ON SPEECH QUALITY ESTIMATION OF PHASE-AWARE SINGLE-CHANNEL SPEECH ENHANCEMENT

Andreas Gaich, Pejman Mowlaee

Signal Processing and Speech Communication Lab
Graz University of Technology, Austria
gaich@student.tugraz.at, pejman.mowlaee@tugraz.at

ABSTRACT

To approximate the speech quality of a given speech enhancement system, most of the existing instrumental metrics rely on the calculation of a distortion metric defined between the clean reference signal and the enhanced signal in the spectral amplitude domain. Several recent studies have demonstrated the effectiveness of employing a phase modification stage in single-channel speech enhancement showing positive impact brought by modifying both amplitude and phase in contrast to the conventional methods where the noisy spectral amplitude is only modified and noisy phase is used for signal reconstruction. In this work we present two contributions; First we study the reliability of the existing instrumental metrics for performance evaluation of phase-aware methods, and second we propose novel phase-aware instrumental metrics and evaluate their reliability in terms of predicting the perceived quality achieved by the phase-aware methods. Our objective and subjective evaluations demonstrate that PESQ and the proposed phase deviation metric perform as reliable speech quality estimators following the subjective results.

***Index Terms*—** Phase estimation, perceived speech quality, phase-aware speech enhancement, subjective listening.

1. INTRODUCTION

The task of quality estimation of speech enhancement systems is of great importance in the development of new methodologies e.g. in mobile communication devices and hearing aids. The ultimate success of a method in these applications highly depends on a robust performance in a noisy environment and on how well it handles the background noise while balancing a trade-off between the remaining residual noise and the amount of introduced speech distortion. To balance such a trade-off, it is crucial to find a fast and reliable performance evaluation tool. Figure 1 demonstrates the configuration for the conventional instrumental metrics used for performance evaluation in speech enhancement [1, Ch. 10].

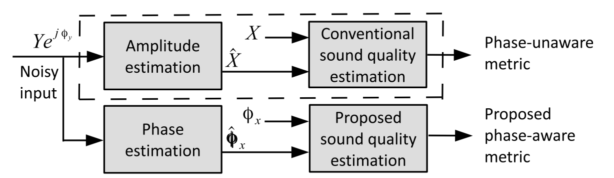

Fig. 1. Conventional (dashed box) versus the proposed evaluation methodology for speech quality estimation.

Although many different instrumental metrics have been proposed, still their reliability has only been demonstrated for limited scenarios including spectral amplitude-only enhancement. As an example, PESQ was shown as a reliable metric to estimate the perceived quality of speech enhancement methods [2]. However, the studied methods only modified the noisy spectral amplitude and copied the noisy phase at the signal reconstruction stage.

Therefore, in this paper, we address two questions; First the reliability of PESQ and other existing instrumental metrics to predict the subjective results achieved by phase-aware enhancement schemes where the spectral phase is also modified (in contrast to the conventional amplitude enhancement schemes). This is inspired by the recent works [3–13] showing improvement in the perceived speech quality by modification of the noisy phase. As our second contribution, we propose new phase-aware instrumental metrics to emphasize on the phase importance in quality estimation of a phase-aware speech enhancement method. Throughout objective and subjective evaluations in various noise scenarios, we evaluate the reliability of the metrics in predicting the subjective results.

The rest of the paper is organized as follow. Section II presents an overview of the conventional metrics. Section III presents the problem definition as the performance evaluation of phase-aware speech enhancement. Section IV presents our proposed metrics for phase-aware performance evaluation in speech enhancement. Section V presents the subjective listening results and statistical analysis to find the correlation to the studied metrics. Section VI concludes on the work.

We thank Rahim Saeidi for his productive discussions. The work was partially funded by the K-Project ASD in the context of COMET Competence Centers for Excellent Technologies by BMVIT, BMWFJ, Styrian Business Promotion Agency (SFG), Province of Styria Government of Styria and the Technology Agency of the City of Vienna (ZIT). The programme COMET is conducted by Austrian Research Promotion Agency (FFG).

