

МЕТОД ПОДАВЛЕНИЯ ЭХО СИГНАЛА И ШУМОВ ОКРУЖАЮЩЕЙ СРЕДЫ НА ОСНОВЕ СПЕКТРАЛЬНОГО ВЫЧИТАНИЯ С ПСИХОАКУСТИЧЕСКОЙ МОТИВАЦИЕЙ

Аношенко А. Е., Петровский А. А.

Белорусский государственный университет информатики и радиоэлектроники
220027, Минск, ул. П.Бровки, 6 (Беларусь), E-mail: palex@it.org.by

Реферат. В данной работе предлагается комбинированный метод подавления эхо сигнала и шумов окружающей среды на основе спектрального вычитания учитывающий психоакустические особенности восприятия человека. Данная система осуществляет обработку сигнала только в частотной области, что позволяет значительно снизить вычислительную сложность алгоритмов, по сравнению с реализацией во временной области. В работе описываются результаты моделирования работы системы, а также проводится сравнение с другими подходами к решению проблемы эхо и шума подавления.

Введение

Громкоговорящая связь (hands free) с каждым годом находит все большее и большее применение в мобильных системах. Особенностью данного вида связи является использование вместо традиционной телефонной трубки системы встроенных в панели управления ненаправленных микрофонов и громкоговорителей. В результате на вход микрофона приходит не только речь говорящего человека, но и многократно отраженный сигнал от громкоговорителей (акустическое эхо) и различные шумы окружающей среды. Это делает невозможным использование громкоговорящих систем без специального оборудования для подавления эхо сигнала и шумов окружающей среды. Анализ различных вариантов построения данных систем проведен в [1].

В данной работе предлагается вариант комбинированной системы подавления эхо сигнала и шумов окружающей среды в которой отсутствует отдельная подсистема подавления эха и подавление всех искажающих составляющих входного сигнала осуществляется на основе метода спектрального вычитания с психоакустической мотивацией.

Система подавления эха и шумов окружающей среды на основе метода спектрального вычитания с психоакустической мотивацией

Блок схема предлагаемой системы показана на рис. 1. Данная система удаляет шумовые эхо составляющие из входного сигнала умножая входной сигнал на взвешивающую функцию. Взвешивающая функция H^{JND} [2] в момент времени n находится на основе спектральной плотности мощности (СПМ) шумовой составляющей сигнала и спектральной плотности акустического эха с учетом психоакустических особенностей восприятия человека.

$$H^{JND}(n, f) = \min \left(\frac{\mathbf{S}_n(n, f)\zeta_n + \mathbf{S}_d(n, f)\zeta_d}{\mathbf{S}_n(n, f) + \mathbf{S}_d(n, f)} + \sqrt{\frac{\mathbf{S}_T(n, f)}{\mathbf{S}_n(n, f) + \mathbf{S}_d(n, f)}}, 1 \right) \quad 1)$$

где ζ_n и ζ_d – требуемый уровень шума и эха соответственно, $\mathbf{S}_n(n)$ – СПМ шума, $\mathbf{S}_d(n)$ – СПМ эхо сигнала, $\mathbf{S}_T(n)$ – абсолютный порог маскирования, f – номер отсчета частотной характеристики. Методика получения абсолютного порога маскирования приведена в [3]. Для нахождения взвешивающей функцию также необходимо знать спектральную плотность мощности шума и эхо сигнала.

Для оценки СПМ шума обычно используется либо метод минимальной статистики, либо алгоритм оценки спектральной плотности мощности в речевых паузах. В предлагаемой системе для оценки мощности шума используется второй метод.

