

## Article

# A Study on the Relationship between the Intelligibility and Quality of Algorithmically-Modified Speech for Normal Hearing Listeners

Yan Tang \* , Christopher Arnold and Trevor J. Cox 

Acoustics Research Centre, University of Salford, Salford M5 4WT, UK; C.T.Arnold@edu.salford.ac.uk (C.A.); T.J.Cox@salford.ac.uk (T.J.C.)

\* Correspondence: y.tang@salford.ac.uk; Tel.: +44-0-161-295-0469

Received: 17 November 2017; Accepted: 7 December 2017; Published: 8 December 2017

**Abstract:** This study investigates the relationship between the intelligibility and quality of modified speech in noise and in quiet. Speech signals were processed by seven algorithms designed to increase speech intelligibility in noise without altering speech intensity. In three noise maskers, including both stationary and fluctuating noise at two signal-to-noise ratios (SNR), listeners identified keywords from unmodified or modified sentences. The intelligibility performance of each type of speech was measured as the listeners' word recognition rate in each condition, while the quality was rated as a mean opinion score. In quiet, only the perceptual quality of each type of speech was assessed. The results suggest that when listening in noise, modification performance on improving intelligibility is more important than its potential negative impact on speech quality. However, when listening in quiet or at SNRs in which intelligibility is no longer an issue to listeners, the impact to speech quality due to modification becomes a concern.

**Keywords:** speech intelligibility; speech quality; normal hearing; speech modification; noise

## 1. Introduction

During the last decade, a considerable number of speech modification algorithms have been proposed in order to boost speech intelligibility in adverse listening environments while maintaining a constant input-output speech intensity. Unlike traditional speech enhancement techniques (e.g., [1–5]), which focus on dealing with the noise-corrupted speech signal (i.e., speech-plus-noise mixture) and on removing background noise from the signal to achieve better listening experiences for listeners, these speech modification algorithms aim to alter the original clean speech signal so that the intelligibility may be preserved even when listened to in non-ideal listening conditions, in which background masking sources may exist. While the majority of modification algorithms operate in the frequency domain, such as enhancing frequency components which are important to speech intelligibility in noise [6–8] and boosting certain spectral regions based on optimising objective intelligibility metrics [9–12], other algorithms make changes in the time domain, including introducing pauses into speech and speeding up or slowing down part of the speech to avoid a temporal clash between the speech and masker [10,13]. Approaches combining both spectral and temporal modifications have achieved better performance than either of the approaches alone [14–16]. With a constant energy constraint, these modification algorithms essentially reallocate the energy of speech across time and frequency.

The performance of speech enhancement and speech modification techniques is usually evaluated using different subjective approaches. For enhanced speech as the output of the former technique, perceptual *quality* ratings are often used. The artefacts due to the under-removal of the noise signal and the over-removal of the speech signal largely affect listener's preferences. In contrast, listeners' word recognition performance for modified speech in noise is measured as *intelligibility* to reflect the

effectiveness of speech modification algorithm. Although many studies pay great attention to the extent to which modifications can improve speech intelligibility in noise, how the perceptual quality of speech or listener preference to modified speech is affected by modifications is much less researched.

Twenty-six speech modification algorithms were evaluated in the Hurricane Challenges [17,18]. The results suggested that the most efficient modification algorithm can boost intelligibility in noise to the equivalent of increasing the intensity of unmodified speech by 5 dB. However, the perceptual quality of the modified speech was not assessed. In [10,19], the intelligibility of speech modified by the proposed algorithms in noise were evaluated both objectively and subjectively, but the quality of the modified speech in quiet (i.e., SNR =  $+\infty$ ) was only inspected using an objective quality measure—the perceptual evaluation of speech quality (PESQ, [20]). Comparing the intelligibility performance of four algorithms, Taal et al. observed the objective quality of the modified speech in noise using PESQ [12]. They also conducted a listening test to study listeners' preference for different types of modified speech with respect to quality at a SNR level which led to maximal intelligibility, but the perceptual quality in quiet was not reported. Thus, it is worth asking whether listeners have the same preference when listening to modified speech in both noise and quiet.

In this study, the relationship between the intelligibility and quality of speech modified by a range of modification algorithms that boost speech intelligibility is investigated. Two listening experiments were conducted to measure listener's word recognition performance as the perceptual intelligibility and the mean opinion score (MOS) as the quality in both stationary and fluctuating noise maskers, and its MOS in quiet. The results and further implications to the design of modification algorithms in practice are further discussed.