2. CONVENTIONAL METRICS

In general, one can divide the existing quality metrics into three groups: SNR-based, speech codec metrics and source separation metrics. Some examples for the SNR-based metrics are: global SNR (GSNR) [14], segmental SNR (SSNR) [1] and frequency weighted SNR (fwSNR) [15]. The SNR-based metrics emphasize on a sample-by-sample comparison. The metrics in the second group were originally proposed to evaluate the performance of a speech codec. Some examples are: log-likelihood ratio (LLR) [16], cepstral distance (CEPS) [17], Itakura-Saito distance (ISa) [18] and perceived speech quality estimation (PESQ) [19]. There the basic idea was to assume that speech follows an auto-regressive process within the short time frames modeled by linear prediction with some psycho-acoustical model on top as it was taken into account in PESQ. The types of distortions quantified by these metrics when used in speech coding are however different than those introduced by a speech enhancement method. For example, the original waveform is distorted due to quantization errors introduced by code excited linear prediction (CELP) while in speech enhancement, in contrast, musical noise as well as speech distortion are common.

The last group is focused on evaluating the performance of a separation algorithm applied on audio mixed signals. A commonly used metric is blind source separation evaluation (BSS EVAL) [20] which consists of three different SNR-based metrics: Signal-to-Distortion Ratio (SDR), Signal-to-Interference Ratio (SIR) and Signal-to-Artifact Ratio (SAR).

3. PROBLEM DEFINITION AND MOTIVATION

3.1. Notations and Problem Definition

Let $y(n) = x(n) + v(n)$ be the noisy signal with $x(n)$ and $v(n)$ denoting the clean and noise signals, respectively. The noisy signal is processed by a speech enhancement algorithm producing the speech enhanced signal $\hat{x}(n)$. Let $Y^c(k,l)$, $X^c(k,l)$, $\hat{X}^c(k,l)$ and $V^c(k,l)$ be the STFT transforms for noisy, clean, enhanced speech and noise signals, respectively, with k and l as the frequency and time indices. The complex spectrum $X^c(k,l)$ consists of spectral amplitude and spectral phase $X^c(k,l) = X(k,l)e^{j\phi_x(k,l)}$ with $X(k,l)$ as the amplitude and $\phi_x(k,l) = \angle X^c(k,l)$ as the spectral phase.

Figure 1 shows the difference between the conventional performance evaluation relying on the spectral amplitude difference ($X(k,l)$ versus $\hat{X}(k,l)$) and our proposed metrics relying on the spectral phase values i.e. $\phi_x(k,l)$ and $\hat{\phi}_x(k,l)$.

3.2. Why Phase-Aware Metrics?

Here we address the incompleteness of the existing instrumental metrics in the speech quality estimation of a phase-aware method. To this end, we present a counterexample to demonstrate why a new phase-aware metric is required. Figure 2 demonstrates the outcome for phase-enhancement using STFT phase improvement (STFTPI) recently proposed in [13]. The method relies on phase reconstruction at harmonics given the fundamental frequency estimate. The signal components between harmonics are entirely removed while strict harmonic structure is enforced across frequency. The improvement in PESQ and fwSNR is obtained at the expense of buzzy speech quality, also reported in [12,21]. The buzzyness is visible by comparing the harmonic structure in the phase-enhanced signal versus the clean original signal as shown in Figure 2. A similar trend is visible in the group delay plot. This counter example and the observed improvement in the existing metrics (PESQ and fwSNR shown in Figure 2) motivates us to address the following two research questions:

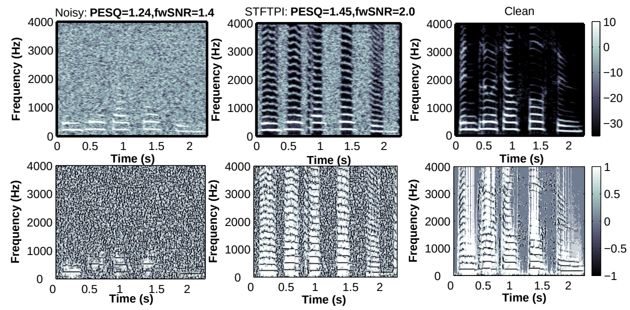

Fig. 2. Counter example: (top) spectrogram (bottom) group delay plots shown for (left) noisy phase unprocessed signal, (middle) phase-enhanced signal using STFTPI [13], (right) clean signal for white noise at 0 (dB).

1. how much the existing metrics (PESQ, fwSNR, $\cdots$) correlate with subjective results for phase-aware speech enhancement,
2. whether some new phase-aware metrics could outperform the existing ones in terms of predicting the subjective listening results.

4. PROPOSED PHASE-AWARE METRICS

4.1. Group Delay (GD)

The group delay is the negative derivative of the spectral phase with respect to frequency and in the discrete domain is defined as

$$\tau(k) = -\Delta_k \phi(k,l) = \phi(k,l) - \phi(k-1,l). \tag{1}$$

As summarized in [5], GD has been reported useful in various speech processing applications. As our phase-aware instrumental metric, here we propose the following distortion metric:

$$\boxed{d_{GD} = \Big(\cos\left(-\Delta_k \phi_x(k,l)\right) - \cos(-\Delta_k \hat{\phi}_x(k,l))\Big)^2} \tag{2}$$

which was first used in [9] and presented in more detail in [6] to resolve the ambiguity in phase estimation for single-channel speech enhancement. To make the distance metric invariant to module of 2π and to avoid wrong error calculations due to the periodicity of phase, in the proposed metric, we employed the cosine function. A similar treatment was employed for phase-based estimators studied in [22] as well as in deriving an estimator for the spectral phase in [23].

4.2. Instantaneous Frequency Deviation (IFD)

In [24, 25], the concept of Instantaneous Frequency Deviation (IFD) was introduced as a useful interpretation of the short-time spectral phase defined as the first-order time-derivative:

$$\mathrm{IFD}_{\phi}(k,l) = \frac{1}{2\pi}\left(\phi(k,l) - \phi(k,l-1)\right) - k. \qquad (3)$$

It was shown that IFD carries information about the vocal-tract excitation [26] and that it is a useful representation for pitch estimation or automatic speech recognition as it resolves the formant frequencies [24, 27].

As our second phase-aware instrumental metric, here we propose the following distortion metric:

$$\boxed{d_{IFD} = \Big(\cos\left(\mathrm{IFD}_{\phi_x}(k,l)\right) - \cos\Big(\mathrm{IFD}_{\hat{\phi}_x}(k,l)\Big)\Big)^2} \qquad (4)$$

which was recently used in [6] to resolve the ambiguity in phase estimation for single-channel speech enhancement.

4.3. Phase Deviation (PD)

A geometric representation of the phase deviation concept is shown in Figure 3. It is defined as the deviation between the noisy and clean phase spectra given by:

$$\phi_{\text{dev}}(k,l) = \phi_y(k,l) - \phi_x(k,l). \qquad (5)$$

Vary first defined phase deviation in [28] and the concept was later employed in [7] for phase estimation and [29] for joint noise reduction and echo cancellation. Here, we propose to employ the PD concept as a new distortion metric, defined as:

$$\boxed{d_{PD} = \Big(\cos(\phi_{\text{dev}}(k,l)) - \cos(\hat{\phi}_{\text{dev}}(k,l))\Big)^2} \qquad (6)$$

where we define $\hat{\phi}_{\text{dev}}(k,l) = \phi_y(k,l) - \hat{\phi}_x(k,l)$ as the estimated phase deviation given the estimated speech phase.

