SUBJECTIVE COMPARISON OF SPEECH ENHANCEMENT ALGORITHMS

Yi Hu and Philipos C. Loizou *

Department of Electrical Engineering
University of Texas at Dallas
Richardson, Texas 75083-0688
{yihuyxy, loizou}@utdallas.edu

ABSTRACT

We report on the development of a noisy speech corpus suitable for evaluation of speech enhancement algorithms. This corpus is used for the subjective evaluation of 13 speech enhancement methods encompassing four classes of algorithms: spectral subtractive, subspace, statistical-model based and Wiener algorithms. The subjective evaluation was performed by Dynastat, Inc. using the ITU-T P.835 methodology designed to evaluate the speech quality along three dimensions: signal distortion, noise distortion and overall quality. This paper reports the results of the subjective tests.

1. INTRODUCTION

Over the past three decades, various speech enhancement algorithms have been proposed to improve the performance of modern communication devices in noisy environments. Yet, it still remains unclear as to which speech enhancement algorithm performs well in real-world listening situations where the background noise level and characteristics are constantly changing. Reliable and fair comparison between algorithms has been elusive for several reasons, including lack of common speech database for evaluation of new algorithms, differences in the types of noise used and differences in the testing methodology. Subjective evaluation of speech enhancement algorithms is further complicated by the fact that the quality of enhanced speech has both signal and noise distortion components, and it is not clear as to whether listeners base their quality judgments on the signal distortion, noise distortion or both. Without having access to a common speech database, it is nearly impossible for researchers to compare at very least the objective performance of their algorithms with that of others.

In this paper, we report on the development of a noisy speech corpus (NOIZEUS) suitable for evaluation of speech enhancement algorithms. This corpus is subsequently used in a comprehensive subjective evaluation of 13 speech enhancement algorithms encompassing four different classes of algorithms: spectral subtractive, subspace, statistical-model based and Wiener algorithms. The enhanced speech files were sent to Dynastat, Inc (Austin, TX) for subjective evaluation using the recently standardized methodology for evaluating noise suppression algorithms based on ITU-T P.835 [1].

2. NOIZEUS: A NOISY SPEECH CORPUS FOR EVALUATION OF ENHANCEMENT ALGORITHMS

NOIZEUS[1] is a noisy speech corpus recorded in our lab to facilitate comparison of speech enhancement algorithms among research groups. The noisy database contains 30 IEEE sentences [2] produced by three male and three female speakers, and was corrupted by eight different real-world noises at different SNRs. Thirty sentences from the IEEE sentence database were recorded in a sound-proof booth using Tucker Davis Technologies (TDT) recording equipment. The sentences were produced by three male and three female speakers (5 sentences/speaker). The IEEE database was used as it contains phonetically-balanced sentences with relatively low word-co-ntext predictability. The thirty sentences were selected from the IEEE database so as to include all phonemes in the American English language. The sentences were originally sampled at 25 kHz and downsampled to 8 kHz. To simulate the receiving frequency characteristics of telephone handsets, the speech and noise signals were filtered by the modified Intermediate Reference System (IRS) filters used in ITU-T P.862 for evaluation of the PESQ measure.

Noise was artificially added to the speech signal as follows. The IRS filter was independently applied to the clean and noise signals. The active speech level of the filtered clean speech signal was first determined using the method B of ITU-T P.56. A noise segment of the same length as the speech signal was randomly cut out of the noise recordings, appropriately scaled to reach the desired SNR level and finally added to the filtered clean speech signal. Noise signals were taken from the AURORA database [3] and included the following recordings from different places: babble (crowd of people), car, exhibition hall, restaurant, street, airport, train station, and train. The noise signals were added to the speech signals

*Research supported in part by NIDCD/NIH.

[1] Available at: http://www.utdallas.edu/~loizou/speech/noizeus/.

Algorithm	Equation/parameters	Ref
KLT	Eq. 14,48	[8]
pKLT	Eq. 34, ν=0.08	[9]
MMSE-SPU	Eq. 7,51, q=0.3	[10]
logMMSE	Eq. 20	[11]
logMMSE-ne	Eq. 20	[11]
logMMSE-SPU	Eq. 2,8,10,16	[12]
pMMSE	Eq. 12	[13]
RDC	Eq. 6,7,10,14,15	[14]
RDC-ne	Eq. 6,7,10,14,15	[14]
MB	Eq. 4-7	[15]
WavThr	Eq. 11,25	[16]
Wiener_as	Eq. 3-7	[4]
AudSup	Eq. 26,38, $\nu_b(i)$=1,2 iterations	[17]

Table 1. List of 13 speech enhancement algorithms evaluated. SPU=speech presence uncertainty, ne=noise estimation.

at SNRs of 0dB, 5dB, 10dB and 15dB.

