Speech Denoising using Common Vector Analysis in Frequency Domain

Erol SEKE*[1], Mehmet Hakan DURAK[2], Kemal OZKAN[3]

Abstract: Signal denoising approaches on data of any dimension largely relies on the assumption that data and the noise components and the noise itself are somewhat uncorrelated. However, any denoising process heavily depending on this assumption retreats when the signal component is adversely affected by the operation. Therefore, several proposed algorithms try to separate the data into two or more parts with varying noise levels so that denoising process can be applied on them with different parameters and constraints. In this paper, the proposed method separates the speech data into magnitude and phase where the magnitude part is further separated into common and difference parts using common vector analysis. It is assumed that the noise largely resides on difference part and therefore denoised by a known algorithm. The speech data is reconstructed by combining common, difference and phase parts. Using Linear Minimum Mean Square Error Estimation algorithm on the difference part, excellent denoising results are obtained. Results are compared with that of the state of the art in well-known speech quality measures.

Keywords: speech, denoising, Common Vector Analysis, CVA, subspace, frequency domain.

1. Introduction

For decades, as cognitive science penetrated into automation systems, voice automated applications like voice directed banking, voice signatures, intelligent homes; voice recognizing mobile-phone apps and such became both possible and popular. All voice applications somewhat necessitate high quality voice signals, mostly in digital form, impetrating voice denoising algorithms. All naturally collected signals carry some noise energy weather from electronics/transmission or unrelated background signals. Voice denoising aims to improve signal quality, voice intelligibility or do both, achieving that with minimal loss in signal energy.

Although the voice denoising methods can be separated into two as single and multiple channel algorithms, researchers mostly focused on single channel because of incomparably higher number of encounters of such. Spectral subtraction, noise estimation, Wiener deconvolution/filtering, statistical and subspace based methods are considered as mainly single channel, notwithstanding the fact that they can be and are also employed in multichannel systems. In this paper, a method based on common vector analysis (CVA) is proposed. Therefore, brief methodology of other well-known subspace is given for comparisons.

Subspace methods rely on the expectation that the noisy data can be separated into two or more component where noise can be handled more efficiently within. A Singular Value Decomposition (SVD) based approach, proposed by Dendrinos et al. [1], uses the expectation that, after factorization of the data into sub-data, noise energy concentrates in vectors corresponding to smaller singular values. In the simplest denoising approach

these are zeroed and voice data is recomposed. This technique is improved by Jensen et al. [2] for colored noise on which the former method has somewhat failed to reduce. On the other hand, their method with high computational complexity had several constraints for controlling residual noise. Ephraim et al. [3] aimed to optimize the estimator that minimizes distortion caused by residual noise. Noisy signal is separated into noise and signal subspaces using Karhunen Loeve Transform (KLT) whereby zeroing the components in noise subspace and restructuring the signal subspace using a gain function. Components in subspaces are recombined again to obtain denoised signal through inverse KLT. Mittal et al. [4] and Rezayee et al. [5] expanded this work for colored noise. They obtained better results by employing different KLT matrices and converging covariance matrix of the noise vectors to a diagonal matrix respectively.

Common Vector Approach (CVA) is a subspace method used in recognition applications. In CVA, training data representing each subject to be discriminated are used to form its own class. In a speech recognition application, ambient noise, ages and genders of speakers result in differences within a class [6]. CVA is based on the common component of those, basically by removing these differences within the class. This component is called the common vector. It has been employed in speaker recognition [7], speech recognition [8]-[10], face recognition [11], fault detection in electrical motors [12], spam e-mail detection [13]. CVA has also been used in image denoising [14].

2. Common Vector Approach

When differences between feature vectors in a class are removed, the remaining vector which consists of features invariant within the class is called the common vector. A feature vector is then, presumed to be a sum of common and difference components. If the number of feature vectors (m) is greater than the dimension (n) of the vectors, then this is called a *sufficient case* ($m > n$). In the sufficient case, the common vector is the mean vector. The *insufficient case* occurs when the vector dimension is larger than or equal to the number of vectors ($n \geq m$), which is the case in most practical applications where, for example, too few image blocks with many pixels each exist. In general, it covers the setups where number of samples is less than the sample

[1] *Electrical and Electronics Engineering Department, Engineering and Architecture Faculty, Eskişehir Osmangazi University, 26480, Eskişehir/Turkey*
[2] *Electrical Electronics Engineering Department, Engineering Faculty, Gazi University, 06570, Ankara/Turkey*
[3] *Computer Engineering Department, Engineering and Architecture Faculty, Eskişehir Osmangazi University, 26480, Eskişehir/Turkey*
* *Corresponding Author: Email: eseke@ogu.edu.tr*

dimension. Since we are interested in such setups, only insufficient case of CVA is described in the following.

