Speech Intelligibility and Quality: A Comparative Study of Speech Enhancement Algorithms

Xiaodong Xu
Software Research Institute
Athlone Institute of Technology
Athlone, Ireland
x.xu@research.ait.ie

Ronan Flynn
Faculty of Engineering & Informatics
Athlone Institute of Technology
Athlone, Ireland
rflynn@ait.ie

Michael Russell
Faculty of Engineering & Informatics
Athlone Institute of Technology
Athlone, Ireland
mrussell@ait.ie

***Abstract*— Mobile devices are widely used today for speech communication. The environments in which these devices are used are widely varied and often the level of background noise in the speaker's environment can be significant. The purpose of speech enhancement is to reduce the level of background noise, ideally to such a level that it is not noticed by the listener. While speech enhancement algorithms can significantly reduce the noise level in a speech signal, improving speech quality, it is widely recognized that enhancement algorithms can have a negative impact on speech intelligibility. This paper compares the effect of three different speech enhancement algorithms on the intelligibility and the quality of speech. This work is the initial phase of an investigation into mitigating the impact of speech enhancement algorithms on speech intelligibility. The speech enhancement algorithms evaluated each use different approaches for noise reduction, namely, a statistical model-based algorithm, a noise estimation algorithm and a wavelet packet decomposition-based algorithm. Two objective speech intelligibility measurements and three objective speech quality measurements are used to assess the performance of the enhancement algorithms. The results of the experiments show that all the speech enhancement algorithms in this study have a negative impact on speech intelligibility to varying degrees.**

Keywords—speech enhancement; speech quality; speech intelligibility

I. INTRODUCTION

Speech communication using some form of mobile device is an essential part of life today. With increased mobility, the users of mobile devices are to be found having conversations in many different environments, with different types and levels of background noise. This background noise when added to the speech can be uncomfortable for a listener, making it difficult to partake in the conversation. The speech quality is used to describe how comfortable it is for someone to listen to the speech. The noise, classed as additive noise, can also affect the speech intelligibility, a measure of how well the content of the speech conversation can be understood. In order to improve the intelligibility and the quality of speech that originates in a noisy background the speech needs to be processed in some way before transmission to a listener.

The aim of speech enhancement algorithms is to remove, or reduce the level of, noise from a noisy speech utterance. Speech enhancement algorithms are to be found in human-to-human communication systems as well as human-to-machine communication systems. Automatic speech recognition (ASR) is an example of an application where an application is designed to recognise words or commands and respond appropriately when interacting with a human. In ASR, speech quality is not as important as in human-to-human communication. For both ASR and human-to-human communication, speech intelligibility, correctly interpreting the meaning or intent of the speech content, is important. Specific speech enhancement algorithms are therefore more suited for use in particular circumstances. For human-to-human communication systems enhancing the noisy speech should not compromise the speech intelligibility and should improve the speech quality.

Speech enhancement algorithms can be classed under different headings, for example, statistical model-based[1], noise-estimation-based [2], subspace algorithms [3] and spectral subtraction algorithms [4]. More recently wavelet packet decomposition [5], machine learning and information theory [6] have been applied by researchers to the area of speech enhancement. Enhancement algorithms can also be categorised based on which domain they operate in. The three domains are the time domain, the frequency domain, and the time-frequency domain. Subspace enhancement algorithms are implemented in the time domain. The principle of subspace enhancement algorithms is to decompose the noisy speech's vector space into two subspaces, one with the noisy signal and the other with the clean speech signal. Statistical model-based, spectral subtraction and noise estimation algorithms operate in the frequency domain. Statistical model-based algorithms estimate the distribution of speech Fourier transform coefficients in order to fit the statistical model in question. Spectral subtraction algorithms subtract the estimated noise power spectrum from the noisy speech power spectrum. Spectral subtraction is based on the concept that the noise and the speech signal are independent of each other so that the clean speech spectrum can be obtained. The conventional way to do noise estimation is to use voice activity detection (VAD) [7], which is used to estimate and update the noise spectrum during periods in which there is considered to be no speech signal. Wavelet-based enhancement algorithms operate in the time-frequency domain. These algorithms typically use threshold techniques applied to wavelet coefficients generated from the noisy speech signal in order to remove the noise from the noisy speech.

