COMPARISON OF POST-PROCESSING METHODS FOR INTELLIGIBILITY ENHANCEMENT OF NARROWBAND SPEECH IN A MOBILE PHONE FRAMEWORK

Emma Jokinen, Marko Takanen and Paavo Alku

Aalto University, Department of Signal Processing and Acoustics, Espoo, Finland
emma.jokinen@aalto.fi

ABSTRACT

Post-processing methods can be used in mobile communications to improve the intelligibility of speech in adverse background noise conditions. This study addresses the improved intelligibility and the speech quality achieved with a well-known approach, dynamic range compression, by comparing it to two other real-time post-processing methods based on energy reallocation. In addition, the effects of utilizing amplitude normalization instead of energy normalization on the performance of the post-processing methods are investigated. The evaluations were conducted using subjective tests in several background noise conditions. The results indicate that the two energy reallocating approaches outperform dynamic range compression both in intelligibility and quality and that amplitude normalization causes the performance of the tested post-processing methods to degrade in some conditions.

Index Terms— Post-processing, dynamic range compression, telephone speech, narrowband speech, normalization

1. INTRODUCTION

Today, mobile phones are used in increasingly difficult background noise conditions, such as in train stations, restaurants and cars. Furthermore, the corrupting noise in such conditions may be in the sending or receiving side of the channel, referred to as far-end and near-end noise, respectively. To improve communication in noisy conditions, post-processing can be used in the receiving mobile device to enhance the intelligibility of speech. This study is focused on the near-end noise case which means that the decoded speech signal is clean and it is processed to stand out over the environmental noise in the listener's surroundings.

Several intelligibility enhancement methods have been developed for near-end noise scenarios and many of them are based on the optimization of objective measures which are known to correlate with subjective intelligibility. For instance, in [1], optimal gains for the sub-bands of the unprocessed speech signal were determined by maximizing the speech intelligibility index (SII) with constrained audio power. This work was further enhanced by modifying the optimization [2] and combining it with adaptive dynamic range compression (DRC) [3]. In [4], frequency regions with low signal-to-noise ratios (SNRs) were enhanced and this work was extended in [5] to different noise types utilizing offline optimization with the glimpse proportion (GP) measure.

Post-filtering, where an adaptive filter is used to reallocate energy in frequency, can also be utilized for intelligibility enhancement. Traditionally, post-filtering has been employed in improving the perceptual quality by utilizing a filter that emphasizes spectral peaks and attenuates spectral valleys [6, 7]. Alternatively, the traditional post-filter may be replaced with a high-pass type filter that effectively attenuates the low frequencies and enhances the high frequencies, resulting in increased speech intelligibility ([8, 9, 10, 11]).

More sophisticated post-filtering algorithms emulate phenomena that occur naturally when humans are trying to overcome communication barriers. Such phenomena include, e.g., the Lombard effect which is observed when talkers modify their speaking style in an effort to make speech more intelligible in the presence of background noise [12]. It consists of multiple modifications to the speech signal, such as increased vocal intensity, fundamental frequency (F_0), formant frequencies, and word durations, and decreased spectral tilt. In [13], energy reallocation was utilized to transfer energy from voiced sounds to unvoiced utterances. In [14], adaptive spectral shaping, aimed at sharpening the formants and reducing the spectral tilt, was combined with DRC. The approach was shown to improve both objective and subjective intelligibility but is partially based on long-term energy normalization which is not suitable for real-time processing. In [15], the spectral tilt reduction and formant sharpening were combined in a post-filtering method that was shown to improve intelligibility in various noise conditions.

Dynamic range compression is used in audio recording and broadcasting [16], digital hearing aids [17], and mobile phones [18] for fitting the audio into a smaller dynamic range. The reduction is done by amplifying low intensity sounds more than loud sounds [17] which can also be utilized for intelligibility enhancement of speech for normal-hearing listeners because it effectively amplifies unvoiced sounds, similarly as in [13]. However, DRC can cause distortion thus disrupting the quality of the speech signal [17].

