Sparse Representation of Human Auditory System

M.Edalatian, A. A. Soltani, and N.Faraji

Imam Khomeini International University, Faculty of Engineering and Technology, Department of Electrical Engineering, Qazvin, Iran

I. Definition and outline

- Biological studies shows that the hair cells in the inner ear of the auditory system generate sparse codes from the output of cochlea filter-bank[1].

II. Sparse Representation

■ Sparse representation problem can be formulated as [2]

$$(P_0): \min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{s.t}\mathbf{Ax}{=}\mathbf{b}, \quad (1)$$

where $\|\mathbf{x}\|_0$ denotes the number of non-zero elements of $\mathbf{x}$ and $\mathbf{A} \in \mathbb{R}^{n\times m}, \mathbf{b} \in \mathbb{R}^{n\times 1}$ and $\mathbf{x} \in \mathbb{R}^{m\times 1}$ indicate the dictionary, the input signal and sparse representation of the input signal, respectively. P_0 can be rewritten as

$$(P_0): \min_{\mathbf{x}} \|\mathbf{Ax}-\mathbf{b}\|_2 \text{s.t} \|\mathbf{x}\|_0 \leq k. \quad (2)$$

In this form, the representation error is minimized with respect to a specified maximum number of nonzero elements in $\mathbf{x}$, i.e. k.

▷ Unfortunately, the problem is NP-hard and its approximate solutions can be obtained through Greedy algorithms.

▷ These algorithms iteratively select the most correlated atom of a basis dictionary with the residual, then updates the residual. The algorithm is stopped whenever the signal to residual error gets less than a predefined value.

▷ Orthogonal matching pursuit (OMP) [2] is used as solution

■ A fundamental element in a sparse representation problem is selecting a suitable dictionary.

▷ Pre-constructed dictionaries such as Discrete Cosine Transform (DCT).

▷ Dictionary can be learned from a training dataset or input data. By this approach, a redundant or an overcomplete dictionary is adapted to the input data type, K-SVD[2] is used.

III. Modeling Human Cochlea

■ Experimentally a **Gammatone (GT)** filter bank is attained whose functions are defined by

$$G_{f_c}(t) = t^{N-1}\exp\left(-2\pi b(f_c)\right)\cos(2\pi f_c t + \phi)u(t), \quad (3)$$

where t, $u(t)$, N, f_c and ϕ are time index, unit step function, order, center frequency and phase of each GT filter, respectively. The equal rectangular bandwidth of each GT filter, ERB, is defined as

$$ERB(f_c) = (1+4.3\times 10^{-3}\times f_c), b(f_c) = 1.019ERB(f_c). \quad (4)$$