Speech enhancement algorithms can significantly improve the speech quality of a noisy speech signal. However, it is recognised that speech enhancement has a negative impact on speech intelligibility [8]. Speech quality and speech intelligibility can be measured using both subjective and objective methods. However, subjective measurements have the disadvantage that the reliability of the listener evaluating the speech intelligibility or quality can be a problem. The decision on intelligibility and quality can be affected by the listener's background, accent, and hearing ability. It can take a long period of time to do subjective evaluations. However, the advantage of subjective testing is that the results are determined by humans as opposed to machines. Objective assessment of speech quality and intelligibility removes any bias that might be introduced using human evaluators. Objective speech quality measurement methods include perceptual evaluation of speech quality (PESQ) [9], segmental signal-to-noise (SNR) ratio [10] and HASQI [11]. Objective speech intelligibility measurement methods include SNR loss [12] and HASPI [11].

The purpose of the work presented here is to compare three different speech enhancement algorithms and to evaluate the intelligibility and the quality of the enhanced speech. The three enhancement algorithms each use different methods to enhance noisy speech, which was taken from the NOIZEUS speech database [13]. The enhancement algorithms are a minimum-mean-square-error short-time spectral amplitude (MMSE STSA) estimator algorithm [1], a noise estimation algorithm [2] and a wavelet packet decomposition algorithm [5]. The MMSE-STSA algorithm [1] was chosen because it is well recognized and is often used as a benchmark for comparison. The noise estimation algorithm [2] was chosen because it was designed for enhancing signal with highly non-stationary noise. The wavelet packet decomposition-based algorithm [5] was chosen because it is an example of a new approach applied to the speech enhancement domain. The intelligibility and quality of the enhanced noisy speech produced by the three enhancement algorithms were evaluated using the objective methods highlighted previously. The motivation for doing this comparative study of speech enhancement algorithms and how they affect speech intelligibility is to inform future work. Of interest to the authors is the impact of enhancement on intelligibility in particular. The results of the work presented will be used to develop a speech enhancement method that has minimal impact on speech intelligibility.

This paper is organized as follows. Section II provides a brief overview of the enhancement algorithms used in this work. The objective intelligibility and quality measurements are discussed in Section III. Section IV presents the results of the experiments with the discussion. Finally, conclusions and plans for future work are given in section V.

II. Speech Enhancement Algorithms

This section provides an overview of the three speech enhancement algorithms used in this comparative study. Each algorithm is different in its approach to speech enhancement. The algorithms include the well-established minimum mean-square error short-time spectral amplitude (MMSE STSA) estimator of Ephraim & Malah [1], the noise-estimation algorithm of Rangachari & Loizou [2] and the very-recent wavelet-based algorithm of Ben messaoud *et al.* [5].

A. Minimum-Mean-Square-Error Short-Time Spectral Amplitude (MMSE STSA) Estimator

The minimum-mean-square-error short-time spectral amplitude (MMSE STSA) estimator of Ephraim & Malah [1] is well known in the field of speech enhancement. This algorithm operates in the frequency domain. A spectral gain value is determined for each frequency bin and applied to the noisy speech. The process is illustrated in Fig. 1.

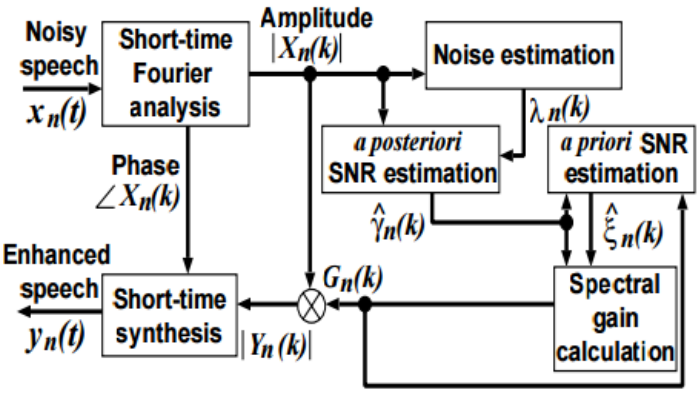


Figure 1: Block diagram illustrating the steps for speech enhancement using the MMSE STSA algorithm [14]