The performance of intelligibility enhancement algorithms is also affected by the normalization which is done after the processing, usually in terms of energy. In post-filtering, normalization is particularly important in order to achieve the energy reallocation from low to high frequencies which is the main cause of the intelligibility gain. Several studies focused on the enhancement of speech off-line utilize long-term energy constraints, such as energy normalization over a sentence of speech ([19, 14, 20]) which allows the energy to be moved from one section of speech to another. This approach, however, is not applicable in a real-time system, such as a mobile phone, where frame-based normalization with small delay has to be used. Usually, energy is used as a constraint but considering the restrictions set by the mobile device, i.e., the maximum amplitude of the signal is constrained by the amplifier, an amplitude constraint could be a more realistic choice.

The purpose of this study was (1) to compare, in a subjective test, three previously proposed post-processing methods, DRC and two post-filtering algorithms, which are realistically implementable in mobile devices and which, to the best knowledge of the authors, have

The research leading to these results was supported by the GETA graduate school, Nokia, the Academy of Finland (research programmes 135003 and 13251770), and the Mide/Ui-art project of Aalto University.

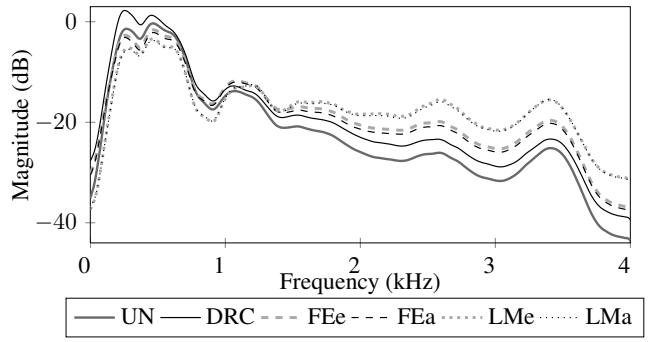

Fig. 1. The long-term average spectra of unprocessed speech and the methods under evaluation (DRC without normalization and the FE and LM post-filters with energy and amplitude normalizations which are denoted by letters e and a, respectively). The spectra have been computed by averaging over all the male speakers in the database used in the study.

not been evaluated jointly in subjective tests previously. The subjective performance evaluation contained both an intelligibility test as well as a quality test which was conducted to study the possible distortions caused by the DRC. A further goal of the study was (2) to evaluate the effects of two frame-based normalization methods, energy and amplitude, on the performance of the different post-filtering methods. The tests were conducted with narrowband speech because it is still prevalent in mobile communications even though wideband speech transmission is becoming increasingly popular. Additionally, intelligibility enhancement is more important for narrowband speech which has fewer speech cues and is thus more severely affected by environmental noise.

2. EVALUATED ALGORITHMS

Three different intelligibility enhancement methods were selected for evaluation: dynamic range compression (DRC) and two post-filtering approaches, the formant equalizing post-filter (FE) and the Lombard-motivated post-filter (LM). The chosen algorithms utilize different techniques that have been shown to improve intelligibility efficiently in previous studies and all of the methods operate with low delay and low computational complexity which makes them implementable in a mobile phone.

DRC reduces the peak-to-average ratio and controls the maximum amplitude of speech whereas both of the post-filtering algorithms are based on energy reallocation from low to high frequencies. Therefore, DRC can be considered as an amplitude normalization method and thus, can be used without additional normalization, but the post-filtering approaches require a level adjustment after the processing. The long-term average spectra (LTAS) of the evaluated algorithms with the different normalization methods used are shown in Fig. 1. In Table 1, the average energies of the different post-processing methods with different normalizations are shown.

2.1. DRC

The DRC method utilized is based on [14] where the authors use a sentence level energy normalization after the compression. However, for the purposes of the current study, the method was adapted to real-time processing by implementing it in frame-based form. In the original study, frame-based processing was used only to compute the envelope of the original sample with different frame lengths for male and female speakers. The frame length used in this study was selected as a compromise between these and was set to 15 ms.

Table 1. The root-mean-square (RMS) energies for DRC without normalization and for the FE and LM post-filters with energy and amplitude normalizations (denoted by letters e and a, respectively). The values are averages over the speech database used in the study normalized with the energy of unprocessed speech.