3. ALGORITHMS EVALUATED

A total of 13 different speech enhancement methods were evaluated based on *our own* implementation. Representative algorithms from four different cla-sses of enhancement algorithms were chosen: three spectral subtractive algorithms, two subspace algorithms, three Wiener algorithms[2] and five statistical-model based algorithms. A subset of those algorithms were evaluated with and without noise-estimation algorithms. The parameters used in the implementation of these algorithms were the same as those published unless stated otherwise[3]. Table 1 shows the list of algorithms evaluated with the associated parameters and Equations given in the references. The decision-directed approach was used for estimating the *a priori* SNR in the statistical methods and the Wiener_as method [4] with a=0.98.

The majority of the algorithms utilized a voice activity detector [5] to update the noise spectrum during the speech-absent periods. The subspace methods used a different VAD method [6] with threshold value set to 1.2. To assess the merit of noise-estimation algorithms [7], two speech-enhancement algorithms were implemented with both VAD and noise estimation algorithms. These algorithms are indicated in Table 1 with the suffix '-ne'.

[2]The Wiener-type algorithms were grouped separately since these algorithms estimate the complex spectrum while the statistical-model algorithms estimate the magnitude spectrum in the mean square sense.

[3]No adjustments were made for algorithms (e.g., [12]) originally designed for 16 kHz.

4. SUBJECTIVE EVALUATION

To reduce the length and cost of the subjective evaluations, only a subset of the NOIZEUS corpus was processed by the 13 algorithms and submitted to Dynastat, Inc. for formal subjective evaluation. A total of 20 sentences corrupted in four background noise environments (car, street, babble and train) at two levels of SNR (5dB and 10dB) were processed and presented to 32 listeners for evaluation. These sentences were spoken by two male speakers and two female speakers.

The subjective tests were designed according to ITU-T recommendation P.835. The P.835 methodology was designed to reduce the listener's uncertainty in a subjective test as to which component(s) of a noisy speech signal, i.e., the speech signal, the background noise, or both, should form the basis of their ratings of overall quality. This method instructs the listener to successively attend to and rate the enhanced speech signal on: a) the speech signal alone using a scale of signal distortion (SIG) - [1= very unnatural, 5=very natural], b) the background noise alone using a scale of background conspicuous/intrusiveness (BAK) - [1=very conspicuous, very intrusive, 5=not noticeable], c) the overall effect using the scale of the Mean Opinion Score (OVRL) - [1=bad, 5=excellent].

The process of rating the signal and background of noisy speech was designed to lead the listener to integrate the effects of both the signal and the background in making their ratings of overall quality. Each trial in a P.835 test involves a triad of speech samples – three samples of the test condition where each sample is a short segment of speech recorded in background noise, e.g., a single sentence. For each sample within the triad, listeners successively used one of the three five-point rating scales, SIG, BAK, and OVRL, to register their judgments of the quality of the test condition. In addition to the experimental conditions, each experiment included a number of reference conditions designed to independently vary the listener's SIG, BAK, and OVRL ratings over the entire five-point range of the rating scales. More details about the testing methodology can be found in [18]. The figures show the mean scores for SIG, BAK, and OVRL scales for the 13 methods evaluated. The mean scores for the noisy speech (unprocessed) files are also shown for reference.

5. DISCUSSION AND CONCLUSIONS

Of the two subspace algorithms examined, the generalized subspace approach [8] performed consistently better in OVRL scale across all SNR conditions and four types of noise. The performance of these two methods was distinctively different in +5dB car noise. Lower signal distortion (i.e., higher SIG scores) were observed with the generalized subspace method in most conditions. Of the five statistical-model based algorithms examined, the log-MMSE and the perceptually motivated MMSE (pM-MSE) algorithms performed the best. Performance of the pMMSE algorithm was comparable to that