The input speech signal $x_n(t)$, where n is the frame index, is converted to the frequency domain $X_n(k)$, where k is the frequency index. The spectral gain function $G_n(k)$ is dependent on the *a posteriori* and the *a priori* SNRs, $\hat{\gamma}_n(k)$ and $\xi_n(k)$ respectively, of the k^{th} spectral components. The noise power estimate $\lambda_n(k)$ in the noisy speech signal is constantly updated on a frame-by-frame basis. Noise estimation is determined based on non-speech periods of the utterances. The enhanced speech $y_n(t)$ is recovered using the original phase from the noisy speech signal and the updated spectral amplitude $|Y_n(k)|$,

B. Noise Estimation combined with Wiener-type Spectral Gain

The noise estimation algorithm proposed by Rangachari & Loizou [2] calculates the speech-presence probability instead of using voice activity detection (VAD). The noise estimation algorithm steps are detailed in Fig. 2, in which k is the frequency bin index and λ is the frame index. A fast Fourier transform is used to convert the time domain speech signal to the frequency domain.

The estimated clean speech spectrum is evaluated as

$$C(\lambda, k) = \max\left\{ |Y(\lambda, k)|^2 - D(\lambda, k), vD(\lambda, k) \right\} \quad (1)$$

where $Y(\lambda, k)$ is the noisy speech signal, $D(\lambda, k)$ is the noise power spectrum estimate according to Fig. 2 and v is a small positive constant. The noise-estimation algorithm in [2] is combined with a Wiener-type speech-enhancement algorithm

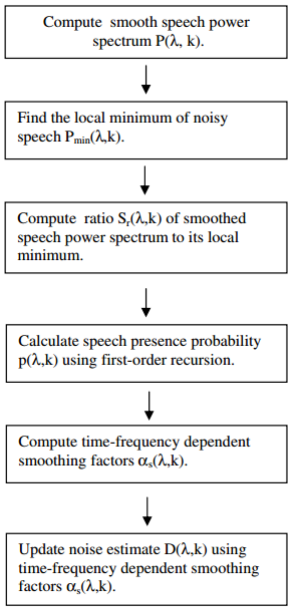


Figure 2: Noise estimation proposed by Rangachari & Loizou [2]

that has the spectral gain function with over subtraction factor μ_k in equation (2).

$$G(\lambda,k) = \frac{C(\lambda,k)}{C(\lambda,k) + \mu_k D(\lambda,k)} \qquad (2)$$

C. Wavelet packet-based Speech Enhancement

Wavelet-based speech enhancement algorithms [15], [16] have been the focus of some research effort in recent years. The noisy speech that undergoes wavelet decomposition produces a coefficient matrix which some form of threshold technique is applied to. Therefore, the threshold method chosen will have an impact on the speech enhancement performance of a wavelet-based enhancement algorithm. For the work presented here, the very recently proposed wavelet packet decomposition algorithm [5] is chosen. A block diagram showing the algorithm steps is given in Fig. 3.

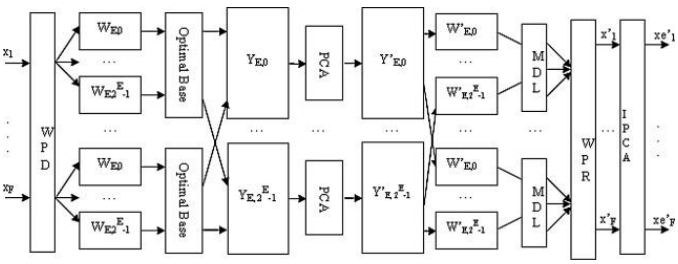


Figure 3: Wavelet packet transform-based speech enhancement algorithm [5]

Noisy speech is divided into frames that are decomposed into wavelet packet (WP) coefficient matrices, $W_{E,2}{}^{E}{}_{-1}$, where E indicates the decomposition level. A best tree function applied to these matrices yields a new set of WP coefficient matrices, $Y_{E,2}{}^{E}{}_{-1}$. Principal component analysis is applied to these matrices, generating score matrices, $Y'_{E,2}{}^{E}{}_{-1}$. Corresponding column vectors from $W_{E,2}{}^{E}{}_{-1}$ are combined to obtain $W'_{E,2}{}^{E}{}_{-1}$. For denoising, the minimum description length (MDL) criterion [17] is used as the thresholding method on matrix $W'_{E,2}{}^{E}{}_{-1}$. Reconstruction is achieved by overlap-adding operation the enhanced speech frames using a WP rebuild (WPR) method. The final step in the generation of the enhamced speech signal is an improved PCA (IPCA), which is discussed in detail in [5].