Let the feature vectors in a pattern class be linearly independent $a_1, a_2, ..., a_m$. Each vector can be written as

$$a_i = a_{i,diff} + a_{comm}, \quad i = 1, 2, ..., m \quad (1)$$

where a_{comm} and $a_{i,diff}$ are common vector of the class and difference vector per a_i respectively. The covariance matrix for the class is

$$\Phi = \sum_{i=1}^{m} (a_i - a_{avg})(a_i - a_{avg})^T \quad (2)$$

where $a_{avg} = \frac{1}{m}\sum_{i=1}^{m} a_i$ denotes the mean of the vectors in the class. In insufficient case ($n \geq m$), $n - m + 1$ of the eigenvalues will be zero and corresponding eigenvectors (u_j) span the indifference subspace $B^\perp$, while the remaining eigenvectors span the difference subspace B where B and $B^\perp$ are orthogonal. The common vector of the class can be found by projecting any feature vector onto the indifference subspace $B^\perp$ using

$$a_{com} = P^\perp a_i, \quad i = 1, 2, ..., m. \quad (3)$$

The projection matrix $P^\perp$ is calculated using eigenvectors u_j that correspond zero valued eigenvalues (spans $B^\perp$) via

$$P^\perp = \sum_{j=1}^{n-m+1} u_j u_j^T. \quad (4)$$

The subspace methods other than CVA requires the inverse of the covariance matrix Φ [15]. However, $n \geq m$ inhibits the calculation of inverse of Φ whereas CVA does not have this problem.

It is expected that noise mainly reside within the difference components when it is uncorrelated between class vectors. Therefore, it is imperative to construct classes as correctly as possible. When class information of the vectors is not available, classes should be constructed by collecting similar vectors into a data set matrix for each evaluated vector whose common vector is to be found. When the input data is a stream or can be handled as a stream with bounds, for example, classes can be constructed by searching similar vectors within a reasonable time range.

Since the raw vectors in speech data are selected to be sample frames of length n, the words *vector* and *frame* are used interchangeably used within this paper.

3. Proposed Algorithm

The denoising algorithm proposed in this paper relies on the intuition that the spectral content of the speech does not abrubtly change and changes are mostly noise related. Approach is similar to time averaging of Fast Fourier Transform (FFT) data in digital spectrum analyzers. On the other hand, the overlapping ratio is the highest in the proposed algorithm. As illustrated in the Fig. 1, frames are picked from original speech data stream by a sliding Hamming window of width n

$$w_i = 0.54 - 0.46\cos(2\pi i/(n-1)), i = 1, 2, ..., n \quad (5)$$

which slides 1 sample for each subsequent frame. Although not required, it is logical to select n to correspond approximately 4 ms speech data since spectral characteristics of speech may

greatly change for longer intervals. We conducted some tests for determining optimal frame length, as explained in the following section and concluded that 4 ms is adequate.

For each frame to be denoised, a class is constructed by picking m most spectrally similar Hamming windowed frames within its neighbourhood. Obviously entire data stream can be used for selecting the frames and constructing the class. In our experiments, it is determined that a neighbourhood size that contains $2n - 1$ frames, x_i', including the current frame to be denoised, is sufficient for both required number of vectors for the class and reasonable computational complexity. Fast Fourier Transform (FFT) is applied on these $2n - 1$ frames and their magnitude and phase components are separated as

$$b_k = |FFT\{x_k'\}|$$
$$p_k = \langle FFT\{x_k'\} \rangle \quad . \quad (5)$$

$m - 1$ magnitude frames that are most similar to the magnitude frame of the current frame (the one to be denoised) are picked and the class is constructed with a total of m frames. When distances to the current magnitude frame calculated using the Euclidean distance

$$d_k = \sqrt{\sum_{i=1}^{n} |b_{k,i} - b_{cur,i}|^2} \quad (6)$$

where $b_{cur,i}$ and $b_{k,i}$ are i th dimension components of the current and k th magnitude frames. m frames with the smallest d_k are selected into the class member set A. Since b_{cur} would have zero distance to itself, it will be assigned index 1 and called a_1 as indicated in Fig. 1.

m frames with dimensions n leads to an insufficient case CVA with the most number of vectors. We noticed that, removal of few class outliers did not have considerable effect on the results, therefore, keeping that option in possibilities, an outlier removal operation is not performed as this would increase complexity.