IV. Sparse Representation of Human Auditory System

■ The proposed sparse models for human auditory systems

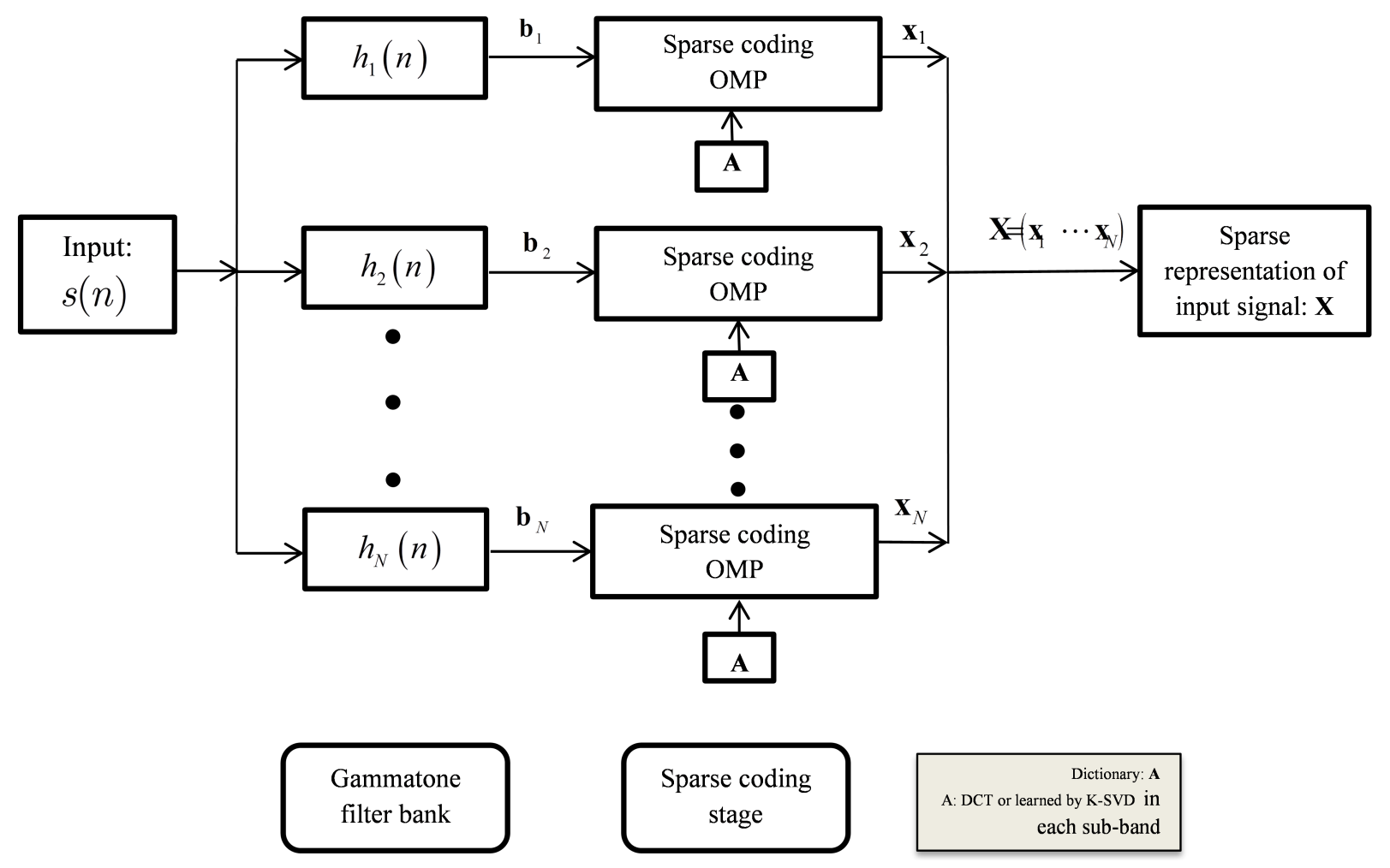

Fig.1 *Block diagram of model (I)*

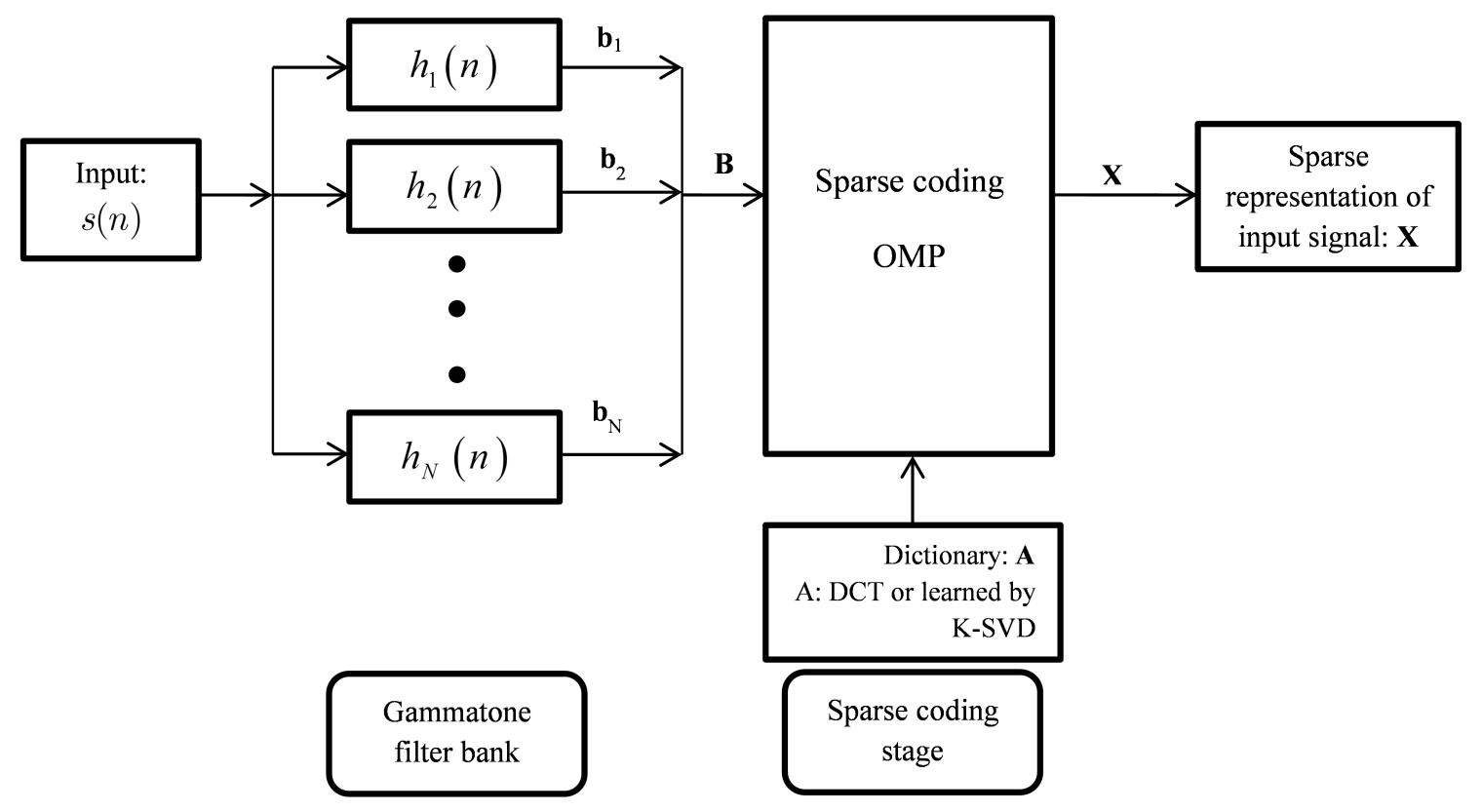