III. Objective Intelligibility and Quality Measurements

This section describes the objective measurements that were used to evaluate the speech intelligibility and speech quality of the enhanced speech produced by the three enhancement methods presented in Section II. The objective speech intelligibility measurements used in the evaluation are HASPI [11] and SNR loss [12]. The objective speech quality measurements used are Segmental SNR [10], PESQ [9] and HASQI [11].

A. Objective measurements for speech intelligibility

1) SNR loss

SNR loss [12] is defined as

$$L(n,k) = SNR_X(n,k) - SNR_{\hat{X}}(n,k) \qquad (3)$$

where $SNR_X(n,k)$ is the input signal-to-noise ratio, $SNR_{\hat{X}}(n,k)$ is signal-to-noise ratio of the enhanced signal, k is the frequency band index of the n^{th} frame. A mapping function is used to limit the SNR loss values to a range between 0 and 1 as described in [12]. The overall SNR loss for a speech utterance is calculated by averaging the limited values across all frequency bands and over all frames in the utterance. A lower SNR loss value indicates higher speech intelligibility. Ideally, the enhanced speech spectrum should equal the spectrum of the clean speech resulting in an SNR loss of zero.

2) Hearing aid speech intelligibility index

The hearing aid speech intelligibility index (HASPI) [11] was created to predict the speech intelligibility performance for speech enhancement algorithms used with assistive listening devices. HASPI defines two measurements of signal distortion. The first distortion measurement compares the change in spectral shape over time between the original speech signal and the processed, or enhanced, speech signal. The second is a cross-correlation measurement, which focuses on the high-level portions of the signal in each frequency band. The score range for this method is between 0 and 1. The higher the value means the better the intelligibility level. The Matlab code for the implementation of HASPI was obtained directly from the authors of [11].

B. Objective measurements of speech quality

1) Segmental SNR

The signal-to-noise ratio (SNR) is a basic objective method to measure the quality of a noisy speech signal. However, this can be extended to give an improved measure of speech quality, namely the segmental SNR [10], which is defined as

$$SNRseg = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\frac{\sum_{n=Nm}^{Nm+N-1}x^2(n)}{\sum_{n=Nm}^{Nm+N-1}(x(n)-\hat{x}(n))^2} \quad (4)$$

where $x(n)$ is the original speech signal, $\hat{x}(n)$ is the enhanced speech, N is the speech frame length, n is the sample index in a frame, M is the total number of frames in the speech utterance, and m is the frame index. A disadvantage of this method is that during silent periods, the signal energy might be very small and this can produce a negative segmental SNR value [18].

2) Perceptual Evaluation of Speech Quality

The method used for the Perceptual Evaluation of Speech Quality (PESQ) [9] is detailed in Fig. 4.

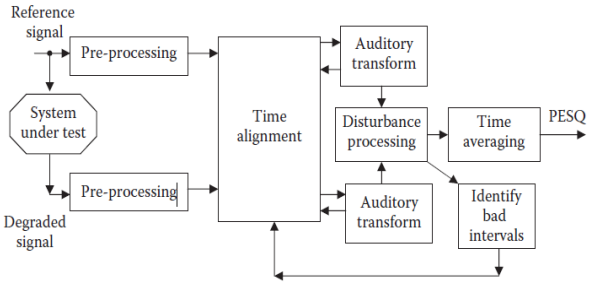


Figure 4: PESQ measurement [9]

PESQ requires the original clean speech and the degraded (noisy) speech as inputs. The score of a PESQ measurement ranges from 1 to 5, indicating a speech quality from bad to excellent.