A CVA operation is performed on the set A via

$$(a_{com}, a_{1,dif}, a_{avg}) = CVA(A, a_1) \quad (7)$$

as described in the previous section. a_{avg} should be added back after the denoising.

Noise largely resides in difference components ($a_{1,dif}$). Therefore, common component (a_{com}) is kept and difference component of the current frame is denoised using a denoising algorithm that involves Principal Component Analysis (PCA) [15]. In fact, any denoising algorithm can be effectively used on $a_{1,dif}$ since large portion of the signal energy is still in the common component, which is considered almost noise-free. After denoising $a_{1,dif}$ and obtaining $\overline{a}_{1,dif}$, the denoised difference frame, current denoised magnitude frame is reconstructed via $\overline{a}_1 = a_{com} + \overline{a}_{1,dif}$. Time domain speech frame is reconstructed by adding phase information and applying inverse FFT.

After applying the described algorithm on each frame, denoised time frames are combined to build the denoised speech data. Since the frames are overlapping, there are several options at the recombining stage. Just adding them onto the appropriate time location is one of them. Here a weighting window can be used to increase the weight of the center of the frame. In our experiments, we noted that just adding the frames (flat window) is sufficient and have least complexity.
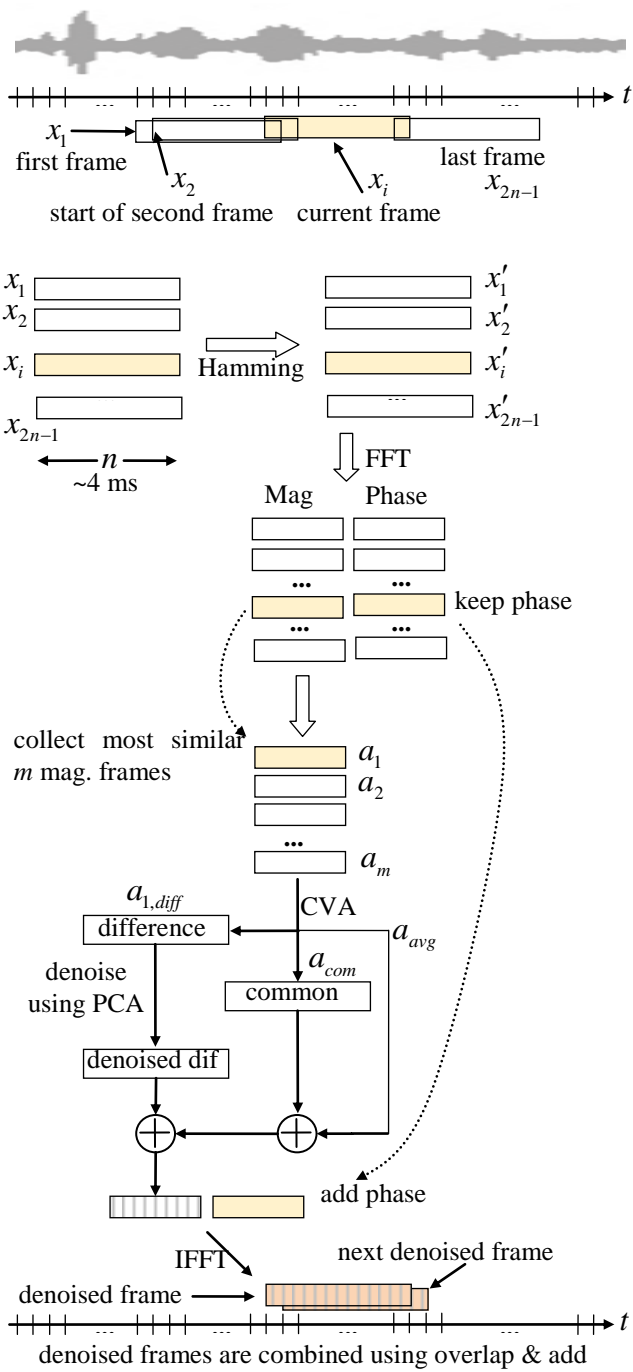
Figure 1. Flow of the proposed algorithm.