## 2. Speech Modification Algorithms

Two speech modification algorithms, SpecShaping+DRC and SelBoost, which best improved intelligibility in noise in the Hurricane Challenge I [17] were chosen. The third modification, ConstBoost was drawn from [21].

- SpecShaping+DRC consists of two separate stages: spectral shaping in the frequency domain followed by dynamic range compression (DRC) in the time domain [14]. Spectral shaping adaptively enhances the formant information and applies a pre-emphasis filter to the voicing segments. A non-adaptive process is then implemented to avoid loss of high frequency components of the speech signal due to low-pass operations at the stage of signal reconstruction. The DRC stage subsequently applies a time-varying gain to the output of the spectral shaper, in order to increase the perceptual loudness of the output signal (e.g., [22,23]). The DRC used in [14] looks up the gain for each temporal window of 6.7 ms from an input/output envelope characteristic curve. The DRC has a release time of 2 ms and an almost instantaneous attack time for an initial dynamic stage. Its peak threshold is then set to be 30% of the max input speech envelope during the further static stage. Finally, the intensity of the output from the compressor is re-adjusted to that of the original unmodified speech signal. The combination of the two components led to a remarkable performance—SpecShaping+DRC outperformed the other modifications in most of the conditions in the Hurricane Challenge I [17]. Online, modification of this system requires no noise information.

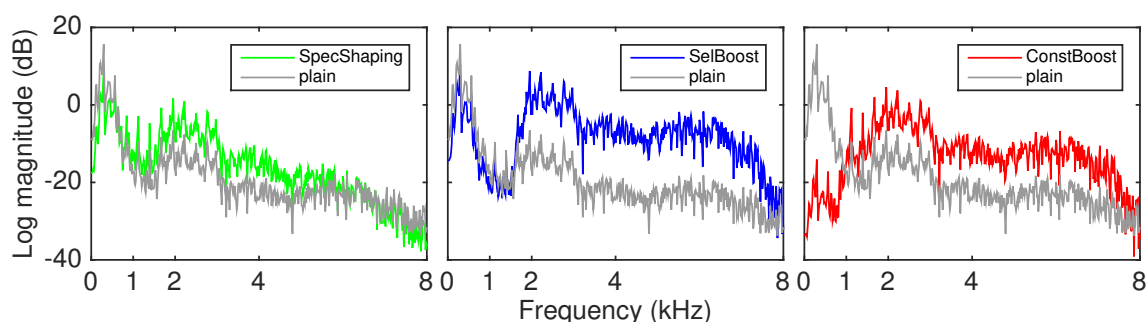
In this study, this modification is treated as two separate systems: SpecShaping only and the original SpecShaping+DRC. The aim is to examine the impact of DRC, as a temporal modification, to speech intelligibility and quality, as well as the combination effect with spectral modifications. Previous objective evaluation using PESQ suggested that temporal modifications appear to negatively affect quality more than spectral modifications [10]. Consequently, the DRC will be imposed on the following two modifications to further form another two modifications.

- SelBoost modifies the speech signal by injecting energy to some time-frequency (T-F) regions between the frequency bands 1800 and 7500 Hz, parts of which are known to be important to

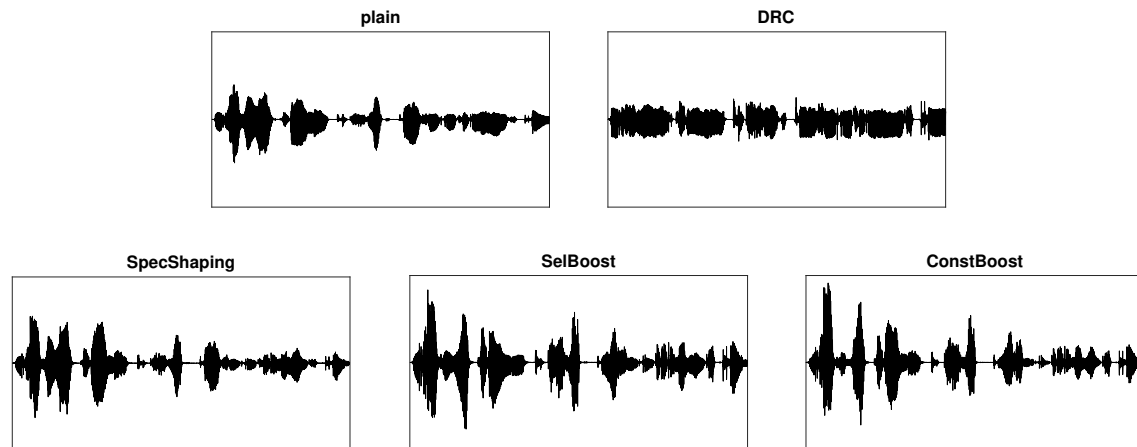
speech intelligibility in noise. The spare energy may come from places where the local SNR is sufficiently high or is less important to speech intelligibility. Two separate optimisation processes were performed in [10], in order to select the T-F regions which are to be boosted. The results of the first optimisation decides the frequency range (1800–7500 Hz) and the boosting amount of 20 dB. The second optimisation, jointly maximising both objective intelligibility and quality, determines the T-F regions within which the local SNR range should be allocated extra energy. It suggests that boosting those regions where local SNR is under or barely above the threshold of audibility (i.e., less than 5 dB) is the effective strategy. As the local SNR of T-F regions needs to be computed as the criterion for boosting, SelBoost requires access to the noise power spectral density. Along with SpecShaping+DRC, SelBoost also demonstrated above-average performance in the stationary masker in the Hurricane Challenge [17].