	DRC	FEe	FEa	LMe	LMa
Male	1.17	1.0	0.94	1.0	0.97
	±0.07	±0.0	±0.03	±0.0	±0.05
Female	1.16	1.0	0.98	1.0	0.93
	±0.07	±0.0	±0.02	±0.0	±0.04

The compression is done in two stages: a dynamic stage and a static stage. In the dynamic stage, the estimated envelope of the signal is smoothed utilizing attack and release time constants adapted to the lower sampling rate of 8 kHz from 16 kHz using the definition given in [21]. In the static stage, a time varying gain is determined based on the decibel value of the smoothed envelope and the input-output envelope characteristic function. The 0 dB reference level needed to determine the decibel value of the envelope was set to 30% of the maximum envelope of the speech database used in the current study.

2.2. FE post-filter

The FE method was introduced by Hall et al. [9]. The algorithm utilizes a fixed high-pass filter which was derived by inverting the average amplitudes of the first two formants measured from adult male speakers. The resulting filter attenuates the frequency range around the first formant with maximum attenuation near 360 Hz. The filter was originally intended for wideband speech with a 22.05-kHz sampling frequency but it was modified for narrowband speech using the z transform given in the original paper [9].

2.3. LM post-filter

The LM post-filter was introduced in [11]. The algorithm aims at modelling spectral changes observed in natural Lombard speech and consists of three parts: spectral tilt compensation with a linear prediction (LP) based approach, formant sharpening, and noise-adaptive high-pass filtering. In the original study, energy normalization was utilized as the final block of the algorithm but for the purposes of this study it was removed.

The first two parts of the algorithm were used directly with parameter values found in the original study [11]. However, the determination of the parameter controlling the smoothing of the high-pass filter depending on the noise was simplified. It was set such that there was no smoothing in severe noise conditions and some smoothing in moderate noise conditions.

3. NORMALIZATION METHODS

Normalization is utilized after the post-processing to control the level of the processed signal. This is especially important with post-filtering where the signal is usually attenuated by the filtering. In frame-based normalization the processed frame is equalized to the level of the original frame in terms of a specific measure. For energy normalization, the measure is the energy of the frame whereas for

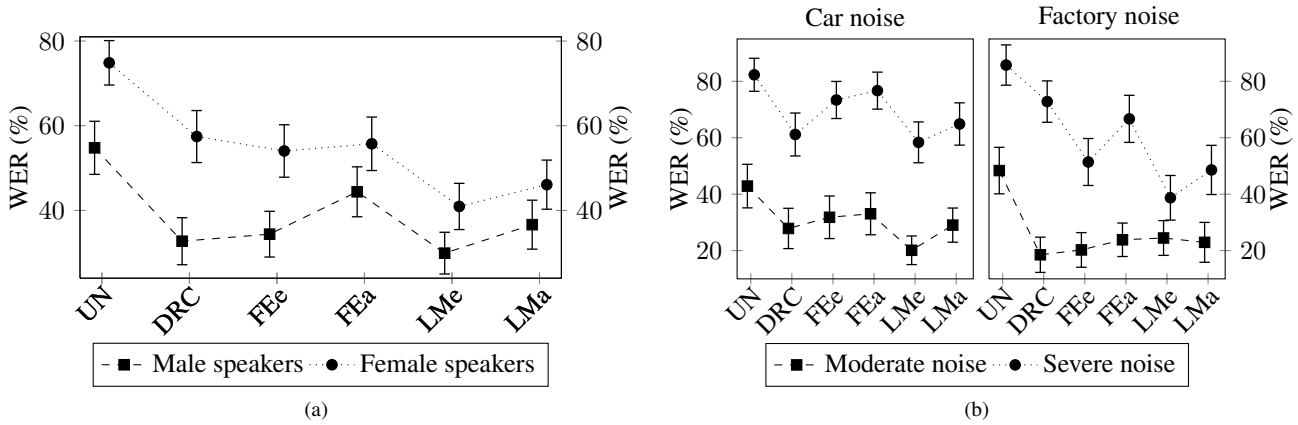

Fig. 2. Results of the intelligibility test (a) for male and female speakers averaged over all noise conditions and (b) averaged over speakers for car and factory noise with moderate and severe noise levels (moderate: car -5 dB, factory 0 dB, severe: car -10 dB, factory -5 dB). The methods under comparison were unprocessed speech (UN), dynamic range compression without normalization (DRC), the formant equalizing post-filter with energy (FEe) and amplitude (FEa) normalization as well as the Lombard-motivated post-filter with energy (LMe) and amplitude (LMa) normalization.