It should be noted here that there are several parameters that can be optimized for the best performance on the speech data to be denoised; m (number of frames in classes), n (frame size), PCA parameters, recombining options. However, since we intended an algorithm that requires no data dependent optimisation parameters, these optimizations are performed for a large training speech data set and best logical parameter set is kept for all.

4. Experiments

For the experimental work on the proposed algorithm, NOISEUS (Hu, and Loizou, 2007) speech database is used. NOISEUS is composed of 30 English sentences spoken by 3 male and 3 female speakers. Recordings are sampled by 8 kHz 16 bits with approximately 2 seconds in length. 8 different noise type (airport, crowd, car, exhibition hall, restaurant, train station, street and train) are added onto each speech data to obtain 4 SNR levels (0 dB, 5 dB, 10 dB, 15 dB). Noise data is itself taken from AURORA database. In addition, the database is extended by adding 4 levels of white noise onto the data.
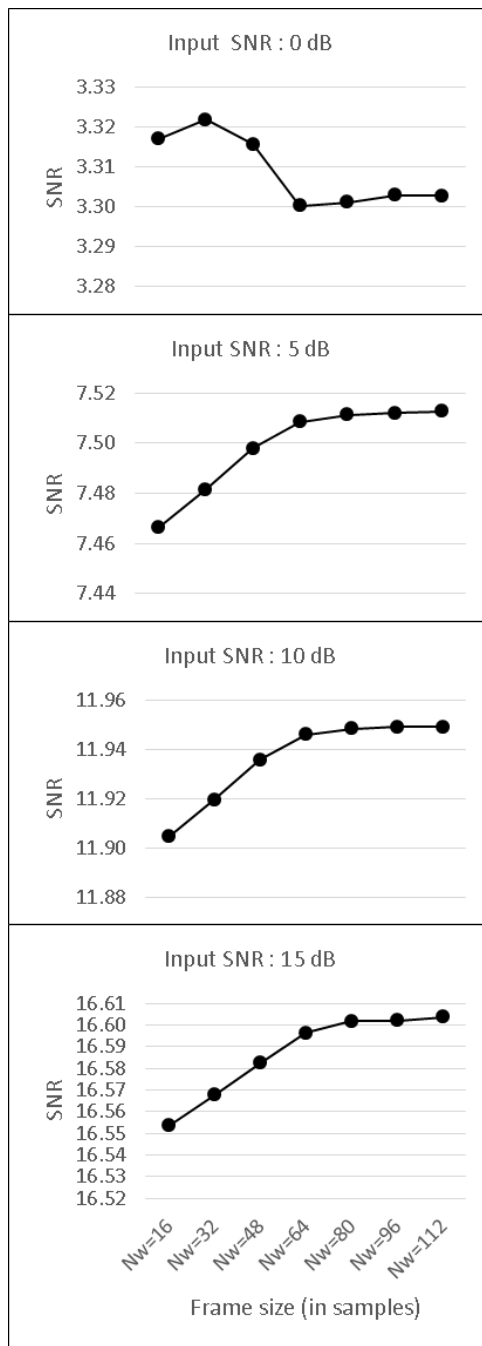
Figure 2. Initial test results; Input/Output SNRs for different frame sizes.

Initial experiments are conducted to determine best or reasonable parameter values for CVA. These are; frame size, overlap ratio and neighbourhood size from which the class member candidates are picked. Fig. 2 shows the performance graphics for various frame sizes and input noise levels. From these tests, it is determined that frame size of 40 samples (corresponding to 4 ms) and highest overlap ratio are adequate for both performance and

complexity. Neighbourhood size tests, on the other hand, were inconclusive for widths greater than three frame sizes. It is seen that algorithm becomes data dependent for larger search areas. In the following tests, we used $N_W = 3N$ as the neighbourhood size, where N is the frame size in samples.

Speech data with 9 different background noise added at 4 levels are denoised using the proposed CVA algorithm with previously determined parameters. The results are compared against 5 state-of-the-art methods found in the literature;

    1. Perceptually motivated subspace algorithm [16], will shortly be called as *sub* from now on.

    2. Psychoacoustically motivated statistical method [17], will be called as *stat*.

    3. Wiener filtering algorithm based on wavelet thresholding multi-taper spectra [18], will be called as *wien*.

    4. A variant of minimum controlled recursive average algorithm [19], will be named as *rec*.

    5. Continuous spectral tracking [20], will be named as *spec* in the following sections.

Performance measures used in comparisons are Perceptual Evaluation of Speech Quality (PESQ), Log Likelihood Ratio (LLR) and Euclidean Distance in Cepstral Domain (CEP), most used measures in the literature.