- ConstBoost is inspired by the optimal spectral weightings found by maximising objective intelligibility metrics [11,21]. In [11], the spectral weightings were derived using a genetic algorithm with the glimpse proportion [24] as the objective function for a range of noise maskers/SNR conditions. It was found that, regardless of the masker type, the suggested weightings always tend to sparsely boost some of the frequencies above 1000 Hz by approximately 10 dB, although the patterns vary in details across maskers. Another attempt was made using a different optimisation algorithm and objective metric in [21]. A similar boosting pattern was observed, but with a boosting amount of 30 dB. Based on these findings, ConstBoost independently imposes a 30 dB gain to all frequency bands above 1000 Hz on the speech, as if applying a high-pass filter to the speech signal. In this way, ConstBoost no longer requires any noise information to operate. After energy renormalisation, the speech energy is effectively transferred to the boosted regions from elsewhere. Further evaluation confirmed that ConstBoost can be as or almost as effective as the noise- and level-dependent spectral weighting in the tested conditions [21].

Figure 1 shows examples of long term average spectra of speech uttered by a male talker, unmodified and spectrally-modified (no DRC applied) by the algorithms introduced above. Figure 2 further illustrates the impact of the modifications to the speech signal in the time domain. Therefore, together with unmodified speech plain as the baseline as well as plain+DRC, the subjective intelligibility and quality of eight types of speech were examined.



**Figure 1.** Long term average spectrum of unmodified and modified speech of a male talker.



**Figure 2.** Waveforms of unmodified and modified utterance of ‘The overweight charmer could slip poison into anyone’s tea’.

### 3. Speech Intelligibility and Quality in Noise

#### 3.1. Experiment Design and Procedure

Subjective intelligibility was measured as the word recognition rate in noise. Ten native British English speakers with normal hearing identified keywords from the Harvard sentences [25], which were uttered by a British male talker. Speech-shaped noise (SSN), babble noise recorded at a cafeteria (BAB) and competing speech (CS) of a female talker were used as the noise maskers. Speech was mixed with each type of noise at two SNR levels (SSN:  $-9$  and  $-3$  dB, BAB:  $-7$  and  $-1$  dB, CS:  $-18$  and  $-12$  dB), forming a low and a high intelligibility condition. The chosen low and high SNR led to subjective recognition rates of approximately 25% and 50% respectively for each noise masker in a pilot test. This setting led to 48 conditions (8 modification  $\times$  3 maskers  $\times$  2 SNRs). Each condition was presented four times, resulting in 192 different sentences being heard by each listener in total. Sentences were divided into six masker/SNR blocks. The presentation order of blocks, and the sentences in each block, were randomised for all listeners.