of the MMSE algorithm which incorporated speech-presence uncertainty (the pMMSE algorithm did not). Lower noise distortion (i.e., high-er BAK scores) was obtained with the pMMSE method in several conditions (5dB train, 5dB car, 10dB street). It was surprising to see that the noise-estimation algorithm [7] did not provide significant improvements to the performance of the log-MMSE algorithm (small improvements were noted only in street noise). Incorporating speech-presence uncertainty as per [12] did not improve the performance of the log-MMSE algorithm. In fact, it degraded performance. Of the two spectral-subtractive algorithms tested, the multi-band spectral subtraction algorithm [15] performed consistently the best across all conditions. Incorporating a noise-estimation algorithm did not improve the performance of the reduced-delay spectral subtraction algorithm. One possible explanation for that is that the speech files were too brief in duration to observe the real benefit of noise-estimation algorithms. Finally, of the three Wiener filtering type of algorithms, the method proposed in [4] based on *a priori* SNR, performed the best. This method also produced consistently the lowest signal distortion comparable to the statistical-model based methods. It did, however, suffer from high noise distortion.

Overall, the statistical-model based methods performed the best across all conditions, followed by the multi-band spectral subtraction method [15].

6. REFERENCES

[1] ITU-T P.835, *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithms*, ITU-T Recommendation P.835, 2003.

[2] IEEE Subcommittee, "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio and Electroacoustics*, pp. 225–246, 1969.

[3] H. Hirsch and D. Pearce, "The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in *ISCA ITRW ASR2000*, Sept. 2000, Paris, France.

[4] P. Scalart and J. Filho, "Speech enhancement based on a priori signal to noise estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1996, pp. 629–632.

[5] J. Sohn, N. Kim, and W. Sung, "A statistical model-based voice activity detection," *IEEE Signal Processing Letters*, pp. 1–3, Jan. 1999.

[6] U. Mittal and N. Phamdo, "Signal/noise KLT based approach for enhancing speech degraded by colored noise," *IEEE Trans. Speech Audio Proc.*, vol. 8, pp. 159–167, Mar. 2000.

[7] S. Rangachari and P. C. Loizou, "A noise estimation algorithm for highly non-stationary environments," *Speech Communication*, vol. 28, pp. 220–231, Feb. 2006.

[8] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Trans. Speech Audio Proc.*, pp. 334–341, July 2003.

[9] F. Jabloun and B. Champagne, "Incorporating the human hearing properties in the signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Proc.*, vol. 11, pp. 700–708, 2003.

[10] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-32, pp. 1109–1121, 1984.

[11] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-33, pp. 443–445, 1985.

[12] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *IEEE Signal Processing Letters*, vol. 9, pp. 113–116, Apr. 2002.

[13] P. C. Loizou, "Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum," *IEEE Trans. Speech Audio Proc.*, pp. 857–869, Sept. 2005.

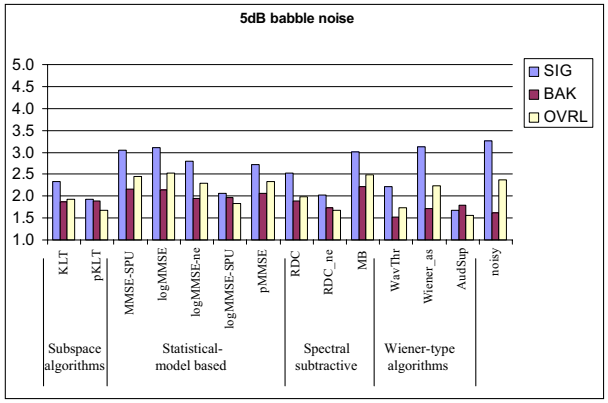
[14] H. Gustafsson, S. Nordholm, and I. Claesson, "Spectral subtraction using reduced delay convolution and adaptive averaging," *IEEE Trans. Speech Audio Proc.*, pp. 799–807, 2001.

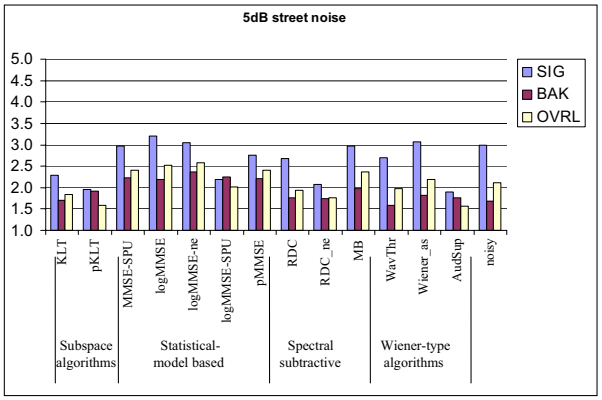
[15] S. Kamath and P. C. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2002.