Fig.2 *Block diagram of modell(II)*

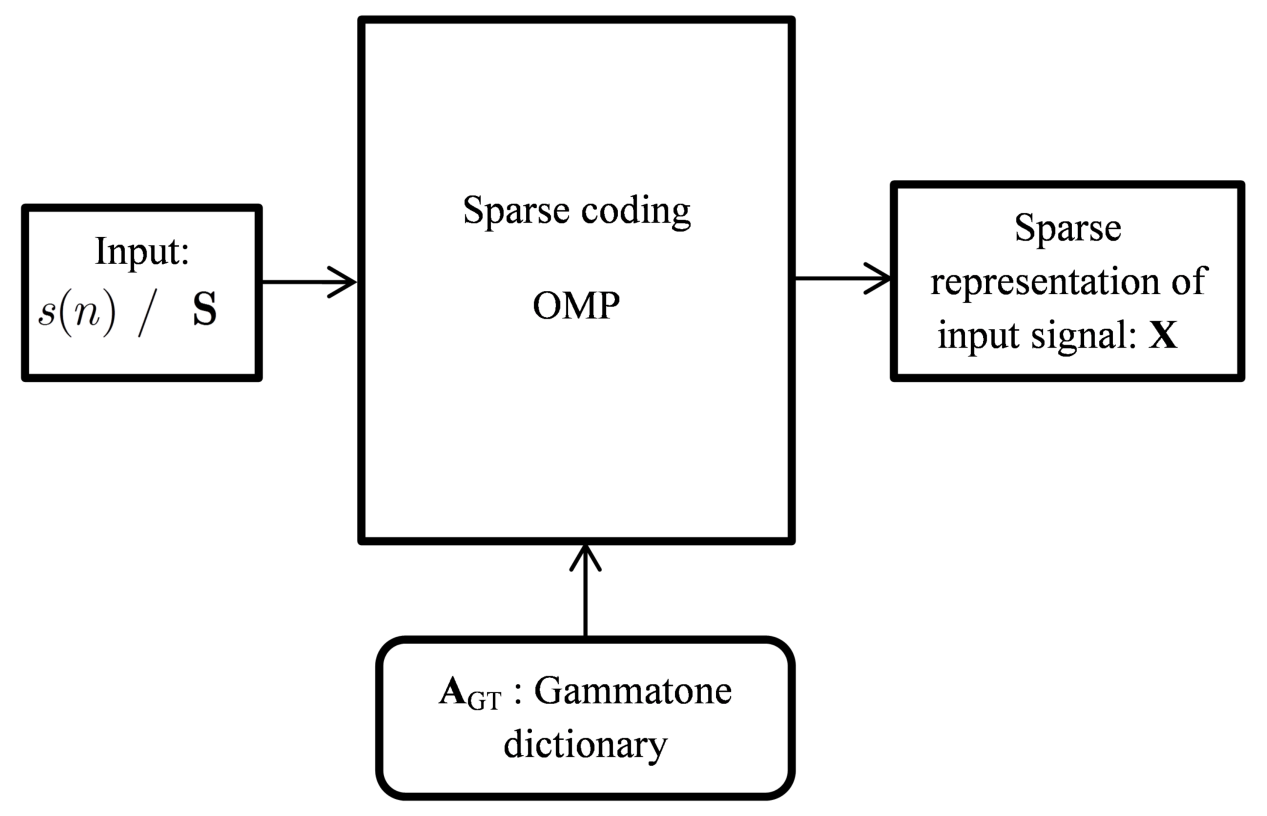
Input:
$s(n)$ / S
Sparse coding
OMP
Sparse representation of input signal: X
$\mathbf{A}_{GT}$: Gammatone dictionary