1) *Perceptual Evaluation of Speech Quality (PESQ)*

Standardized by ITU-T P.862 (02/01), PESQ is a commonly used in telecommunication. It is defined as a linear combination

$$PESQ = a_0 + a_1 D_{avg} + a_2 A_{avg} \qquad (8)$$

of the average disturbance value D_{avg} and the average asymmetrical disturbance values A_{avg} where $a_0 = 4.5$, $a_1 = -01$ and $a_2 = -0.0309$ are the optimized for speech processed through networks [21]. Table 1 gives the results for CVA along with *sub* [16], *stat* [17], *wien* [18], *rec* [19], *spec* [20] in PESQ measure. Best PESQ values are marked with boldfaced characters. Each number in the table is the average of PESQ values for 180 recordings (30 sentences spoken by 6 individuals).

2) *Log-Likelihood Ratio (LLR)*

In LLR, distorted and denoised data are compared statistically [22] and is defined as

$$d_{LLR}\left(c_x, c_{\hat{x}}\right) = \log\left(\frac{c_x R_{\hat{x}} c_x^T}{c_{\hat{x}} R_{\hat{x}} c_{\hat{x}}^T}\right). \qquad (9)$$

Here c_x and $c_{\hat{x}}$ are Linear Prediction Coefficient (LPC) vectors of the distorted and denoised speech data respectively. $R_{\hat{x}}$ is the autocorrelation matrix of the distorted speech signal [23]. Lower LLR values mean higher quality speech signal. In Table II, LLR values for CVA and other 5 methods are compared with boldfaces indicating the best/lowest LLR value for a test input. It is notable that CVA is superior to the compared methods since it generated the lowest LLR for all background noise tests. However, for white noise cases CVA failed to be the best even though the scores are close. CVA performed best among all methods in all PESQ tests except 5. That indicates about 83% success. In white noise cases, although not the best, CVA performed close to the best.

3) *Cepstral Distance Measure (CEP)*

CEP too is based on LPCs and is defined as the distance between LPCs of original and enhanced speech frames as

$$d_{CEP}\left(c_x, c_{\hat{x}}\right) = 6.1421 \sqrt{\sum_{k=1}^{N}(c_{x,k} - c_{\hat{x},k})^2} \ . \qquad (10)$$

where N is the dimension of LPCs. Lower CEP values indicate higher speech quality [24]. As shown in Table III, proposed CVA method is superior against all other denoising methods for all background noise types and levels applied in tests, except for white noise cases.

TABLE I
COMPARISON OF 6 METHODS ON PESQ MEASURE

		sub	stat	wien	rec	spec	cva
airport	0 dB	1.43	1.31	1.40	1.24	1.32	**1.69**
	5 dB	1.87	1.68	1.84	1.77	1.82	**2.05**
	10 dB	2.23	2.09	2.24	2.20	2.26	**2.40**
	15 dB	2.63	2.59	2.68	2.72	2.73	**2.74**
crowd	0 dB	1.37	1.20	1.34	1.13	1.34	**1.64**
	5 dB	1.84	1.73	1.82	1.62	1.78	**2.02**
	10 dB	2.23	2.16	2.25	2.24	2.24	**2.39**
	15 dB	2.63	2.60	2.65	2.69	2.70	**2.74**
car	0 dB	1.46	1.27	1.41	1.23	1.40	**1.68**
	5 dB	1.91	1.92	1.89	1.67	1.86	**1.97**
	10 dB	2.24	2.25	2.25	2.17	2.28	**2.34**
	15 dB	2.57	2.70	2.60	2.66	**2.71**	2.70
ex. hall	0 dB	1.42	1.29	1.39	1.15	1.31	**1.62**
	5 dB	1.91	1.89	1.89	1.68	1.85	**1.98**
	10 dB	2.25	2.17	2.25	2.21	2.28	**2.32**
	15 dB	2.62	2.61	2.60	2.64	2.64	**2.67**
restrnt	0 dB	1.37	1.24	1.42	1.21	1.26	**1.70**
	5 dB	1.84	1.74	1.83	1.71	1.80	**2.02**
	10 dB	2.23	2.08	2.22	2.22	2.22	**2.41**
	15 dB	2.60	2.57	2.64	2.67	2.63	**2.72**
station	0 dB	1.44	1.35	1.41	1.20	1.36	**1.66**
	5 dB	1.88	1.87	1.89	1.75	1.83	**2.07**
	10 dB	2.24	2.25	2.27	2.24	2.31	**2.38**
	15 dB	2.53	2.61	2.59	2.65	2.66	**2.72**
street	0 dB	1.32	1.18	1.24	1.04	1.24	**1.62**
	5 dB	1.79	1.68	1.78	1.62	1.76	**2.00**
	10 dB	2.23	2.13	2.21	2.19	2.21	**2.38**
	15 dB	2.57	2.59	2.57	2.60	2.56	**2.68**
train	0 dB	1.22	1.10	1.16	0.81	1.09	**1.57**
	5 dB	1.86	1.67	1.75	1.42	1.65	**1.92**
	10 dB	2.28	2.14	2.21	2.07	2.19	**2.29**
	15 dB	2.64	2.63	2.62	2.64	2.63	**2.65**
white	0 dB	**1.77**	1.76	1.75	1.23	1.73	1.70
	5 dB	**2.17**	**2.17**	2.16	1.77	2.16	2.06
	10 dB	2.44	2.55	2.47	2.31	**2.57**	2.42
	15 dB	2.75	**2.95**	2.80	2.76	2.87	2.75