4.4. Mean Square Error (MSE) of Phase

In estimation theory, the mean square error (MSE) is commonly chosen to quantify the amount of estimation error introduced by an estimator [30], and is described as below:

$$\boxed{d_{MSE} = \Big(\cos(\phi_x(k,l) - \hat{\phi}_x(k,l))\Big)^2} \qquad (7)$$

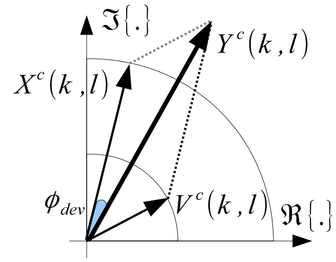

Fig. 3. Geometric representation for the single-channel speech enhancement problem; showing noisy, clean and noise complex spectra denoted by $Y^c(k,l)$, $X^c(k,l)$ and $V^c(k,l)$, respectively. The phase deviation ϕ_{dev} is shown as the phase difference between the clean and the noisy speech signal.

5. RESULTS AND DISCUSSION

5.1. Experimental Setup

As the test material, we selected 50 utterances from the GRID corpus [31] composed of male and female speakers downsampled to 8 kHz. White and babble noise were selected from NOISEX-92 [32] and were added to the clean speech signals at SNRs of 0, 5 and 10 dB. The noisy files (unprocessed, UP) were processed by four speech-enhancement algorithms: *Conventional (C)* (MMSE-LSA [33]) *, Conventional + Clean phase (C + clean), Conventional + Phase-Enhanced (PE)* [3] and *Phase-Aware (PA)* [7]. Including the unprocessed files, we had an overall number of 1500 speech files in the analysis. The results of the instrumental metrics were averaged over all utterances for each method and SNR.

5.2. Subjective Listening Test

The subjective listening test was conducted in a quiet room. AKG K-240 Studio Headphones and a Hoontech DSP24Value 24Bit/96kHz soundcard as audio interface were used. Following the MUSHRA standard [34], we included a Hidden Reference as well as an Anchor (defined as the 2.5 kHz low-pass filtered reference signal). A panel of 11 listeners were recruited for the subjective test, all experienced listeners at the Graz University of Technology. For each noise type and SNR two randomly selected utterances were presented to each participant.

Figure 4 shows the Mean Opinion Scores (MOS) and 95% confidence intervals differentiated in terms of noise type and SNR. For all noise types and SNRs, similar rankings were observed where the PA method performed best followed by the C + Clean phase method. The PE method was ranked as the third with a short gap with respect to the C method. The UP, as expected, had the lowest ranking. T-tests were conducted to justify the significance of these rankings. Except between the C and the PE method, all other rankings were significant with respect to each other with $p < 0.05$. However, PE outperforms C significantly for white noise scenario

	PESQ [19]			**ISa** [18]			**GD** [proposed]			**IFD** [proposed]			**PD** [proposed]			**MSE**		
	ρ	σ	τ	ρ	σ	τ	ρ	σ	τ	ρ	σ	τ	ρ	σ	τ	ρ	σ	τ
Noise=Babble	0.87	0.07	0.77	0.87	0.07	-0.73	0.84	0.07	-0.70	0.84	0.07	-0.68	**0.91**	0.06	-0.77	0.83	0.08	0.70
Noise=White	0.86	0.07	0.79	0.91	0.06	-0.75	**0.96**	0.04	-0.87	0.89	0.07	-0.81	0.94	0.05	-0.89	0.90	0.06	0.81
SNR=0(dB)	0.91	0.05	0.91	0.89	0.05	-0.73	0.81	0.07	-0.69	0.87	0.06	-0.87	**0.92**	0.05	-0.87	0.87	0.06	0.78
SNR=5(dB)	**0.95**	0.04	0.91	0.83	0.07	-0.69	0.83	0.07	-0.73	0.92	0.05	-0.69	0.93	0.05	-0.87	0.90	0.06	0.70
SNR=10(dB)	**0.93**	0.05	0.91	0.92	0.06	-0.78	0.91	0.06	-0.78	0.88	0.07	-0.82	0.90	0.06	-0.82	0.86	0.08	0.78
(SNRs, Noise)	0.86	0.07	0.73	0.86	0.07	-0.68	0.87	0.07	-0.71	0.86	0.07	-0.71	**0.91**	0.06	-0.77	0.86	0.07	0.74
Mean	0.90	0.06	0.84	0.88	0.06	-0.73	0.87	0.06	-0.75	0.88	0.07	-0.76	**0.92**	0.06	-0.83	0.87	0.07	0.75