Для оценки СПМ шума сначала рассчитывается среднее значение мощности входного сигнала во время речевых пауз \mathbf{S}_{av} по формуле

$$\mathbf{S}_{av}(n, f) = \begin{cases} \alpha \cdot \mathbf{S}_y(n, f) + (1 - \alpha) \cdot \mathbf{S}_{av}(n - 1, f), & \text{silence} = 1 \\ \mathbf{S}_{av}(n - 1, f), & \text{silence} = 0 \end{cases} \quad 2)$$

где $\mathbf{S}_y(n)$ – спектральная плотность мощности входного сигнала y , α – коэффициент в диапазоне [0;1], $\mathbf{S}_{av}(n)$ – новое значение среднего СПМ входного сигнала, $\mathbf{S}_{av}(n-1)$ – предыдущее значение \mathbf{S}_{av} , silence – управляющий сигнал от детектора тишины, который принимает значение 1 если отсутствует речевая активность и 0 – в противоположном случае. СПМ шума \mathbf{S}_n находится на основе \mathbf{S}_{av} по формуле

$$\mathbf{S}_n(n, f) = \min(\mathbf{S}_{av}(n, f), \mathbf{S}_y(n, f)) \quad 3)$$

Спектральную плотность мощности эхо сигнала \mathbf{S}_d можно представить

$$S_d(n, f) = \sum_{i=0}^N \mathbf{W}(i, f) \cdot S_x(n-i, f), \quad (4)$$

где N – определяет максимальную длительность эхо сигнала, \mathbf{W}_i – коэффициент, характеризующий ослабление сигнала в результате многократного отражения и поглощения в момент времени $n-i$. Таким образом матрица \mathbf{W} характеризует акустическую обстановку помещения.

Формула (4) используется для расчета СПМ эхо сигнала в предлагаемом методе. Для нахождения акустической характеристики помещения \mathbf{W} используется метод наискорейшего спуска, а в частности нормализованный метод наименьших квадратов (НМНК). Выбор данного метода обусловлен тем, что он имеет наибольшую скорость сходимости по сравнению с другими модификациями метода наискорейшего спуска.

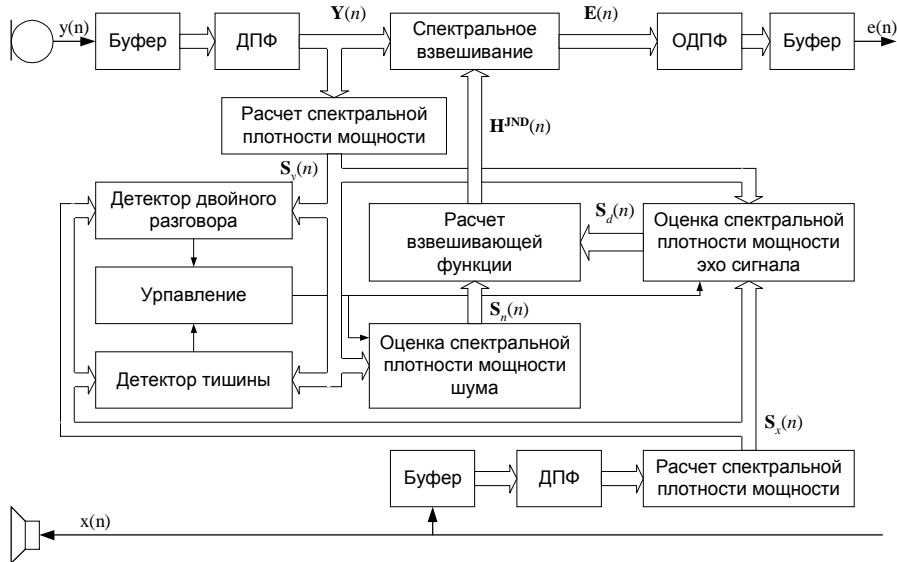


Рис. 1. Одномикрофонная комбинированная система подавления эха и редактирования шума на основе метода спектрально вычитания с психоакустической мотивацией.

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

$$\mathbf{W}(i, j) = \max \left(\mathbf{W}(i, j) + \frac{\mu \cdot (S_y(n, i) - S_d(n, i)) \cdot S_x(n - j, i)}{\|S_x(n - j)\|^2}, 0 \right) \quad (6)$$

где μ – коэффициент сходимости алгоритма в диапазоне $[0;1]$.