amplitude normalization, it is the maximum amplitude. The two normalization methods are conceptually very different but in practice the difference is small as can be seen from Fig. 1 and Table 1.

3.1. Energy normalization

The energy normalization is the adaptive gain control (AGC) of the adaptive multi-rate (AMR) speech codec [22] with non-overlapping frames of 10 ms. First, a gain factor, γ_{E}, is computed for the present frame and it is updated sample-by-sample with $\beta_{\mathrm{E}}(n) = \alpha_{\mathrm{E}}\beta_{\mathrm{E}}(n-1) + (1-\alpha_{\mathrm{E}})\gamma_{\mathrm{E}}$, where $\beta_{\mathrm{E}}(n)$ is the final scaling factor for sample n and $\alpha_{\mathrm{E}} = 0.9$ as in [22].

3.2. Amplitude normalization

For the amplitude normalization, 10 ms frames are used. First, the maximum amplitude of the processed signal is compared with the original signal and a scaling factor is determined. Similarly with the energy normalization, each sample is then multiplied with an updated scaling factor $\beta_{\mathrm{A}}(n) = \alpha_{\mathrm{A}}\beta_{\mathrm{A}}(n-1) + (1-\alpha_{\mathrm{A}})\gamma_{\mathrm{A}}$, where γ_{A} is the original amplitude scaling factor for the frame and α_{A} was chosen to be 0.9. The value of α_{A} is based on informal listening where amplitude scaling was done after compression type processing. The selected value provided a smooth transition between frames without audible distortion.

4. SUBJECTIVE EVALUATION

A subjective test was conducted to evaluate the performances of the different post-processing and normalization methods. The test consisted of two parts: an intelligibility evaluation, a word-error rate (WER) test, followed by a pair comparison test with questions on overall quality and listening preference. In the WER test, clean speech was corrupted with two types of additive noise (stationary car noise and unstationary factory noise [23]) each with two SNR levels which were selected based on informal listening to create noise conditions characterized as moderate, and severe. The SNR levels for car noise were -5 dB and -10 dB and for factory noise 0 dB and -5 dB. The methods under evaluation in the WER test were unprocessed speech (UN), DRC without normalization (DRC) and the FE and LM post-filters with energy (FEe and LMe, respectively) and amplitude normalizations (FEa and LMa, respectively). For the pair comparison test, the methods had to be normalized to make comparison possible. Therefore, the post-processing methods (DRC, FE, and LM) were all used with energy normalization which was done with SV56 [24, 25].

The speech material consisted of phonetically balanced sentence material from two male and two female speakers which has been calibrated in terms of intelligibility in a previous study [26]. The speech material developed in [26] consists of meaningful sentences both in Finnish and English but for the present study only the Finnish material was used. The sentences contained 4-5 words and had an average duration of approximately 2 seconds. All speech samples were first downsampled to 16 kHz, filtered with the MSIN filter [24] to simulate mobile station input characteristics, downsampled to 8 kHz and AMR encoded and decoded [27]. After this, the samples were equalized to -26 dBov with SV56 [24, 25], processed with one of the methods (DRC, FE, LM), and normalized using amplitude or energy normalization. Finally, car or factory noise was added according to the noise condition.

12 normal-hearing listeners, all native speakers of Finnish, participated in the listening tests. The tests took place in a soundproofed listening booth with Sennheiser HD 650 headphones. The test was divided into two parts and a short practice session preceded each part. During the practice, the listeners were able to adjust the volume to a comfortable listening level. For the test, the volume setting was kept constant.