3) Hearing aid speech quality indexI

The hearing aid speech quality indices (HASQI) are objective quality measurements recently proposed in [11]. Similar to HASPI, it was proposed for use originally with assistive listening devices. The quality measurement method requires the original speech signal and the degraded speech signal. The score range for the measurement is from 0 to 1, the higher the better. The HASQI method uses two measurements. The first measurement is a comparison of the time-frequency envelope between the original and degraded speech. The second is a cross-correlation measurement. The Matlab code for the implementation of HASQI was obtained directly from the authors of [11].

IV. Results and Discussion

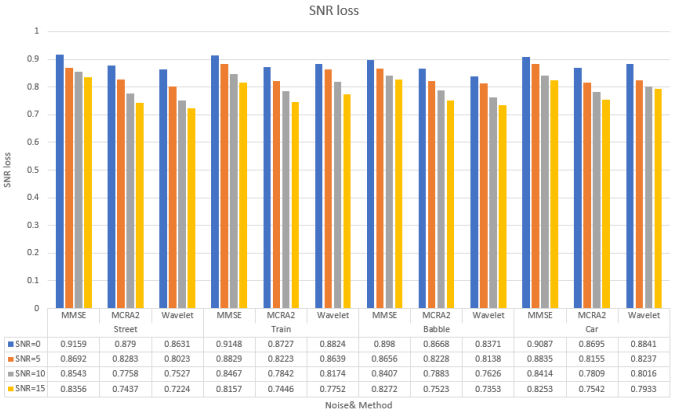
Speech with four types of noise (street, train, babble and car) at different SNRs was taken from the Noizeus database [13]. The noisy speech was processed by each of the three speech enhancement algorithms described in Section II. The objective intelligibility and quality measurements described in Section III were used to evaluate the performance of the enhancement algorithms.

A. Result of Objective intelligibility measurements

Speech intelligibility as measured by SNR loss is shown in Fig. 5. The lower SNR loss figures indicate a higher speech intelligibility. Across all noise types and noise levels, intelligibility evaluation using SNR loss indicates that speech enhanced by the MMSE STSA algorithm [1] has the poorest intelligibility compared to the other two enhancement algorithms used in this study. The results in Fig. 5 also show that speech enhanced using the noise estimation algorithm of Rangachari & Loizou [2], referred to as MCRA2 [18] in Figs. 6-10, has the highest intelligibility (lowest SNR loss) for stationary noise (train and car). For non-stationary noise (street and babble) the wavelet-based [5] enhancement algorithm results in speech with the highest intelligibility. The results from Fig. 6 suggest that for speech intelligibility, the noise type (stationary or non-stationary) should be considered when selecting a speech enhancement algorithm.

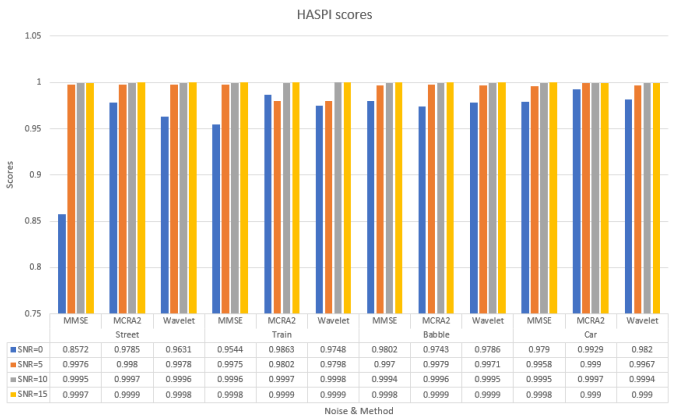
For speech intelligibility as measured by HASPI, Fig. 6 indicates that all three speech enhancement algorithms are very similar for SNRs of 15 dB, 10 dB and 5 dB across all noise types. For an SNR of 0 dB, intelligibility as measured by HASPI shows that speech enhanced using the MMSE STSA algorithm [1] is poor across all noise types except bubble noise, compared to the other two enhancement algorithms. For speech with an SNR of 0 dB enhanced using the noise estimation algorithm of Rangachari & Loizou [2], the speech intelligibility is marginally higher than the wavelet-based algorithm [5] for street, train and car noise.

The results in Fig. 5 and Fig. 6 suggest that speech enhancement using the noise estimation method of [2] or the wavelet-based speech enhancement [5] should be considered if intelligibility of the enhanced speech is important.