Блок схема системы оценки СПМ эхо сигнала показана на рис.2.

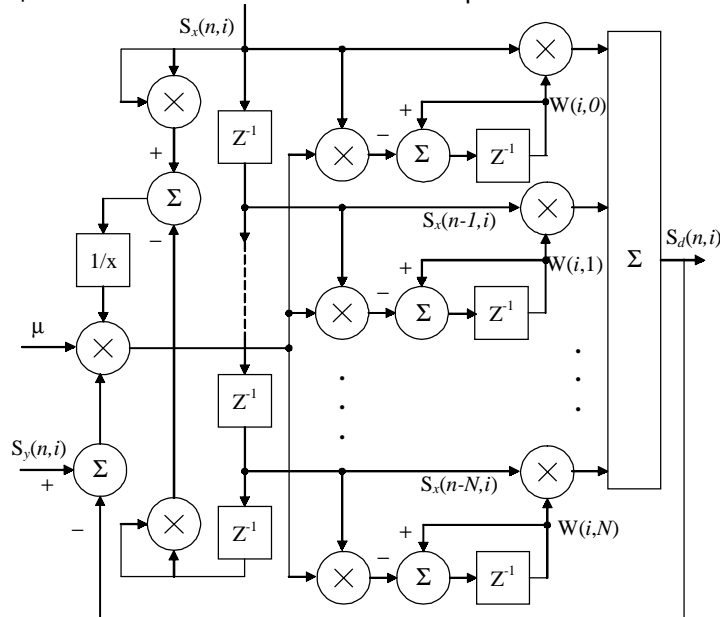


Рис. 2. системы оценки СПМ эхо сигнала.

Заключение

Было проведено моделирование работы двух систем подавления эха и шумов окружающей среды. а третья – на схеме показанной на рис.2. В первой системе для подавления эхо сигнала использовался адаптивный фильтр на основе НМНК алгоритма, а вторая была построена по схеме показанной на рис. 1. Параметры систем следующие: частота дискретизации 8 кГц, размер блока – 32 отсчета, длина ДПФ – 256 (временная задержка не более 0,01 с), порядок фильтра $N=1120$ (длительность эха 0,14 с).

Результаты моделирования показали, что все системы позволяет получить примерно одинаковое ослабление эха порядка $ERLE=30-35$ дБ и ослабление уровня шума также порядка 30-35 дБ. Вторым критерием сравнения данных систем являлась вычислительная сложность. Система на основе спектрального вычитания имеет в 22 раза меньшую вычислительную сложностью по сравнению с алгоритмом на базе НМНК.

Библиография

1. A.Petrovsky, K.Bielawski, A.Anoshenko, Hands-free radiotelephony communication devices with combine front end processing systems: global approaches in the time and frequency domain. Journal of the University of Applied Sciences Mittweida Wissenschaftliche Zeitschrift der Hochschule Mittweida (FH). Proc. IWKM'98, Band C, Nr.3, 1998, pp.135-142.
2. S.Gustafsson, R.Martin, P.Vary, Combined acoustic echo control and noise reduction for hands-free telephony. Signal Processing, N.64, 1998, pp. 21-32.
3. Петровский А.А., Аношенко А.Е. Комбинированная система подавления эха и шумового сигналов на основе обработки в частотной области с использованием психоакустического подхода // 2-я Международная конференция "Цифровая обработка сигналов и ее применение", Москва, 21-24 сентября 1999, том. 1 – с. 166-169.



METHOD OF ECHO CANCELLATION AND NOISE REDUCTION ON THE BASIS OF SPECTRAL SUBTRACTION WITH PSYCHOACOUSTICS MOTIVATION

Anoshenko A., Petrovsky A.