In the WER test, the subjects were allowed to play each sample only once after which they typed the sentence on the computer. The percentage of correct words was computed by scoring the stems and suffices of inflected words separately after obvious spelling errors had been corrected. In the pair comparison test, the listeners were able to freely listen to two samples, A and B, and were asked to answer the following questions:

Q1: Which sample is of better quality?

Q2: Which sample would you prefer to listen to?

They could answer by selecting one of the options: *A*, *B* or *No difference* and were instructed to select *No difference* if they had no preference even if they heard a difference between the samples.

Table 2. Pairwise comparison between methods in terms of overall quality (Q1) and listening preference (Q2). The methods have all been normalized with SV56 [24, 25] to make comparison possible. Only pairs with statistically significant differences are shown. The preferred method is highlighted with the letters in boldface.

	Speaker	Comparison	W	p
Q1	Male	**UN**-DRC	-5.02	0.00
		DRC-**FE**	3.18	0.00
		DRC-**LM**	2.90	0.00
		FE-LM	-3.81	0.01
	Female	**UN**-DRC	-5.61	0.00
		UN-**FE**	3.20	0.00
		DRC-**FE**	6.65	0.00
		DRC-**LM**	6.93	0.00
Q2	Male	**UN**-DRC	-5.02	0.00
		UN-LM	-3.18	0.00
		FE-LM	-2.54	0.01
	Female	**UN**-DRC	-5.86	0.00
		DRC-**FE**	6.93	0.00
		DRC-**LM**	4.63	0.00

The results of the WER test were analyzed with a five-way analysis of variance (ANOVA) procedure using 5% significance level. The test subject was modelled as a random factor while the noise type (car noise, factory noise), the SNR level (moderate, severe), the method (UN, DRC, FEe, FEa, LMe, and LMa), and the speaker gender (male, female) were modelled as fixed factors. The normality of the residuals and the normality of the random effects were verified using the one-sample Kolmogov-Smirnov test with the significance level of 5%. The ANOVA indicated that the noise type [F(1,14) = 23.68, $p < 0.001$], the method [F(5,70) =30.60, $p < 0.001$], the SNR level [F(1,14) = 724.00, $p < 0.001$], the speaker gender [F(1,14) = 140.10, $p < 0.001$] as well as the interactions between the noise type and the method [F(5,70) = 5.76, $p < 0.001$], between the SNR level and the method [F(5,70) = 4.36, $p < 0.01$], between the method and the speaker gender [F(5,70) = 4.36, $p < 0.01$], and between the noise type, the SNR level and the method [F(5,70) = 5.10, $p < 0.01$] had a significant effect on the WER scores.

In order to gain more insight into the nature of the effects, the marginal means and the 95% confidence intervals were computed, and Dunnett's T3 post-hoc test with the significance level of 5% was applied to confirm the statistical significance of the findings. The values shown in Fig. 2(a) illustrate that the WER scores were, on average, slightly lower with male speakers. Additionally, the difference between UN and the other methods was larger with female speakers. Moreover, LMe had a lower WER score than UN, DRC and FEa with female speakers. The difference between LMe and FEe failed to reach statistical significance.

On the other hand, the values shown in Fig. 2(b) illustrate that the differences between the methods were, on average, larger in the factory noise conditions, and that the decrease in SNR level resulted in more pronounced differences between the methods as well as in increased WER scores, on average. Interestingly, only LMe provides a better WER score than UN in all noise conditions. Furthermore, LMe performs also over the DRC and FEa in the severe factory noise. The differences between LMe and FEe, and between LMe and FEa in the severe car noise were not statistically significant.

The responses given in the pair comparison test were arranged into preference matrices and the Bradley-Terry method was used to fit generalized linear models to the data obtained. Thereafter, a two-way ANOVA procedure was used to test the dependence of the preference score from the method-pair under comparison or from the gender of the speaker. For Q1, the method-pair [χ^2 = 100.22, d.f. = 3, $p < 0.001$], and the interaction between the method-pair and the speaker gender [χ^2 = 30.42, d.f. = 3, $p < 0.001$] had a significant effect on whether one of the compared samples was chosen. The same factors ([χ^2 = 73.13, d.f. = 3, $p < 0.001$], and [χ^2 = 24.71, d.f. = 3, $p < 0.001$]) affected also Q2.