5. Conclusions

Tests conducted on 30 sentences spoken by 6 individuals with added 4 levels of 8 structured background noise recordings (total of 5760 recordings per method per quality measure) let us safely conclude that the proposed CVA method is superior against other 5 methods. In additional tests using white noise, on the other hand, CVA has failed to be the best (total of 720 recordings per method per quality measure). However, in most of the tests that CVA was not the best, its scores were close to the best.

Acknowledgements

TABLE III
COMPARISON OF 6 METHODS ON LLR MEASURE

		sub	stat	wien	rec	spec	cva
airport	0 dB	1.29	1.09	1.37	1.03	0.96	**0.85**
	5 dB	1.10	0.92	1.16	0.86	0.83	**0.66**
	10 dB	0.86	0.70	0.94	0.67	0.69	**0.49**
	15 dB	0.66	0.56	0.72	0.54	0.56	**0.35**
crowd	0 dB	1.30	1.14	1.35	1.08	1.03	**0.89**
	5 dB	1.10	0.92	1.17	0.90	0.88	**0.70**
	10 dB	0.88	0.71	0.93	0.69	0.72	**0.50**
	15 dB	0.67	0.56	0.72	0.55	0.59	**0.36**
car	0 dB	1.35	1.10	1.47	1.08	1.03	**0.94**
	5 dB	1.11	0.89	1.22	0.86	0.86	**0.73**
	10 dB	0.98	0.75	1.10	0.67	0.71	**0.54**
	15 dB	0.71	0.57	0.82	0.51	0.57	**0.37**
ex.hall	0 dB	1.34	1.19	1.52	1.17	1.14	**1.06**
	5 dB	1.13	0.98	1.31	0.98	0.97	**0.85**
	10 dB	1.02	0.86	1.19	0.78	0.81	**0.64**
	15 dB	0.81	0.74	1.01	0.61	0.66	**0.47**
restrnt	0 dB	1.35	1.12	1.39	1.02	0.98	**0.85**
	5 dB	1.04	0.88	1.08	0.84	0.82	**0.66**
	10 dB	0.88	0.73	0.94	0.68	0.68	**0.49**
	15 dB	0.70	0.60	0.77	0.56	0.60	**0.36**
station	0 dB	1.28	1.10	1.39	1.06	1.00	**0.89**
	5 dB	1.08	0.87	1.21	0.84	0.83	**0.68**
	10 dB	0.89	0.72	1.02	0.66	0.69	**0.50**
	15 dB	0.72	0.59	0.83	0.52	0.56	**0.35**
street	0 dB	1.39	1.19	1.48	1.09	1.07	**0.94**
	5 dB	1.16	0.99	1.28	0.89	0.90	**0.75**
	10 dB	0.93	0.78	1.05	0.72	0.75	**0.57**
	15 dB	0.78	0.68	0.94	0.59	0.63	**0.42**
train	0 dB	1.41	1.18	1.43	1.14	1.14	**1.05**
	5 dB	1.20	1.06	1.28	0.99	1.01	**0.88**
	10 dB	1.01	0.86	1.10	0.82	0.86	**0.68**
	15 dB	0.80	0.66	0.87	0.64	0.69	**0.50**
white	0 dB	1.44	1.50	1.64	**1.36**	1.49	1.43
	5 dB	1.30	1.27	1.52	**1.22**	1.37	1.26
	10 dB	1.11	1.12	1.36	**1.03**	1.23	1.07
	15 dB	0.91	0.96	1.17	**0.84**	1.06	0.88