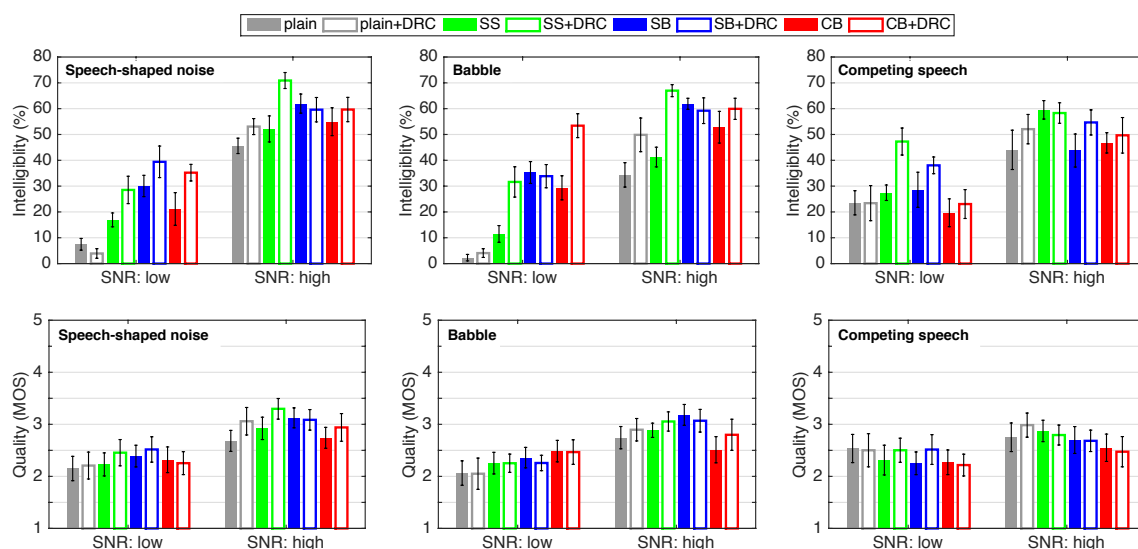
The playback of the stimuli was controlled by a MATLAB graphic programme. Stimuli were presented to listeners monaurally over a pair of Sennheiser (Wedemark, Germany) HD650 headphones after being pre-amplified by a Focusrite (High Wycombe, UK) Scarlett 2i4 USB audio interface. Listeners were allowed to listen to each sentence only once. The experiment took place in a semi-anechoic room with a background noise level lower than 15 dBA. The presentation level of speech was calibrated and fixed to 72 dBA; the noise level was then adjusted to the required SNRs.

After responding to each stimulus by typing what they heard using a physical keyboard, listeners also rated the quality of the sentence. Listeners were only told to give their ‘total impression’ of the speech they heard without being provided with any specific definition of quality, nor any examples of ‘good or bad quality’ as a reference. Because speech quality involves aspects such as intelligibility, pleasantness or naturalness [26], loudness and even listening effort [27], quality rating is rather subjective and very much up to the individual’s judgment. Therefore, we decided to let listeners make a free judgement on how the signal sounded to them according to their listening experience, rather than to guide them to listen for specific aspects of speech quality. The quality rating was performed on the scale of MOS, which falls into the range from 1 to 5. Listeners could choose any number between 1 and 5 using a continuous slider.

### 3.2. Results

The mean subjective intelligibility score across listeners is presented in the upper row of Figure 3. All the three modifications alone (without DRC) improved intelligibility over the unmodified speech plain between 6.6 and 22.6 percentage points in SSN, and between 6.9 and 33.0 percentage points

in BAB across SNR levels. SelBoost and ConstBoost appeared less beneficial in CS than in the other two maskers. Nevertheless, an improvement of between 2.7 and 15.5 percentage points was received across all modifications and SNR levels. Except in the low SNR condition of SSN, DRC alone was more beneficial than harmful to plain. Furthermore, DRC seemed to always yield extra gain in addition to SpecShaping and ConstBoost, especially in SSN and BAB. It boosted the two noise-independent modifications up to 2.9 to 25.7 percentage points over that achieved on their own. SelBoost did not benefit from DRC in BAB and the high SNR condition of SSN. Inclusion of DRC however did not largely compromise the performance of SelBoost in these cases.



**Figure 3.** Subjective speech intelligibility (**upper row**) and quality (**lower row**) in noise. Scores are grouped by SNR level. Error bars indicate  $\pm 1$  standard error. SS, SB and CB denote SpecShaping, SelBoost and ConstBoost, respectively.

A three-way repeated measures ANOVA with within-subjects factors of masker type, SNR level and modification on rationalised arcsine units [28] converted from the keyword scores supported visual impressions: modifications significantly altered the intelligibility of modified speech [ $F(7, 63)=27.58, p < 0.001, \eta^2 = 0.33$ ]. As one of the dominant factors for speech intelligibility in noise, the SNR effect is also significant [ $F(1, 9) = 272.42, p < 0.001, \eta^2 = 0.55$ ]. All significant bi-factor and three-way interactions suggested that the performance of the modifications varies with masker type and SNR level [all  $p < 0.001$ ].