Fig.3 *Block diagram of model(III)*

■ **Model (I)-Sub-band sparse representation**:

- It assumes that the auditory system processes the outputs of cochlea filters independently to generate sparse codes, as shown in Fig.(1).
- Indeed, hair-cells of each channel generate spike codes independently to the other channels.

■ **Model (II)-Spectro-temporal sparse representation**:

- It assumes that the haircells in the inner ear of the auditory system uses the output processed information of all cochlea filters together to generate spike codes, as shown in Fig.(2).
- Sparse coding problem is solved through finding a sparse linear combination of these spectro-temporal elements.

■ **Model (III)-Gammatone dictionary as the basis dictionary for sparse representation**:

- Time-frequency shifts of GT filters are considered as a basis dictionary matrix for sparse representation, as shown in Fig.(3).
- Our method differs in two ways from the one in [3]. First, as this basis matrix becomes huge, the time shifts are selected randomly or some of the least important time shifts are omitted. In our experiments both approaches are tested that lead to similar results. Second, OMP sparse representation algorithm is employed instead of MP.

V. Experiments

■ Performance measures

▷ Reconstruction Error

$$\delta = \left\|\mathbf{x_{rec}} - \mathbf{x_{orig}}\right\|_2, \quad (5)$$

where $\mathbf{x_{rec}}$ and $\mathbf{x_{orig}}$ are the normalized reconstructed signal and the normalized original signal, respectively.

▷ Perceptual Evaluation of Speech Quality (PESQ)

$*$ As this measure uses an auditory model in the core of its qualification process and also it is able to measure the quality near to subjective quality measures, quality of the reconstructed speech signals is measured by using PESQ measure.

■ Experiments setup

▷ Approximately three minutes of randomly selected clean speech files from TIMIT database are used. Also, white and pink noise signals from NOISEX-92 database are used to make the noisy speech signals at the SNR levels of -5, 0, 5, and 10 dB.

VI. Results

■ Clean Speech signal

– PESQ scores for the reconstructed clean speech signals using three models, OMP is used for sparse coding.

Model	PESQ	Reconstruction error
Model (I)	3.5	15×10^{-4}
Model (II)	3.8	15×10^{-4}
Model (III)	3.35	20×10^{-4}

– PESQ scores for the reconstructed clean speech signals using three models, OMP is used for sparse representation of input signal over a learned dictionary by K-SVD.

Model	PESQ	Reconstruction error
Model (I)	3.8	8×10^{-5}
Model (II)	4.1	6×10^{-5}
Model (III)	3.9	16×10^{-4}