TABLE IIIII
COMPARISON OF 6 METHODS ON CEP MEASURE

		sub	stat	wien	rec	spec	cva
airport	0 dB	7,27	6,29	7,44	6,11	5,85	**5,27**
	5 dB	6,47	5,57	6,54	5,42	5,27	**4,52**
	10 dB	5,39	4,62	5,55	4,59	4,62	**3,73**
	15 dB	4,41	3,90	4,45	3,91	4,02	**3,02**
crowd	0 dB	7,26	6,47	7,32	6,22	6,04	**5,37**
	5 dB	6,43	5,55	6,58	5,50	5,42	**4,60**
	10 dB	5,47	4,63	5,50	4,61	4,69	**3,76**
	15 dB	4,43	3,89	4,47	3,92	4,08	**3,00**
car	0 dB	7,46	6,23	7,77	6,15	5,96	**5,43**
	5 dB	6,49	5,40	6,74	5,30	5,31	**4,66**
	10 dB	5,92	4,76	6,19	4,50	4,62	**3,87**
	15 dB	4,63	3,90	4,88	3,74	3,94	**3,06**
ex.hall	0 dB	7,44	6,59	8,02	6,62	6,43	**5,95**
	5 dB	6,60	5,76	7,12	5,87	5,78	**5,16**
	10 dB	6,08	5,23	6,61	5,03	5,09	**4,33**
	15 dB	5,11	4,61	5,70	4,21	4,39	**3,53**
restrnt	0 dB	7,50	6,45	7,58	6,07	5,89	**5,27**
	5 dB	6,19	5,42	6,21	5,29	5,18	**4,52**
	10 dB	5,46	4,74	5,55	4,58	4,57	**3,73**
	15 dB	4,57	4,06	4,66	4,05	4,16	**3,08**
station	0 dB	7,20	6,27	7,46	6,14	5,91	**5,35**
	5 dB	6,37	5,28	6,68	5,25	5,18	**4,51**
	10 dB	5,49	4,63	5,83	4,48	4,55	**3,75**
	15 dB	4,65	3,99	4,93	3,81	3,95	**3,00**
street	0 dB	7,63	6,65	7,89	6,35	6,24	**5,62**
	5 dB	6,65	5,82	7,00	5,49	5,52	**4,84**
	10 dB	5,69	4,92	6,00	4,78	4,85	**4,08**
	15 dB	4,99	4,44	5,45	4,15	4,29	**3,37**
train	0 dB	7,65	6,60	7,73	6,53	6,51	**6,09**
	5 dB	6,78	6,19	7,11	5,98	6,02	**5,45**
	10 dB	5,99	5,31	6,26	5,25	5,37	**4,63**
	15 dB	5,05	4,47	5,26	4,43	4,60	**3,81**
white	0 dB	8,00	8,45	8,60	**7,78**	8,66	8,45
	5 dB	7,36	7,64	8,02	**7,32**	8,29	7,97
	10 dB	**6,57**	7,01	7,36	6,69	7,79	7,36
	15 dB	**5,65**	6,25	6,48	5,91	7,09	6,61

References

[1] M. Dendrinos, S. Bakamidis and G. Carayannis, "Speech enhancement from noise: a regenerative approach," Speech Comm., vol. 10 (1), pp. 45-57, 1991.

[2] S. H. Jensen, P. C. Hansen, S. D. Hansen and J. A Sorensen, "Reduction of broadband noise in speech by truncated QSVD," IEEE Trans. Speech Audio Processing, vol. 3, pp. 439-448, 1995.

[3] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech Enhancement," IEEE Trans. on Speech Audio Processing, 3, 251. 166, 1995.

[4] U. Mittal and N. Phamdo, "Signal/noise KLT based approach for enhancing speech degraded by colored noise," IEEE Trans. Speech and Audio Proc., vol. 8 (2), pp. 159-167, 2000.