The mean subjective quality in noise across listeners is shown in the lower row of Figure 3. Despite large variation in intelligibility among speech modified by different algorithms, listeners seemed to perceive the quality very similarly at the same SNR condition (low or high) across maskers. The mean quality ratings across modifications for each masker/SNR combination are listed in Table 1.

**Table 1.** Mean mean opinion score (MOS) across modifications at each signal-to-noise ratio (SNR) level, with 95% confidence interval in the parentheses.

	SSN	BAB	CS
SNR: high	3.0 ( $\pm 0.13$ )	2.9 ( $\pm 0.14$ )	2.7 ( $\pm 0.11$ )
SNR: low	2.3 ( $\pm 0.08$ )	2.3 ( $\pm 0.11$ )	2.4 ( $\pm 0.09$ )

A separate ANOVA with the same three main factors was performed on the MOS. The results confirmed that there was no significant main effect of modification type nor noise type, except that of SNR level [ $F(1, 9) = 17.19$ ,  $p < 0.01$ ,  $\eta^2 = 0.23$ ]. The bi-factor interactions between masker type and SNR level [ $F(2, 18) = 4.14$ ,  $p < 0.05$ ,  $\eta^2 = 0.02$ ], and between masker type and modification [ $F(14, 126) = 2.23$ ,  $p < 0.05$ ,  $\eta^2 = 0.03$ ] were significant, but that between SNR level and modification was not. There was no three-way interaction either.

Post-hoc analyses using Fisher's least significant difference (FLSD) were further conducted on the quality ratings separately for masker/SNR combination with the single factor of modification. The results in Table 2 confirmed that listeners provided similar quality ratings to all speech types at the low SNR level in SSN and CS. Speech modified by ConstBoost with or without DRC was rated as being lower quality than some of the others in the remaining conditions, except at the low SNR level in BAB, in which both ConstBoost with and without DRC led to better quality than plain with and without DRC.

**Table 2.** Fisher's least significant difference (FLSD) on MOS in each masker/SNR condition.

	SSN	BAB	CS
SNR: high	0.42	0.46	0.40
SNR: low	0.41	0.39	0.40

## 4. Speech Quality in Quiet

### 4.1. Experiment Design and Procedure

The same 10 participants from the first experiment also rated the speech quality in quiet. As SelBoost is noise-dependent, when noise masker varies the algorithm may affect the speech quality differently according to the objective evaluations using PESQ [10]. Therefore, the quality of speech modified by SelBoost in the three maskers were evaluated separately. In each masker, the modification was performed at 'high' SNR only, in which case the objective quality appears to be somewhat lower than when SNR is low [10]. With plain, SpecShaping and ConstBoost, 12 types of speech were evaluated including the 6 DRC-compressed versions.

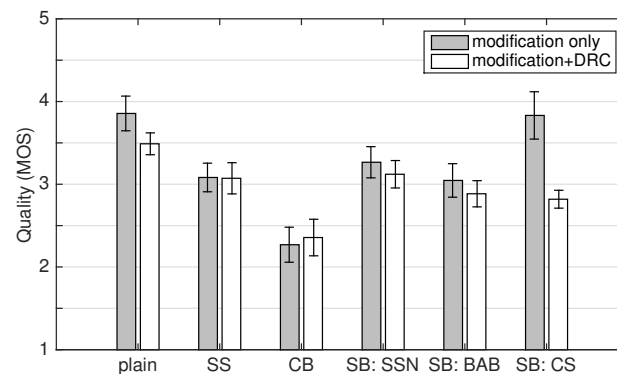
The 12 conditions formed 6 groups, each of which consisted of an unmodified sentence and two modified sentences by the same modification with and without DRC. An identical utterance was used within a group so that the listeners would be able make a relatively fair judgement on the three conditions. Within a group, the participant could choose in which order to play the recordings and repeat them if necessary. Listeners listened to each group three times with different sentences each time (i.e., 18 groups); All the groups were presented to listeners in a random manner.

### 4.2. Results

Figure 4 displays the MOS scores of each modified speech rated by listeners in quiet. Unlike in noise, listeners tended to rate the quality of the modified speech rather differently across modifications, and across the noise maskers in the presence of which the modification was performed for Selboost. Overall, listeners rated plain speech with the highest scores in conditions both with (MOS = 3.9)



and without (MOS = 3.5) DRC, followed by SpecShaping and SelBoost. For SelBoost, the quality was expected to decrease from stationary (SSN) to fluctuating (CS) masker, according to the PESQ predictions in [10]. However, listeners rated the quality without DRC in CS (MOS = 3.8) almost as high as the original unmodified speech.