■ Noisy speech signal

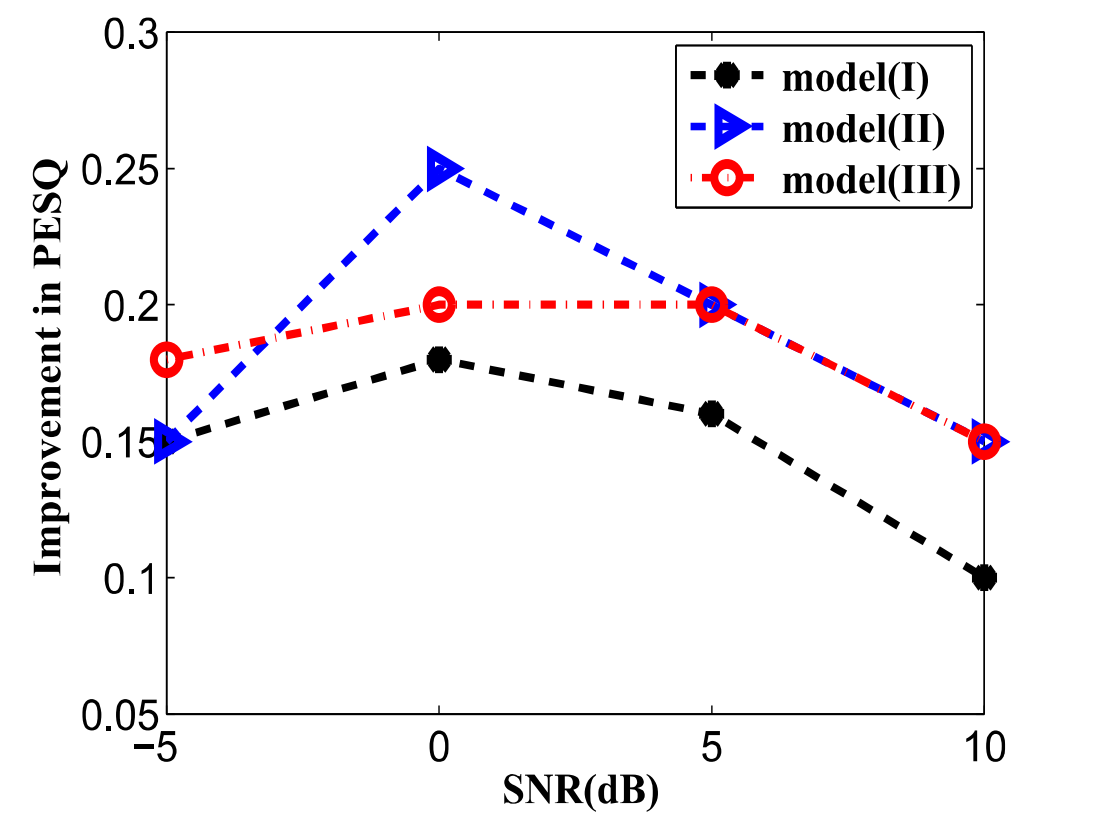

Fig. 4: *PESQ value improvement for the reconstructed speech signals in additive white Gaussian noise case. OMP is used in the sparse coding stage.*