[5] A. Rezayee and S. Gazor, "An Adaptive KLT Approach for Speech Enhancement," IEEE Trans. Speech and Audio Proc., vol. 9 (2), pp. 87-95, 2001.

[6] S. Ergin, "The Improvement and Recognition of the Noisy Speech Parameters," M. Eng. Thesis, Eskisehir Osmangazi University, Eskisehir, Turkey, 2004, [Online]. Available: http://ulusaltezmerkezi.com/gurultulu-ses-parametrelerinin-iyilestirilmesi-ve-taninmasi/ .

[7] M. B. Gülmezoglu. and A. Barkana, "Text-dependent speaker recognition by using Gram-Schmidt orthogonalization method," Proc. of IASTED Int. Conf. on Sig. Proc. & App., 438-440, 1998.

[8] M. B. Gülmezoglu, V. Dzhafarov, M. Keskin and A. Barkana, "A novel approach to isolated word recognition," IEEE Trans. on Acoustic Speech and Signal Processing, vol. 7(6), pp. 620-628, 1999.

[9] M.B.Gülmezoglu, V. Dzhafarov and A. Barkana, "The Common Vector Approach and its relation to Principal Component Analysis," IEEE Trans. On Speech and Audio Processing, vol. 9(6), pp. 655-662, 2001.

[10] M. B. Gülmezoglu, V. Dzhafarov and A. Barkana, "The Common Vector Approach and its Comparison with Other Subspace Methods in Case of Sufficient Data," Computer Speech and Language, vol. 21, pp. 266-281, 2007.

[11] H. Çevikalp, M. Neamtu, M. Wilkes and A. Barkana, "Discriminative Common Vectors for Face Recognition," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 27(1), pp. 4-13, 2005.

[12] M. B. Gülmezoglu and S. Ergin, "An Approach for Bearing Fault Detection in Electrical Motors," European Transactions on Electrical Power, vol. 17(6), pp. 628-641, 2007.

[13] S. Günal, S. Ergin, M. B. Gülmezoglu and Ö. N. Gerek, "On Feature Extraction for Spam E-Mail Detection," Lecture Notes in Computer Science, vol. 4105, pp. 635-642, 2006.

[14] K. Özkan and E. Seke, "Image Denoising Using Common Vector Approach," IET Image Processing, vol. 9(8), pp. 709-715, 2015.

[15] L. Zhang, W. S. Dong, D. Zhang, and G. M. Shi, "Two-stage image denoising by principal component analysis

with local pixel grouping," Pattern Recognition, vol. 43, pp. 1531-1549, 2010.

[16] F. Jabloun and B. Champagne, "Incorporating the human hearing properties in the signal subspace approach for speech enhancement," IEEE Trans. on Speech and Audio Processing, vol. 11(6), pp. 700-708, 2003.

[17] Y. Hu and P. Loizou, "Incorporating a psychoacoustical model in frequency domain speech enhancement," IEEE Signal Processing Letters, vol. 11(2), pp. 270-273, 2004.

[18] Y. Hu and P. Loizou, "Speech enhancement based on wavelet thresholding the multitaper spectrum," IEEE Trans. on Speech and Audio Processing, vol. 12(1), pp. 59-67, 2004.

[19] S. Rangachari and P. Loizou, "A noise estimation algorithm for highly nonstationary environments," Speech Communication, vol. 28, pp. 220-231, 2006.

[20] G. Doblinger, "Computationally Efficient Speech Enhancement By Spectral Minima Tracking in Subbands," Proc. EuroSpeech, vol. 2, p. 1513-1516, 1995.

[21] Y. Hu and P. Loizou, "Evaluation of objective quality measures for speech enhancement," IEEE Trans. Speech Audio Processsing, vol. 16 (1), pp. 229-238, 2008.

[22] S. Quackenbush, T. Barnwell and M. Clements, "Objective measures of speech quality," NJ: Prentice-Hall, Eaglewood Cliffs, 1988.

[23] J. H. L. Hansen and B. L. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms," Int'l. Conf. on Spoken Language Processing, vol. 7, pp. 2819–2822 , 1998.

[24] N. Kitawaki, H. Nagabuchi and K. Itoh, "Objective quality evaluation for low bitrate speech coding systems," IEEE J. Select. Areas in Comm., vol. 6(2), pp. 262-273, 1988.