Table 1. Statistical analysis of top performing metrics for different noise, SNRs and averaged over both SNRs and noise.

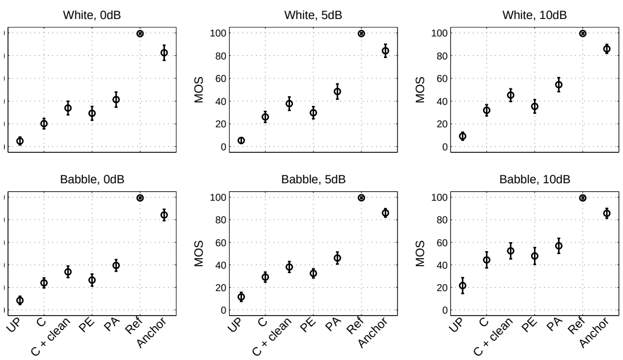

Fig. 4. Mean Opinion Scores (MOS) of the MUSHRA test (top) white (bottom) babble shown for eleven participants.

at SNR $= 0$ (dB) as well as at SNR $= 5$ (dB) with $p = 0.077$.

Another observation is that the overall perceptual quality improvement between the UP and the PE method is more pronounced in white noise than babble. In contrast, the quality improvement achieved by the C method is approximately the same for both noise types. This is due to the fact that blind f_0 estimation [35] works better in white noise, leading to a more accurate estimated phase by the PE method [3, 36].

5.3. Correlation Analysis

For performance evaluation, we used the same approach as presented in [37] where the Pearson's correlation coefficient (ρ), the normalized root-mean-square error (σ) and Kendall's Tau (τ) were used as three figures of merrit. We followed the same procedure in [37] by applying two different logistic functions to account for the nonlinear relationship between the objective and subjective scores.

Figure 5 shows the results separated by the noise types and ranked by the correlation increasing from left to right. The proposed phase-aware metrics performed better than the conventional ones: fwSNR, CEPS, LLR, GSNR and SAR. The PD metric outperforms all the conventional metrics including PESQ in both noise scenarios. The GD metric is the most reliable predictor in white noise. This could be explained due to the robustness of group delay representation against additive white noise as reported in [38].

To further evaluate the reliability of the results, we conducted another correlation analysis at each SNR as well as on the whole data (labeled as (SNRs, Noise) in Table 1). In Table 1 only those metrics with a correlation $\rho > 0.8$ are shown. The metrics which fulfilled this constraint were the proposed phase-aware metrics as well as PESQ and ISa among the conventional metrics. To get the last row (labeled as Mean), we averaged the correlation coefficients. The PD metric showed the highest correlation ($\rho = 0.92$) on average followed by PESQ ($\rho = 0.9$). Furthermore, the PD metric performed most reliable as it showed a more stable ρ across noise types. Following the correlation analysis in [39], the ranking of the proposed PD metric on (SNRs, Noise), as labeled in Table 1, is significant versus global SNR, SSNR, fwSNR, BSS-EVAL, MSE, LLR, IFD and CEPS.