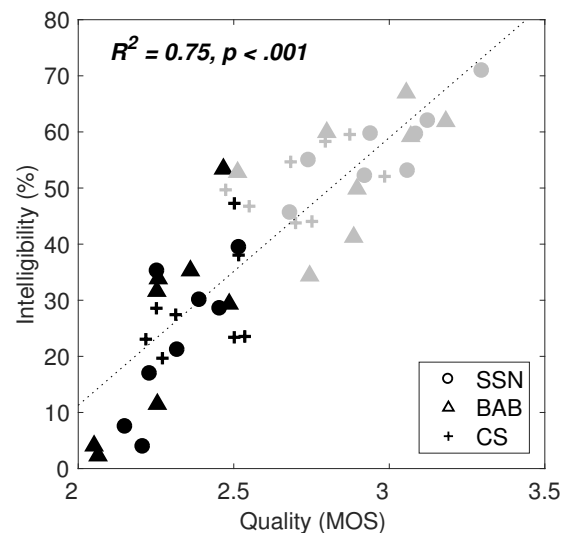
**Figure 4.** Subjective speech quality rating in quiet. Error bars indicate  $\pm 1$  standard error.

A one-way ANOVA analysis revealed a strong effect of modification type [ $F(11, 99) = 10.82$ ,  $p < 0.001$ ,  $\eta^2 = 0.38$ ]. Further FLSD [ $FLSD = 0.4$ ] analysis confirmed that the quality of unmodified speech and speech modified by Selboost in CS is significantly better than that of others. While SpecShaping and SelBoost in SSN and BAB resulted in similar speech quality, ConstBoost led to the poorest speech quality according to listeners' preference. Except for SelBoost in CS, where DRC drastically deteriorated the quality, DRC did not further decrease quality over the modifications.

## 5. Discussion

This study sought to explore the relationship between the intelligibility and quality of speech processed by intelligibility-boosting modification algorithms. In noise, the chosen modifications led to different level of intelligibility. The perceptual quality, however, varied little across most modifications at the same SNR, but did vary significantly when the SNR changed. The results therefore suggest that the noise effect determines speech quality more than do other factors. Another possibility could be that at certain SNR, listeners may not perceive the changes to the speech signal that would lead to a degradation in speech quality. This is likely because when the noise masker masks the audibility of speech, it could simultaneously disguise the artefacts or distortions on the speech due to the modifications, as argued in [12].

Figure 5 further plots the intelligibility scores against the quality rating. It demonstrates a strong positive linear relationship [ $R^2 = 0.75$ ,  $p < 0.001$ ] between speech intelligibility and quality in noise. We interpret this relationship as suggesting that despite the similar quality of the modified speech in noise, speech signals with better intelligibility tend to be rated as higher in speech quality by listeners compared to signals with poorer intelligibility. This may further imply that intelligibility is one of the important factors that listeners use to make a judgement on speech quality in noise. This is consistent with findings in [29] for hearing listeners with normal speech: when intelligibility varies greatly, the quality rating as the total impression of loudness and listening effort can be reflected in listeners' intelligibility performance.



**Figure 5.** Perceptual speech quality versus intelligibility in noise, coded by markers. Black and grey represent low and high SNR, respectively. The overall  $R^2$  and  $p$  value are provided.

When rating speech quality in quiet, where intelligibility is the greatest it can be, listeners mostly preferred plain over modified speech, implying that modifications indeed harm perceptual quality to different degrees. Interestingly, the listeners rated unmodified speech the best in terms of quality even though they were not given any direct reference for quality rating in this study. They seem to have learned from their experiences and have formed a consistent ‘standard’ of quality. However, it is surprising to see that speech processed by SelBoost only in CS (SB: CS) was rated as being as high quality as plain. As the modification needs to adapt to the large fluctuation of CS, a poorer quality would be expected compared to that of SSN and BAB. Thus, the reason behind the listener’s rating in this conditions is unclear here and needs further investigation. Compared to the other conditions, the larger error bar in Figure 4 indicates a larger variation in listeners’ opinions in this condition.

