (2001) Amendment 2 (11/05) [Online]. Available: <http://www.itu.int/rec/T-REC-P.862-200511-I!Amd2/en>.
10. Hu Y. Evaluation of objective quality measures for speech enhancement / Y. Hu, P. Loizou // IEEE Transactions on Speech and Audio Processing. – 2008. – 16(1). – P. 229-238.
11. Ma J. Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions / J. Ma, Y. Hu, P. Loizou // J. Acoust. Soc. Am. – May 2009. – Vol. 125, No. 5. – P. 3387-3405.
12. Beerends J. PESQ, the new ITU standard for objective measurement of perceived speech quality. Part II – Perceptual model / J. Beerends, A. Hekstra, A. Rix, M. Hollier // J. Audio Eng. Soc., Oct. 2002. – Vol. 50. – P. 765-778.
13. Продеус А. Вычисления показателя качества речи PESQ в среде MATLAB / А. Продеус // Труды 14-й международной научно-практической конференции «Новейшие сетевые технологии в Украине», 17-19 сентября 2012 года, м.т. Партемит, научно-производств. сборник "Вестник УНИИИС", Киев. – 2012. – С. 70-76.

Поступила в редколлегию 14.08.2015

Рецензент: д-р техн. наук, проф. А.И. Рыбин, Национальный технический университет Украины «КПИ», Киев.

ОЦІНКА АПРІОРНОГО ВІДНОШЕННЯ СИГНАЛ-ШУМ В АЛГОРИТМАХ ШУМОЗАГЛУШЕННЯ

А.М. Продеус, В.С. Дідковський

Зіставлені три методи оцінювання априорного відношення сигнал-шум (SNR) в алгоритмах шумозаглушення. Показано, що оцінка за методом «керування рішенням» дозволяє забезпечити максимально високу якість звучання мовного сигналу. «Груба» оцінка при інтегральному $SNR > 15$ дБ виявилася кращою для використання в системах автоматичного розпізнавання мовлення. Оцінка максимальної правдоподібності займає проміжне місце між оцінками за методом «керування рішенням» та «грубою» оцінкою.

Ключові слова: шумова перехода; алгоритм шумозаглушення; априорне відношення сигнал-шум; показник якості; мовленнєвий сигнал.

ASSESSMENT OF A PRIORI SIGNAL-TO-NOISE RATIO IN NOISE REDUCTION ALGORITHMS

A.N. Prodeus, V.S. Didkovskiy

Three methods of estimation of the a priori signal-to-noise ratio (SNR) in noise reduction algorithms were compared. It is shown that "decision-directed" method ensures the best sound quality of the speech signal. "Rough" estimate was the best for use in automatic speech recognition for the overall $SNR > 15$ dB. The maximum likelihood estimation occupies an intermediate position between estimates by "decision-directed" and "rough" estimator.

Keywords: noise interference; noise reduction algorithm; a priori signal-to-noise ratio; quality indicator; speech signal.