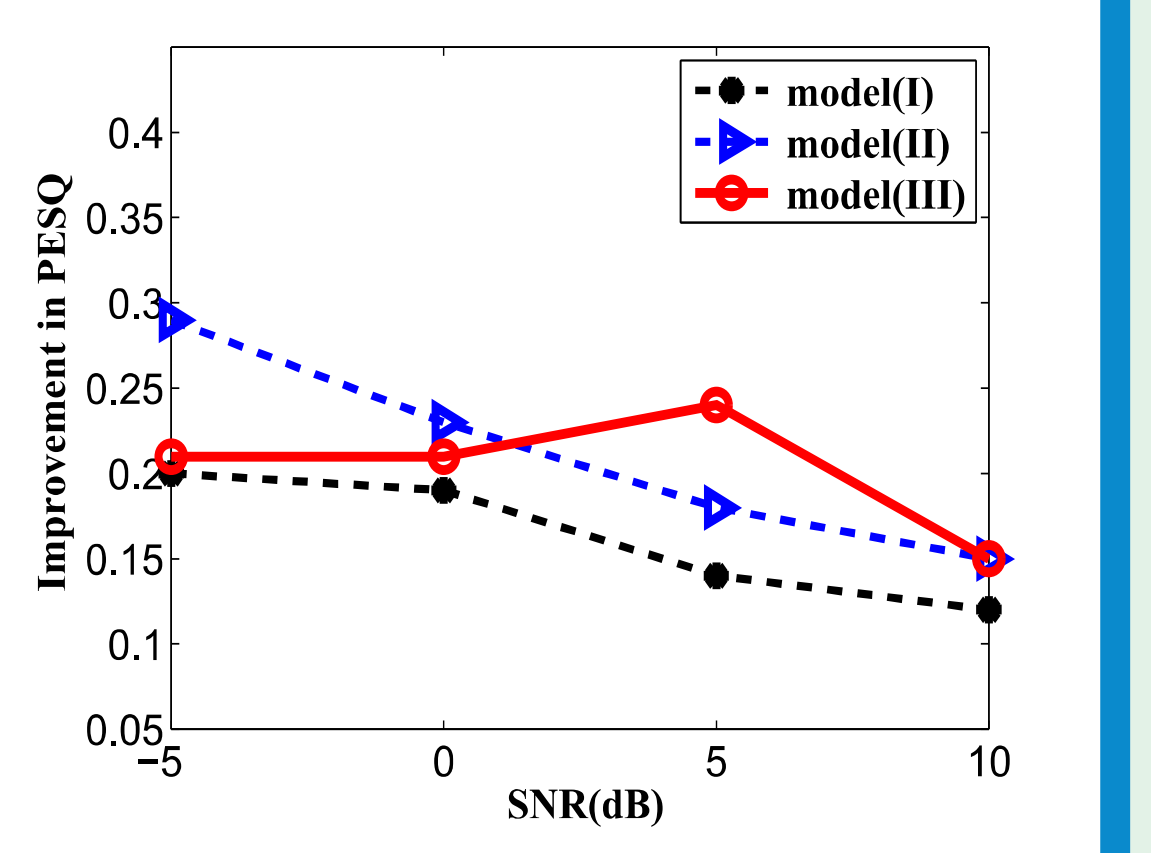

Fig. 6: *PESQ value improvement for the reconstructed speech signals in additive pink noise case. OMP is used in the sparse coding stage.*

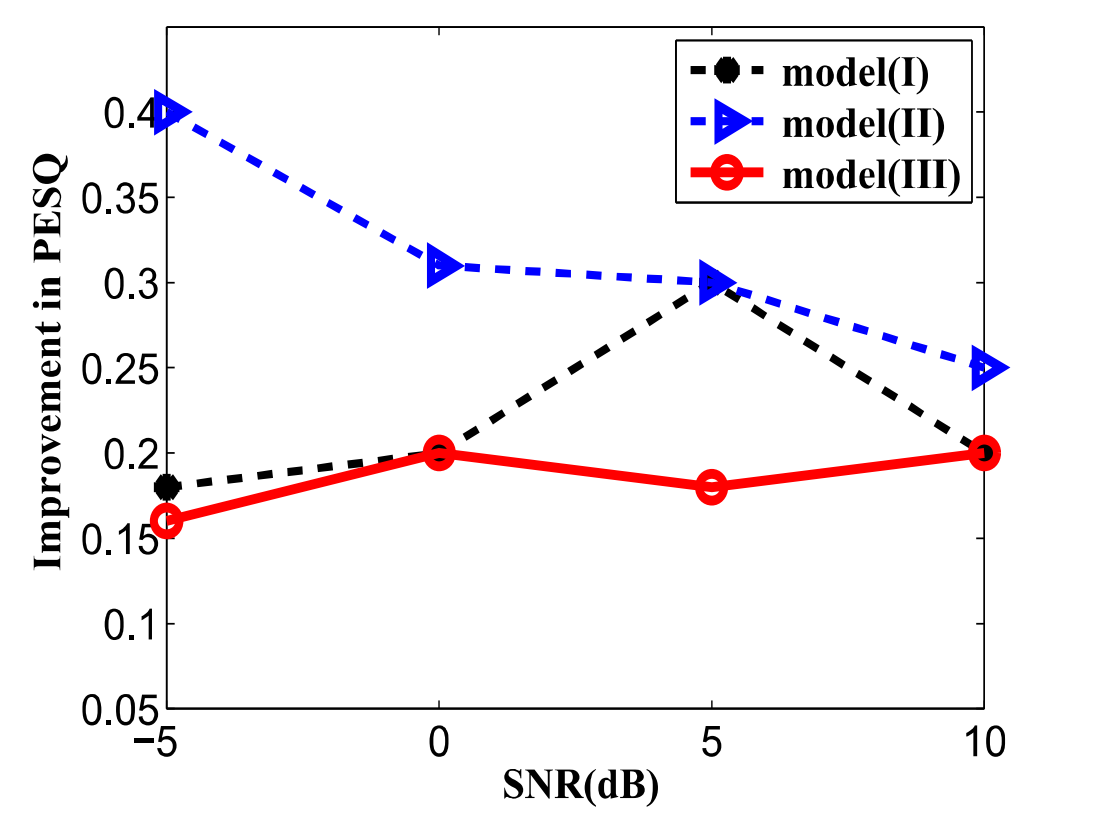

Fig. 5: *PESQ value improvement for the reconstructed speech signals in additive white Gaussian noise case.OMP is used over a learned dictionary by K-SVD in the sparse coding stage.*