In [12] the quality of the unmodified speech and the speech modified by two independent algorithms was compared, when intelligibility converged in a noisy condition (SSN, 5 dB SNR). It was found that listeners mostly preferred one of the modified speech signals to the unmodified speech signals, which was rated better than the other modified signals. In this case, listeners’ judgement was probably still affected by the noise although intelligibility was no longer an issue at the chosen SNR. In addition, some modifications may introduce greater artefacts or distortions to the speech signal than others. When the SNR is high, the traces left by the modification start to stand out, resulting in low perceived quality—the same may apply to ConstBoost here—despite its significant intelligibility gain in noise. As shown in Figure 1, ConstBoost boosts the mid-high frequencies by sacrificing energy from 1000 Hz below where the pitch and harmonic information exists. Speech signals in which this frequency range is largely attenuated tend to have poor perceptual quality [26].

## 6. Conclusions

By comparing the impact of the speech modifications to speech quality in quiet and intelligibility performance in noise, we revealed that speech with good quality (e.g., plain) in quiet does not necessarily ensure high intelligibility in noise, and vice versa (e.g., ConstBoost). When the SNR is low, speech intelligibility appears to be the dominant factor to the overall perceptual quality, suggesting that speech modification algorithms should be primarily designed for achieving large intelligibility gain in this case. However, when the SNR is high or speech is presented in quiet, there is a tradeoff to make between the intelligibility and quality of the modified speech. This further implies that for the deployment of speech modification techniques in practice, it may be essential to perform instant



SNR estimation online, in order to determine the threshold for modification (de)activation in respect to speech quality.

**Acknowledgments:** This work was supported by the EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1) and the BBC as part of the BBC Audio Research Partnership. Data underlying the findings are fully available without restriction, details are available from <https://dx.doi.org/10.17866/rd.salford2201281>.

**Author Contributions:** Yan Tang and Christopher Arnold designed all the experiments in this study. Christopher Arnold conducted the experiments, and Yan Tang composed this manuscript. All authors approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Nørholm, S.M.; Jensen, J.R.; Christensen, M.G. Enhancement of non-stationary speech using harmonic chirp filters. In Proceedings of the Interspeech, 16th Annual Conference of the International Speech Communication Association, 6–10 September 2015, Dresden, Germany, 2015; pp. 1755–1759.
2. Mohammadiha, N.; Smaragdis, P.; Leijon, A. Supervised and unsupervised speech enhancement using nonnegative matrix factorization. *IEEE Trans. Audio Speech Lang. Process.* **2013**, *21*, 2140–2151.
3. Paliwal, K.; Wójcicki, K.; Schwerin, B. Single-channel speech enhancement using spectral subtraction in the short-time modulation domain. *Speech Commun.* **2010**, *52*, 450–475.
4. Martin, R. Speech Enhancement Based on Minimum Mean-Square Error Estimation and Supergaussian Priors. *IEEE Trans. Speech Audio Process.* **2005**, *13*, 845–856.
5. Rangachari, S.; Loizou, P.C. A noise-estimation algorithm for highly non-stationary environments. *Speech Commun.* **2005**, *48*, 220–231.
6. Brouckxon, H.; Verhelst, W.; Schuymer, B.D. Time and frequency dependent amplification for speech intelligibility enhancement in noisy environments. In Proceedings of the Interspeech, 9th Annual Conference of the International Speech Communication Association, 22–26 September Brisbane Australia, 2008; pp. 557–560.
7. Villegas, J.; Cooke, M. Maximising objective speech intelligibility by local  $f_0$  modulation. In Proceedings of the Interspeech, 13th Annual Conference of the International Speech Communication Association, Portland, OR, USA, 9–13 September 2012; pp. 1704–1707.
8. Godoy, E.; Koutsogiannaki, M.; Stylianou, Y. Assessing the Intelligibility Impact of Vowel Space Expansion via Clear Speech-Inspired Frequency Warping. In Proceedings of the Interspeech, 14th Annual Conference of the International Speech Communication Association, Lyon, France, 25–29 August 2013; pp. 1169–1173.
9. Sauert, B.; Vary, P. Recursive closed-form optimization of spectral audio power allocation for near end listening enhancement. In Proceedings of the ITG-Fachtagung Sprachkommunikation, Bochum, Deutschland, 6–8 October 2010; pp. 955–958.
10. Tang, Y.; Cooke, M. Energy reallocation strategies for speech enhancement in known noise conditions. In Proceedings of the Interspeech, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, 26–30 September 2010; pp. 1636–1639.
11. Tang, Y.; Cooke, M. Optimised spectral weightings for noise-dependent speech intelligibility enhancement. In Proceedings of the Interspeech, 13th Annual Conference of the International Speech Communication Association, Portland, OR, USA, 9–13 September 2012; pp. 955–958.
12. Taal, C.; Hendriks, R.C.; Heusdens, R. Speech energy redistribution for intelligibility improvement in noise based on a perceptual distortion measure. *Comput. Speech Lang.* **2014**, *28*, 858–872.
13. Aubanel, V.; Cooke, M. Information-preserving temporal reallocation of speech in the presence of fluctuating maskers. In Proceedings of the Interspeech, 14th Annual Conference of the International Speech Communication Association, Lyon, France, 25–29 August 2013; pp. 3592–3596.
14. Zorila, T.C.; Kandia, V.; Stylianou, Y. Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression. In Proceedings of the Interspeech, 13th Annual Conference of the International Speech Communication Association, Portland, OR, USA, 9–13 September 2012; pp. 635–638.