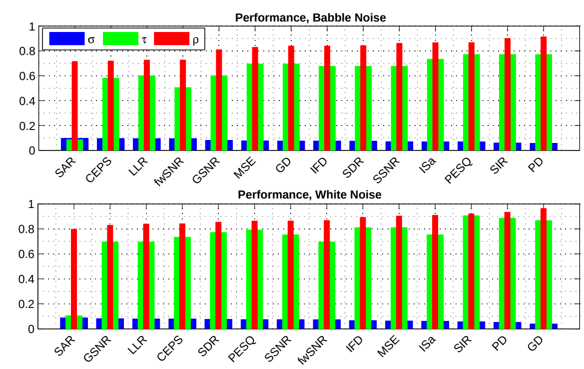

Fig. 5. Performance evaluation of the instrumental metrics categorized to (top) babble (bottom) white noise.

6. CONCLUSION

In this paper, we addressed the following two open questions; which existing instrumental metrics reliably estimate the perceived quality of the enhanced speech when spectral phase is modified and whether a new phase-aware metric would outperform existing ones in terms of predicting subjective results. We quantified the correlation between instrumental metrics and human listening results via conducting statistical analysis. Results showed that both PESQ and the proposed phase deviation metric are reliable estimators of the perceived speech quality when the noisy spectral phase is processed apart from conventional amplitude enhancement schemes.

7. REFERENCES

[1] P. Loizou, *Speech Enhancement: Theory and Practice*, CRC Press, Boca Raton, 2007.

[2] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 229–238, Jan 2008.

[3] J. Kulmer, P. Mowlaee, and M. Watanabe, "A probabilistic approach for phase estimation in single-channel speech enhancement using von mises phase priors," in *IEEE Workshop on Machine Learning for Signal Processing*, Sept. 2014.

[4] C. Chacon and P. Mowlaee, "Least squares phase estimation of mixed signals," *in Proceedings of the 15th International Conference on Spoken Language Processing*, pp. 2705–2709, 2014.

[5] P. Mowlaee, R. Saeidi, and Y. Stylianou, "Interspeech 2014 phase importance in speech processing applications," *in Proceedings of the 15th International Conference on Spoken Language Processing*, pp. 1623–1627, 2014.

[6] P. Mowlaee and R. Saeidi, "Time-frequency constraint for phase estimation in single-channel speech enhancement," *The International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 338–342, 2014.

[7] P. Mowlaee and R. Saeidi, "Iterative closed-loop phase-aware single-channel speech enhancement," *IEEE Signal Processing Letters*, vol. 20, no. 12, pp. 1235–1239, Dec. 2013.

[8] P. Mowlaee and R. Saeidi, "On phase importance in parameter estimation in single-channel speech enhancement," in *ICASSP*, 2013, pp. 7462–7466.

[9] P. Mowlaee, R. Saeidi, and R. Martin, "Phase estimation for signal reconstruction in single-channel speech separation," in *Proc. Interspeech*, 2012, pp. 1–4.

[10] P. Mowlaee and M. Watanabe, "Iterative sinusoidal-based partial phase reconstruction in single-channel source separation," in *Proc. Interspeech*, 2013, pp. 832–836.

[11] T. Gerkmann and M. Krawczyk, "MMSE-optimal spectral amplitude estimation given the STFT-phase," *IEEE Signal Processing Letters*, vol. 20, no. 2, pp. 129 –132, Feb. 2013.

[12] T. Gerkmann, "Bayesian estimation of clean speech spectral coefficients given a priori knowledge of the phase," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4199–4208, Aug 2014.

[13] M. Krawczyk and T. Gerkmann, "STFT phase improvement for single channel speech enhancement," in *International Workshop on Acoustic Signal Enhancement; Proceedings of IWAENC*, 2012, pp. 1–4.

[14] J. R. J. Deller, J. H. L. Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals*, An IEEE Press classic reissue. Wiley, 2000.

[15] J. M. Tribolet, P. Noll, B. McDermott, and R. E. Crochiere, "A study of complexity and quality of speech waveform coders," in *ICASSP.*, Apr 1978, vol. 3, pp. 586–590.

[16] A. Gray and J. Markel, "Distance measures for speech processing," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, no. 5, pp. 380–391, Oct 1976.