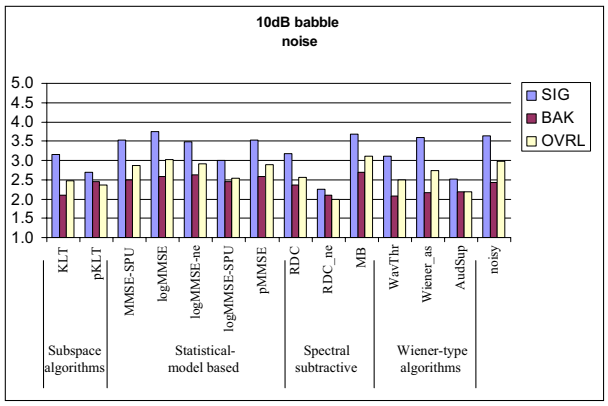
[16] Y. Hu and P. C. Loizou, "Speech enhancement based on wavelet thresholding the multitaper spectrum," *IEEE Trans. Speech Audio Proc.*, pp. 59–67, Jan. 2004.

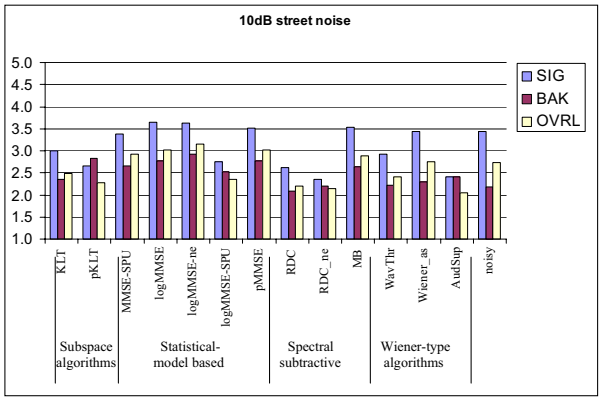
[17] D. E. Tsoukalas, J. N. Mourjopoulos, and G. Kokkinakis, "Speech enhancement based on audible noise suppression," *IEEE Trans. Speech Audio Proc.*, vol. 5, pp. 479–514, Nov. 1997.

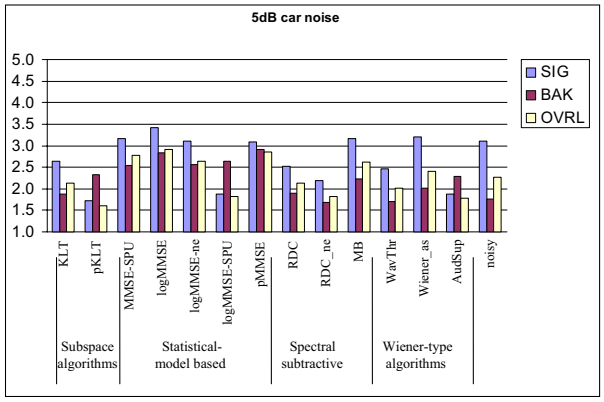
[18] Y. Hu and P. C. Loizou, "Subjective evaluation and comparison of speech enhancement algorithms," submitted to *Speech Communication*.

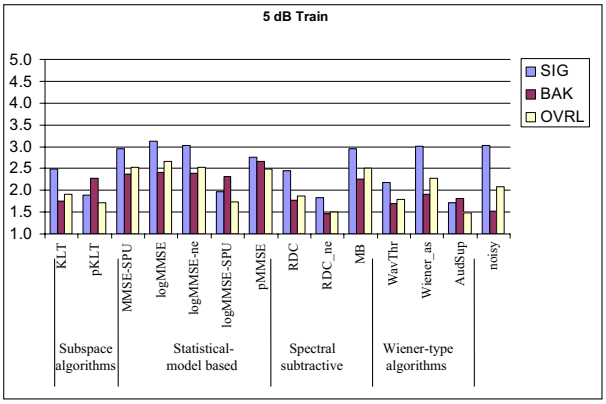

5dB babble noise
SIG
BAK
OVRL
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
KLT
pKLT
MMSE-SPU
logMMSE
logMMSE-ne
logMMSE-SPU
pMMSE
RDC
RDC_ne
MB
WavThr
Wiener_as
AudSup
noisy
Subspace algorithms
Statistical-model based
Spectral subtractive
Wiener-type algorithms