	MMSE	MCRA2	Wavelet	MMSE	MCRA2	Wavelet	MMSE	MCRA2	Wavelet	MMSE	MCRA2	Wavelet
		Street			Train			Babble			Car	
SNR=0	0.9159	0.879	0.8631	0.9148	0.8727	0.8824	0.898	0.8668	0.8371	0.9087	0.8695	0.8841
SNR=5	0.8692	0.8283	0.8023	0.8829	0.8223	0.8639	0.8656	0.8228	0.8138	0.8835	0.8155	0.8237
SNR=10	0.8543	0.7758	0.7527	0.8467	0.7842	0.8174	0.8407	0.7883	0.7626	0.8414	0.7809	0.8016
SNR=15	0.8356	0.7437	0.7224	0.8157	0.7446	0.7752	0.8272	0.7523	0.7353	0.8253	0.7542	0.7933

Figure 5: SNR loss objective intelligibility evaluation of enhanced speech



	MMSE	MCRA2	Wavelet	MMSE	MCRA2	Wavelet	MMSE	MCRA2	Wavelet	MMSE	MCRA2	Wavelet
		Street			Train			Babble			Car	
SNR=0	0.8572	0.9785	0.9631	0.9544	0.9863	0.9748	0.9802	0.9743	0.9786	0.979	0.9929	0.982
SNR=5	0.9976	0.998	0.9978	0.9975	0.9802	0.9798	0.997	0.9979	0.9971	0.9958	0.999	0.9967
SNR=10	0.9995	0.9997	0.9996	0.9996	0.9997	0.9998	0.9994	0.9996	0.9995	0.9995	0.9997	0.9994
SNR=15	0.9997	0.9999	0.9998	0.9998	0.9999	0.9999	0.9998	0.9999	0.9999	0.9998	0.999	0.999

Figure 6: HASPI objective intelligibility evaluation of enhanced speech

B. *Result of objective quality measurements*

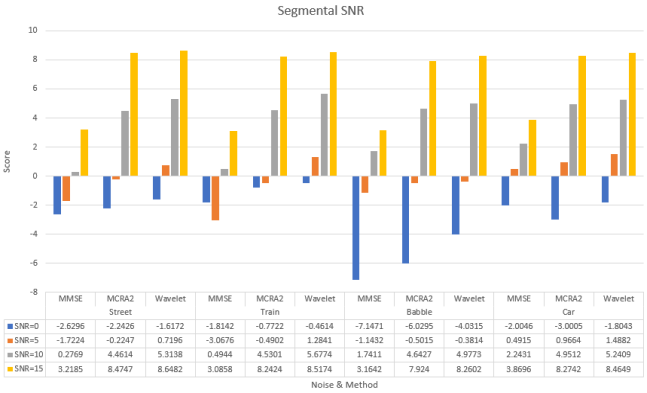
Figs. 7-9 present objective speech quality measurements using Segmental SNR, PESQ and HASQI respectively.

When using Segmental SNR to evaluate speech quality, Fig. 7 shows that speech enhanced using wavelet packet decomposition algorithm [5] has highest quality across all noise types and SNR values. Speech enhanced by the MMSE STSA algorithm [1] has the poorest quality according to the Segmental SNR measurements.

When using PESQ to evaluate speech quality Fig. 8 indicates that speech enhanced by the wavelet packet-based enhancement algorithm has the highest quality overall. This is in agreement with that suggested by the Segmental SNR scores in Fig. 8.

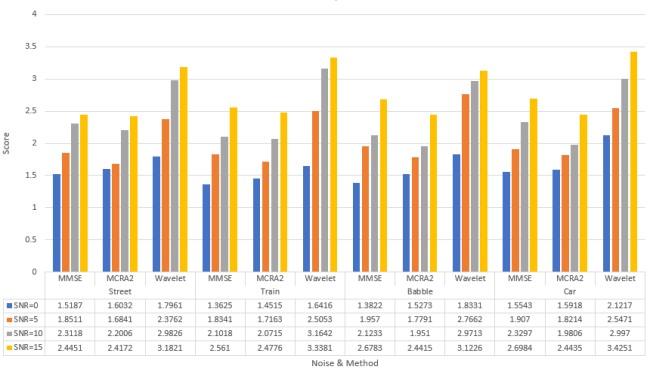
From the HASQI measurements in Fig. 9 the MMSE-STSA enhancement algorithm produces speech with the lowest quality scores for street, train and car noise at SNRs of 0 dB and 5 dB. For the same three noise types at SNRs of 10 dB and 15 dB, overall the wavelet-based enhancement method [5] is outperforms the noise estimation-based enhancement method [2]. All three enhancement algorithms produce speech of a similar quality for babble noise across all four SNR levels.