[17] N. Kitawaki, H. Nagabuchi, and K. Itoh, "Objective quality evaluation for low-bit-rate speech coding systems," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 2, pp. 242–248, Feb 1988.

[18] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," *Electron. Commun. Japan*, vol. 53, no. 1, pp. 36–43, 1970.

[19] 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 *ICASSP*, 2001, vol. 2, pp. 749–752.

[20] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1462 –1469, Jul 2006.

[21] S. P. Patil and J. N. Gowdy, "Exploiting the baseband phase structure of the voiced speech for speech enhancement," in *ICASSP*, May 2014, pp. 6092–6096.

[22] M. Lagrange and S. Marchand, "Estimating the instantaneous frequency of sinusoidal components using phase-based methods," *J. Audio Eng. Soc*, vol. 55, no. 5, pp. 385–399, 2007.

[23] I. Cohen, "Relaxed statistical model for speech enhancement and a priori SNR estimation," *IEEE Trans. on Speech and Audio Processing*, vol. 13, no. 5, pp. 870–881, 2005.

[24] L. D. Alsteris and K. K. Paliwal, "Short-time phase spectrum in speech processing: A review and some experimental results," *Digital Signal Processing*, vol. 17, no. 3, pp. 578 – 616, 2007.

[25] L. D. Alsteris and K. K. Paliwal, "Further intelligibility results from human listening tests using the short-time phase spectrum," *Speech Communication*, vol. 48, no. 6, pp. 727 – 736, 2006.

[26] M. R. Portnoff, "Short-time fourier analysis of sampled speech," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 3, pp. 364–373, Jun 1981.

[27] P. Maragos, J.F. Kaiser, and T.F. Quatieri, "Energy separation in signal modulations with application to speech analysis," *IEEE Transactions on Signal Processing*, vol. 41, no. 10, pp. 3024–3051, Oct 1993.

[28] P. Vary, "Noise suppression by spectral magnitude estimation mechanism and theoretical limits," *Signal Processing*, vol. 8, no. 4, pp. 387 – 400, 1985.

[29] S. Gustafsson, R. Martin, Peter Jax, and P. Vary, "A psychoacoustic approach to combined acoustic echo cancellation and noise reduction," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 245–256, Jul 2002.

[30] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation theory*, Fundamentals of Statistical Signal Processing. Prentice-Hall PTR, 1993.

[31] M. Cooke, J. Barker, S. Cunningham, and X. Shao, "An audio-visual corpus for speech perception and automatic speech recognition," *Acoustical Society of America Journal*, vol. 120, pp. 2421, 2006.

[32] A. Varga, H. J. M. Steeneken, M. Tomlinson, and D. Jones, "The NOISEX–92 Study on the Effect of Additive Noise on Automatic Speech Recognition," *Technical Report, DRA Speech Research Unit*, 1992.

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

[34] "ITU-R BS.1534-1, Method for the subjective assessment of intermediate quality level of coding systems," 2001.

[35] S. Gonzales and M. Brookes, "PEFAC - a pitch estimation algorithm robust to high levels of noise," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 2, pp. 518–530, Feb 2014.

[36] J. Kulmer and P. Mowlaee, "Phase estimation in single channel speech enhancement using phase decomposition," *IEEE Signal Processing Letters*, vol. 22, no. 5, pp. 598–602, May. 2015.

[37] C. H. Taal, R. C. Hendriks, R. Heusdens, J. Jensen, and U. Kjems, "An evaluation of objective quality measures for speech intelligibility prediction," in *Proc. Interspeech 2009*, 2009, pp. 1947–1950.

[38] P. S. Murthy and B. Yegnanarayana, "Robustness of group-delay-based method for extraction of significant instants of excitation from speech signals," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 6, pp. 609–619, Nov 1999.

[39] J. H Steiger, "Tests for comparing elements of a correlation matrix.," *Psychological Bulletin*, vol. 87, no. 2, pp. 245, 1980.