The Byelorussian State University of Informatics and Radioelectronics
220027, Minsk, P.Brovki st., 6 (Belarus)
E-mail: palex@it.org.by

In the given paper the variant of the combined system of an echo and noise cancellation is presented. In proposed system cancellation of an echo and of all noise components is carried out on the basis of a method of spectral subtraction with psychoacoustics motivation. For suppression of noise and echo components the input signal is multiplied on weighing function. Weighing function H^{JND} [1] at the moment of time n is calculated on the basis of the power spectral density (PSD) of signal's noise component and of the power spectral density of acoustic echo taking into account the psychoacoustics features of the human perception.

$$H^{JND}(n, f) = \min \left(\frac{\mathbf{S}_n(n, f)\zeta_n + \mathbf{S}_d(n, f)\zeta_d}{\mathbf{S}_n(n, f) + \mathbf{S}_d(n, f)} + \sqrt{\frac{\mathbf{S}_T(n, f)}{\mathbf{S}_n(n, f) + \mathbf{S}_d(n, f)}}, 1 \right), \quad 1)$$

where ζ_n and ζ_e – required noise and echo levels accordingly, $\mathbf{S}_n(n)$ – PSD of a noise signal, $\mathbf{S}_d(n)$ – PSD of an echo signal, $\mathbf{S}_T(n)$ – absolute masking threshold, f – number of a sample of the frequency characteristics. The technique of calculating the absolute masking threshold is given in [2].

The method of the PSD estimation during speech pauses is usually used for an estimation of the noise PSD. For this purpose an average value of the input signal spectral power during speech pauses \mathbf{S}_{av} is calculated by the formula:

$$\mathbf{S}_{av}(n, f) = \begin{cases} \alpha \cdot \mathbf{S}_y(n, f) + (1 - \alpha) \cdot \mathbf{S}_{av}(n - 1, f), & \text{silence} = 1 \\ \mathbf{S}_{av}(n - 1, f), & \text{silence} = 0' \end{cases} \quad 2)$$

where $\mathbf{S}_y(n)$ – PSD of the input signal y , α – a coefficient from the set [0;1], silence – output signal from the silence detector, which is set to 1 if there is no speech activity and to 0 in other case. PSD of noise \mathbf{S}_n is calculated on a basis of \mathbf{S}_{av} by the formula:

$$\mathbf{S}_n(n, f) = \min(\mathbf{S}_{av}(n, f), \mathbf{S}_y(n, f)), \quad 3)$$

Power spectral density of an echo signal \mathbf{S}_d can be presented as:

$$\mathbf{S}_d(n, f) = \sum_{i=0}^N \mathbf{W}(i, f) \cdot \mathbf{S}_x(n - i, f), \quad 4)$$

where N defines the maximum duration of an echo signal, \mathbf{W}_i – coefficient describing attenuation of a signal as a result of repeated reflecting and absorbing. Initially the matrix \mathbf{W} must contain zero values. While processing each block of an input signal, the matrix \mathbf{W} is corrected according to the formula below only when the far end user speaks.

$$\mathbf{W}(i, j) = \max \left(\mathbf{W}(i, j) + \frac{\mu \cdot (\mathbf{S}_y(n, i) - \mathbf{S}_d(n, i)) \cdot \mathbf{S}_x(n - j, i)}{\|\mathbf{S}_x(n - j)\|^2}, 0 \right) \quad 5)$$

where μ – coefficient of algorithm convergence in the range [0;1].

The modeling results show that the system allows echo attenuation around ERLE = 30-35 dB and noise attenuation level around 30-35 dB. System parameters are as follows: sampling frequency is 8 kHz, the block size is 32 samples, FFT length is 256, and the filter tap is 1120.

REFERENCE

1. S.Gustafsson, R.Martin, P.Vary, Combined acoustic echo control and noise reduction for hands-free telephony. Signal Processing, N.64, 1998, pp. 21-32.
2. Petrovsky A.A., Anoshenko A.E. Combined echo and noise cancellation system based on the frequency domain processing with using the psychoacoustically approach. // The 2nd International Conference "Digital signal processing and its applications", Moscow, 1999 September 21-24, Vol. 1 – pp. 166-169.