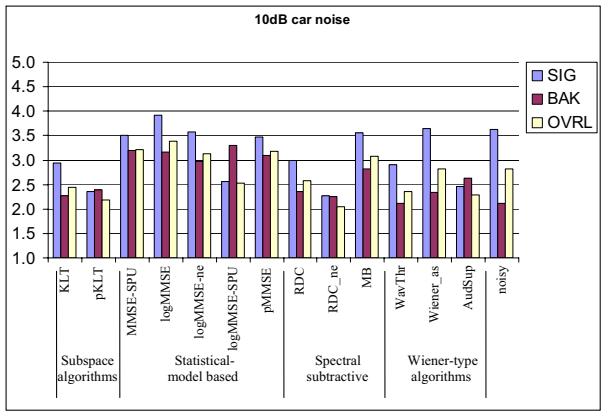

5dB street noise
SIG
BAK
OVRL
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
KLT
pKLT
MMSE-SPU
logMMSE
logMMSE-ne
logMMSE-SPU
pMMSE
RDC
RDC_ne
MB
WavThr
Wiener_as
AudSup
noisy
Subspace algorithms
Statistical-model based
Spectral subtractive
Wiener-type algorithms

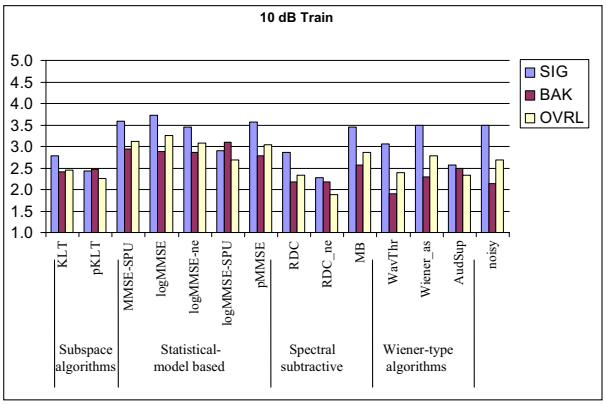

10dB babble noise
SIG
BAK
OVRL
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
KLT
pKLT
MMSE-SPU
logMMSE
logMMSE-ne
logMMSE-SPU
pMMSE
RDC
RDC_ne
MB
WavThr
Wiener_as
AudSup
noisy
Subspace algorithms
Statistical-model based
Spectral subtractive
Wiener-type algorithms


10dB street noise
SIG
BAK
OVRL
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
KLT
pKLT
MMSE-SPU
logMMSE
logMMSE-ne
logMMSE-SPU
pMMSE
RDC
RDC_ne
MB
WavThr
Wiener_as
AudSup
noisy
Subspace algorithms
Statistical-model based
Spectral subtractive
Wiener-type algorithms


5dB car noise
SIG
BAK
OVRL
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
KLT
pKLT
MMSE-SPU
logMMSE
logMMSE-ne
logMMSE-SPU
pMMSE
RDC
RDC_ne
MB
WavThr
Wiener_as
AudSup
noisy
Subspace algorithms
Statistical-model based
Spectral subtractive
Wiener-type algorithms


5 dB Train
SIG
BAK
OVRL
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
KLT
pKLT
MMSE-SPU
logMMSE
logMMSE-ne
logMMSE-SPU
pMMSE
RDC
RDC_ne
MB
WavThr
Wiener_as
AudSup
noisy
Subspace algorithms
Statistical-model based
Spectral subtractive
Wiener-type algorithms


10dB car noise
SIG
BAK
OVRL
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
KLT
pKLT
MMSE-SPU
logMMSE
logMMSE-ne
logMMSE-SPU
pMMSE
RDC
RDC_ne
MB
WavThr
Wiener_as
AudSup
noisy
Subspace algorithms
Statistical-model based
Spectral subtractive
Wiener-type algorithms


10 dB Train
SIG
BAK
OVRL
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
KLT
pKLT
MMSE-SPU
logMMSE
logMMSE-ne
logMMSE-SPU
pMMSE
RDC
RDC_ne
MB
WavThr
Wiener_as
AudSup
noisy
Subspace algorithms
Statistical-model based
Spectral subtractive
Wiener-type algorithms