15. Godoy, E.; Stylianou, Y. Increasing speech intelligibility via spectral shaping with frequency warping and dynamic range compression plus transient enhancement. In Proceedings of the Interspeech, 14th Annual Conference of the International Speech Communication Association, Lyon, France, 25–29 August 2013; pp. 3572–3576.
16. Schepker, H.; Rennies, J.; Doclo, S. Speech-in-noise enhancement using amplification and dynamic range compression controlled by the speech intelligibility index. *J. Acoust. Soc. Am.* **2015**, *138*, 2692–2706.
17. Cooke, M.; Mayo, C.; Valentini-Botinhao, C.; Stylianou, Y.; Sauert, B.; Tang, Y. Evaluating the intelligibility benefit of speech modifications in known noise conditions. *Speech Commun.* **2013**, *55*, 572–585.
18. Cooke, M.; Mayo, C.; Valentini-Botinhao, C. Intelligibility-enhancing speech modifications: The Hurricane Challenge. In Proceedings of the Interspeech, 14th Annual Conference of the International Speech Communication Association, Lyon, France, 25–29 August 2013; pp. 3552–3556.
19. Tang, Y.; Cooke, M. Subjective and objective evaluation of speech intelligibility enhancement under constant energy and duration constraints. In Proceedings of the Interspeech, 12th Annual Conference of the International Speech Communication Association, Florence, Italy, 27–31 August 2011; pp. 345–348.
20. Rix, A.; Hollier, M.; Hekstra, A.; Beerends, J. Perceptual evaluation of speech quality (PESQ): The new ITU standard for end-to-end speech quality assessment. Part I. Time-delay compensation. *J. Audio Eng. Soc.* **2002**, *50*, 755–764.
21. Tang, Y.; Cooke, M. Learning static spectral weightings for speech intelligibility enhancement in noise. *Comput. Speech Lang.* **2017**. <https://doi.org/10.1016/j.csl.2017.10.003>.
22. Blesser, B.A. Audio dynamic range compression for minimum perceived distortion. *IEEE Trans. Audio Electroacoust.* **1969**, *17*, 22–32.
23. Schmidt, J.C.; Rutledge, J.C. Multichannel dynamic range compression for music signals. In Proceedings of the 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Atlanta, GA, USA, 9 May 1996; pp. 1013–1016.
24. Cooke, M. A glimpsing model of speech perception in noise. *J. Acoust. Soc. Am.* **2006**, *119*, 1562–1573.
25. Rothauser, E.H.; Chapman, W.D.; Guttman, N.; Silbiger, H.R.; Hecker, M.H.L.; Urbanek, G.E.; Nordby, K.S.; Weinstock, M. IEEE Recommended practice for speech quality measurements. *IEEE Trans. Audio Electroacoust.* **1969**, *17*, 225–246.
26. Gabrielsson, A.; Schenkman, B.N.; Hagerman, B. The effects of different frequency responses on sound quality judgments and speech intelligibility. *J. Speech Lang. Hear. Res.* **1988**, *31*, 166–177.
27. Hafter, E.; Schlauch, R. *Noise-Induced Hearing Loss*; Chapter Cognitive Factors and Selection of Auditory Listening Bands; Mosby-Year Book: St Louis, MO, USA, 1992.
28. Studebaker, G.A. A ‘rationalized’ arcsine transform. *J. Speech Hear. Res.* **1985**, *28*, 455–462.
29. Preminger, J.E.; Tasell, D.J.V. Quantifying the relation between speech quality and speech intelligibility. *J. Speech Hear. Res.* **1995**, *38*, 714–725.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).