The pair comparisons were further analyzed in a pairwise manner in order to obtain detailed knowledge on whether a particular method was preferred significantly over another method. These analyses were performed using Barnard's exact test with the significance level of 5%. Table 2 summarizes the results for the comparisons in which one method was significantly preferred over the other. Inspection of the results reveals that UN was always preferred over DRC both in terms of Q1 and Q2. Moreover, FE and LM were also always preferred in terms of Q1 over DRC, although in terms of Q2, their preferences over DRC were significant only with female speakers. Interestingly, it seems that FE is preferred in terms of both Q1 and Q2 over LM with male speakers, but not with female speakers. In contrast, detailed inspection of the results revealed that the test subjects had a tendency to prefer LM over FE when female speech was used, but the preference failed to reach a statistical significance.

5. DISCUSSION

Dynamic range compression (DRC) was compared to two other intelligibility-enhancing post-processing methods (FE and LM post-filters) in a WER test with two noise types and multiple noise conditions and in a pair comparison test with clean speech. Additionally, two frame-based normalization methods, energy and amplitude normalization, were utilized in the intelligibility test to evaluate their impact on the performance of the post-processing algorithms.

The results of the WER test indicate that energy-normalized LM-post-filter (LMe) was the only post-processing method that improved the speech intelligibility compared to unprocessed speech (UN) in all of the noise conditions. DRC, which in advance was expected to provide the largest intelligibility increase because the processed speech had on average much more energy, did not outperform the other methods. On the other hand, in the pair comparison test, DRC was rated worse in almost all of the comparisons, in terms of both overall quality and listening preference. The most likely reason for this is that DRC contained distortions and lacked the clarity provided by the other methods. Interestingly, FE was rated over LM for both questions in the case of male speakers but there were no differences for female speakers. This can be explained by noting that the spectral tilt compensation in LM is based on LP which has difficulties in modelling speech with high F_0 [28]. Therefore, the tilt estimation likely follows the true envelope of the spectrum better for male speakers leading to more efficient spectral tilt compensation at high frequencies than in the case of female speakers.

Whether energy or amplitude normalization is used has a clear, although not statistically significant, impact on the performance of the two post-filtering methods. The results indicate that energy normalization provides larger intelligibility gains even though the difference between the normalization methods in terms of RMS energy of the processed speech is relatively small. Regardless, amplitude normalization is a well-justified approach from a device perspective and could be useful with post-processing methods that knowingly take advantage of the different restriction in their design.

6. REFERENCES

[1] B. Sauert and P. Vary, "Recursive closed-form optimization of spectral audio power allocation for near end listening enhancement," in *ITG-Fachtagung Sprachkommunikation*, 2010.

[2] C.H. Taal, J. Jensen, and A. Leijon, "On optimal linear filtering of speech for near-end listening enhancement.," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 225–228, 2013.

[3] H. Schepker, J. Rennies, and S. Doclo, "Improving speech intelligibility in noise by SII-dependent preprocessing using frequency-dependent amplification and dynamic range compression," in *Proc. Interspeech*, 2013, pp. 3577–3581.

[4] Y. Tang and M. Cooke, "Subjective and objective evaluation of speech intelligibility enhancement under constant energy and duration constraints," in *Proc. Interspeech*, 2011, pp. 345–348.

[5] Y. Tang and M. Cooke, "Optimised spectral weightings for noise-dependent speech intelligibility enhancement," in *Proc. Interspeech*, 2012.

[6] J.-H. Chen and A. Gersho, "Adaptive postfiltering for quality enhancement of coded speech," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 1, pp. 59–71, 1995.

[7] V. Grancharov, J.H. Plasberg, J. Samuelsson, and W.B. Kleijn, "Generalized postfilter for speech quality enhancement," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 1, pp. 57–64, 2008.

[8] R.J. Niederjohn and J.H. Grotelueschen, "The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, pp. 277–282, 1976.

[9] J.L. Hall and J.L. Flanagan, "Intelligibility and listener preference of telephone speech in the presence of babble noise," *J. Acoust. Soc. Amer.*, vol. 127, no. 1, pp. 280–285, 2010.