The results in Figs. 7-9 demonstrate that speech enhanced by the wavelet packet-based algorithm [5] has a higher speech quality compared to speech enhanced by the MMSE-STSA [1] and noise estimation-based [2] algorithms.



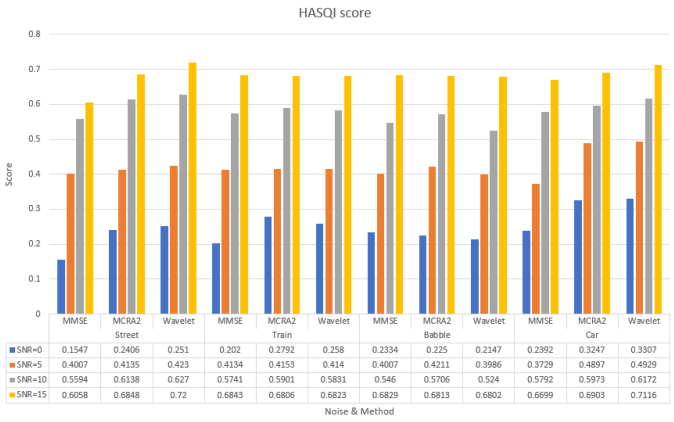
	MMSE	MCRA2 Street	Wavelet	MMSE	MCRA2 Train	Wavelet	MMSE	MCRA2 Babble	Wavelet	MMSE	MCRA2 Car	Wavelet
SNR=0	-2.6296	-2.2426	-1.6172	-1.8142	-0.7722	-0.4614	-7.1471	-6.0295	-4.0315	-2.0046	-3.0005	-1.8043
SNR=5	-1.7224	-0.2247	0.7196	-3.0676	-0.4902	1.2841	-1.1432	-0.5015	-0.3814	0.4915	0.9664	1.4882
SNR=10	0.2769	4.4614	5.3138	0.4944	4.5301	5.6774	1.7411	4.6427	4.9773	2.2431	4.9512	5.2409
SNR=15	3.2185	8.4747	8.6482	3.0858	8.2424	8.5174	3.1642	7.924	8.2602	3.8696	8.2742	8.4649

Figure 7: Segmental SNR objective quality evaluation of enhanced speech



	MMSE	MCRA2 Street	Wavelet	MMSE	MCRA2 Train	Wavelet	MMSE	MCRA2 Babble	Wavelet	MMSE	MCRA2 Car	Wavelet
SNR=0	1.5187	1.6032	1.7961	1.3625	1.4515	1.6416	1.3822	1.5273	1.8331	1.5543	1.5918	2.1217
SNR=5	1.8511	1.6841	2.3762	1.8341	1.7163	2.5053	1.957	1.7791	2.7662	1.907	1.8214	2.5471
SNR=10	2.3118	2.2006	2.9826	2.1018	2.0715	3.1642	2.1233	1.951	2.9713	2.3297	1.9806	2.997
SNR=15	2.4451	2.4172	3.1821	2.561	2.4776	3.3381	2.6783	2.4415	3.1226	2.6984	2.4435	3.4251

Figure 8: PESQ objective quality evaluation of enhanced speech



	MMSE	MCRA2 Street	Wavelet	MMSE	MCRA2 Train	Wavelet	MMSE	MCRA2 Babble	Wavelet	MMSE	MCRA2 Car	Wavelet
SNR=0	0.1547	0.2406	0.251	0.202	0.2792	0.258	0.2334	0.225	0.2147	0.2392	0.3247	0.3307
SNR=5	0.4007	0.4135	0.423	0.4134	0.4153	0.414	0.4007	0.4211	0.3986	0.3729	0.4897	0.4929
SNR=10	0.5594	0.6138	0.627	0.5741	0.5901	0.5831	0.546	0.5706	0.524	0.5792	0.5973	0.6172
SNR=15	0.6058	0.6848	0.72	0.6843	0.6806	0.6823	0.6829	0.6813	0.6802	0.6699	0.6903	0.7116