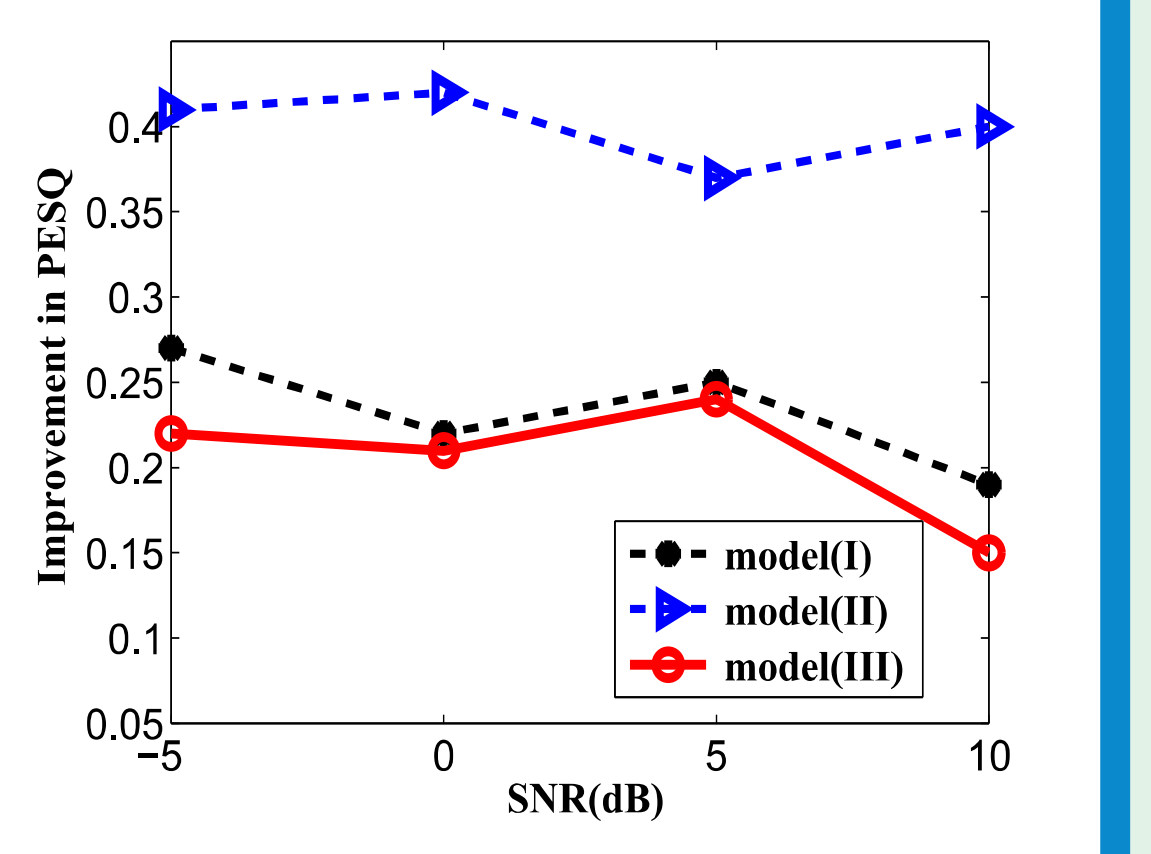

Fig. 7: *PESQ value improvement for the reconstructed speech signals in additive pink noise case. OMP is used over a learned dictionary by K-SVD in the sparse coding stage.*

VII. Conclusion

- All three sparse models have good performance in reconstructing clean speech signal.
- Three proposed models in noisy conditions lead to improvement in quality of the reconstructed noisy speech signals.
- Experiments show that these models represent good sparse models for the human peripheral auditory system.

VIII. References

[1] E. C. Smith and M. S. Lewicki, “Efficient auditory coding,” *Nature*, vol. 439, pp. 978–982, February 2006.

[2] M. Elad, *Sparse and redundant representation from theory to application*, Springer-Verlag New York, 2010.

[3] R. Pichevar, H. Najaf-Zadeh, L. Thibault, and H. Lahdili, “Auditory-inspired sparse representation of audio signals,” *Speech Communication*, vol. 53, pp. 643–657, September 2011.