[10] E. Jokinen, S. Yrttiaho, H. Pulakka, M. Vainio, and P. Alku, "Signal-to-noise ratio adaptive post-filtering method for intelligibility enhancement of telephone speech," *J. Acoust. Soc. Amer.*, vol. 132, no. 6, pp. 3990–4001, 2012.

[11] E. Jokinen, M. Takanen, M. Vainio, and P. Alku, "An adaptive post-filtering method producing an artificial Lombard-like effect for intelligibility enhancement of narrowband telephone speech," *Computer, speech and language*, vol. 28, no. 2, pp. 619–628, 2014.

[12] W. Van Summers, D.B. Pisoni, R.H. Bernacki, R.I. Pedlow, and M.A. Stokes, "Effects of noise on speech production: Acoustic and perceptual analyses," *J. Acoust. Soc. Amer.*, vol. 84, no. 3, pp. 917–928, 1988.

[13] M.D. Skowronski and J.G. Harris, "Applied principles of clear and Lombard speech for automated intelligibility enhancement in noisy environments," *Speech Commun.*, vol. 48, no. 5, pp. 549–558, 2006.

[14] T.-C. Zorilă, V. Kandia, and Y. Stylianou, "Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression," in *Proc. Interspeech*, 2012.

[15] E. Jokinen, P. Alku, and M. Vainio, "Lombard-motivated postfiltering method for the intelligibility enhancement of telephone speech," in *Proc. Interspeech*, 2012.

[16] D. Giannoulis, M. Massberg, and J.D. Reiss, "Digital dynamic range compressor design - a tutorial and analysis," *J. Audio Eng. Soc*, vol. 60, no. 6, pp. 399–408, 2012.

[17] P.E. Souza, "Effects of compression on speech acoustics, intelligibility, and sound quality," *Trends Amplif.*, vol. 6, no. 4, pp. 131–165, 2002.

[18] Janne Kivinen, Ari Koski, Jari Sjöberg, and Mauri Vaananen, "Method and circuit arrangement for adjusting the level or dynamic range of an audio signal," US Patent US 5907823 A, issued May 25, 1999.

[19] D. Erro, T.C. Zorilă, Y. Stylianou, E. Navas, and I. Hernaez, "Statistical synthesizer with embedded prosodic and spectral modifications to generate highly intelligible speech in noise," in *Proc. Interspeech*, 2013, pp. 3557–3561.

[20] A. Suni, R. Karhila, T. Raitio, M. Kurimo, M. Vainio, and P. Alku, "Lombard modified text-to-speech synthesis for improved intelligibility: Submission for the Hurricane challenge 2013," in *Proc. Interspeech*, 2013, pp. 3562–3566.

[21] B.A. Blesser, "Audio dynamic range compression for minimum perceived distortion," *IEEE Trans. Audio Electoacoust.*, vol. AU-17, no. 1, pp. 22–32, 1969.

[22] 3rd Generation Partnership Project, Valbonne, France, *Adaptive multi-rate (AMR) speech codec; Transcoding functions*, 2008, version 8.0.0.

[23] A. Varga and H.J.M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, vol. 12, no. 3, pp. 247–251, 1993.

[24] Int. Telecommun. Union, Geneva, Switzerland, *Recommendation G.191: Software tools for speech and audio coding standardization*, September 2005.

[25] Int. Telecommun. Union, Geneva, Switzerland, *Recommendation P.56: Objective measurement of active speech level*, March 1993.

[26] M. Vainio, A. Suni, H. Järveläinen, J. Järvikivi, and V.-V. Mattila, "Developing a speech intelligibility test based on measuring speech reception thresholds in noise for English and Finnish," *J. Acoust. Soc. Amer.*, vol. 118, no. 3, pp. 1742–1750, 2005.

[27] 3rd Generation Partnership Project, Valbonne, France, *ANSI-C code for the floating-point Adaptive Multi-Rate (AMR) speech codec*, 2009, version 9.0.0.

[28] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," *IEEE Trans. Signal Process.*, vol. 39, no. 2, pp. 411–423, 1991.