Figure 9: HASQI objective quality evaluation of enhanced speech

V. Conclusion and Future Work

This paper compared the speech intelligibility and speech quality of noisy speech that was processed by three different speech enhancement algorithms, for a range of noise types and SNR levels. The objective intelligibility assessment of the enhanced speech indicates that the noise type should be considered when selecting an enhancement algorithm so that the intelligibility of the enhanced speech is maximized. When assessing the quality of the speech produced by the three enhancement algorithms, the three objective quality measurements used show that, overall, the wavelet packet-based enhancement algorithm results in speech with the best quality. These results will inform future work that will seek to mitigate the effects of speech enhancement algorithms on speech intelligibility and speech quality.

Acknowledgment

This work was supported by the Athlone Institute of Technology President's Seed Fund. The authors would also like to thank Professor James Kates for providing the HASPI and HASQI code.

References

[1] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 32, no. 6, pp. 1109–1121, 1984.

[2] S. Rangachari and P. C. Loizou, "A noise-estimation algorithm for highly non-stationary environments," *Speech Commun.*, vol. 48, no. 2, pp. 220–231, Feb. 2006.

[3] S. Surendran and T. K. Kumar, "Perceptual Subspace Speech Enhancement with Variance Normalization," *Procedia Comput. Sci.*, vol. 54, pp. 818–828, 2015.

[4] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *ICASSP*, 2002, vol. 4, pp. 44164–44164.

[5] M. anouar Ben messaoud, A. Bouzid, and N. Ellouze, "Speech enhancement based on wavelet packet of an improved principal component analysis," *Comput. Speech Lang.*, vol. 35, pp. 58–72, Jan. 2016.

[6] N. Mohammadiha, P. Smaragdis, and A. Leijon, "Supervised and unsupervised speech enhancement using nonnegative matrix factorization," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 10, pp. 2140–2151, 2013.

[7] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," *IEEE Signal Process. Lett.*, vol. 6, no. 1, pp. 1–3, 1999.

[8] P. C. Loizou and G. Kim, "Reasons why Current Speech-Enhancement Algorithms do not Improve Speech Intelligibility and Suggested Solutions," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, no. 1, pp. 47–56, Jan. 2011.

[9] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, 2001, vol. 2, pp. 749–752.

[10] J. H. Hansen and B. L. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms.," in *ICSLP*, 1998, vol. 7, pp. 2819–2822.

[11] T. H. Falk *et al.*, "Objective quality and intelligibility prediction for users of assistive listening devices: Advantages and limitations of existing tools," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 114–124, 2015.

[12] J. Ma and P. C. Loizou, "SNR loss: A new objective measure for predicting the intelligibility of noise-suppressed speech," *Speech Commun.*, vol. 53, no. 3, pp. 340–354, 2011.

[13] "NOIZEUS: Noisy speech corpus - Univ. Texas-Dallas," Available: http://ecs.utdallas.edu/loizou/speech/noizeus/. [Accessed: 08-Mar-2017].

[14] K. Masanori, A. Sugiyama, and M. Serizawa, "Noise suppression with high speech quality based on weighted noise estimation and MMSE STSA," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. 85, no. 7, pp. 1710–1718, 2002.

[15] Y. Ghanbari and M. R. Karami-Mollaei, "A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets," *Speech Commun.*, vol. 48, no. 8, pp. 927–940, 2006.

[16] M. T. Johnson, X. Yuan, and Y. Ren, "Speech signal enhancement through adaptive wavelet thresholding," *Speech Commun.*, vol. 49, no. 2, pp. 123–133, 2007.

[17] T. Roos, P. Myllymaki, and J. Rissanen, "MDL denoising revisited," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3347–3360, 2009.

[18] P. C. Loizou, *Speech enhancement: theory and practice*. CRC press, 2013.