MULTIZONE NEAR-END SPEECH ENHANCEMENT UNDER OPTIMAL SECOND-ORDER MAGNITUDE DISTORTION

*João B. Crespo**

Delft University of Technology
Signal & Information Processing lab
j.b.crespo@tudelft.nl

Richard C. Hendriks

Delft University of Technology
Signal & Information Processing lab
r.c.hendriks@tudelft.nl

ABSTRACT

In this article, we address near-end speech enhancement for a scenario where there are several playback zones. A signal model is explored, where effects of noise, reverberation and zone crosstalk are taken into account simultaneously. Through the symbolic usage of a general smooth distortion measure, necessary optimality conditions are derived. The conditions are applied to a DFT magnitude-based distortion measure and an algorithm follows, which applies per-zone spectral subtraction followed by channel inversion. Simulations validate the optimality of the algorithm and show a clear benefit in multizone processing, as opposed to the iterated application of a single-zone algorithm.

***Index Terms*—** Near-end speech enhancement, multizone, second-order magnitude distortion, public address system

1. INTRODUCTION

Recently, the field of source-based (near-end) speech enhancement has gained increasing interest in the research community. While traditional speech enhancement systems apply a time-frequency weighting to a *received* noisy speech signal to enhance speech components with respect to the noise [1], *source-based* or *speech reinforcement* systems (*e.g.*, [2, 3, 4]) apply the weighting at a clean speech source in the hope that, when played back in – and corrupted by some acoustic communication channel, degradation is minimized at the listener. Examples of applications which could benefit from source-based speech enhancement range from mobile telephones or hearing aids to conference or public address systems.

Current source-based systems only consider a single playback region where some kind of speech reinforcement is applied. However, many practical scenarios consist of multiple regions or zones, *e.g.*, consider airports, train stations or shoping malls. In such a multizone situation, signals played back in one region can leak ("cross-talk") into another region. As an example of this phenomenon, consider a two-zone system installed on two platforms of a train station (Fig. 1), with some reciprocal cross-talk between zones. Consider also a single-zone signal recovery strategy in each zone (*e.g.*, [2]). In this scenario, each zone working autonomously will (potentially) consider the cross-talk speech coming from the other zone as noise, trying to amplify its own speech to mask the speech from the other zone. Due to the increase in leakage upon an increase in playback level, a competition effect arises between the two zones. This undesirable behaviour motivates the importance of the study of reinforcement taking multiple zones into account.

*This research is partly supported by the Dutch Technology Foundation STW and Bosch Security Systems B.V., The Netherlands.

As far as the authors' knowledge goes, multizone reinforcement as described here has not been previously studied, and constitutes the main novelty of this paper.

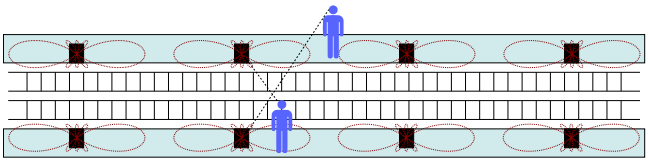

Figure 1: Two-zone speech reinforcement system example (train station). Speech in one platform can leak into the other. Black boxes are loudspeakers, with corresponding directivity diagrams.

Also, many speech reinforcement schemes are designed taking empirical considerations into account without a formal quantitative methodology for pre-processing. For instance, algorithms applying high-pass filtering [5] or some kind of consonant enhancement [6] are based on empirical studies that motivate the relatively higher importance of high frequencies and consonants to intelligibility. Other approaches use heuristic arguments to motivate the construction of the algorithm, *e.g.*, [2], where a signal-to-noise ratio (SNR) recovery method is proposed, which sets the received speech level to a target log distance with respect to the noise.

In contrast to the empirical and somewhat heuristic approaches described above, we take a more formal quantitatively defined approach. We quantify the degradation between source and received speech by a general distortion measure d, which could either model intelligibility or quality degradation using an adequate model. From there on, we formulate a minimization problem on the expected distortion, which we want to solve with respect to the signals played back in each zone.

We split the solution of the minimization problem in two steps. In the first place, we derive general necessary conditions for the optimality of *any* smooth distortion measure. Subsequently, we apply the general conditions to a particular distortion measure, namely, to the second-order magnitude distortion. Thereby, we derive a processing scheme based on a simple frequency-flat transfer of the acoustic channel. The resulting scheme depends on both noise and channel characteristics, applying per-zone spectral subtraction as a first step, followed by channel inversion. The latter is a procedure similarly found and studied in the subject of crosstalk cancellation [7]. Note that speech reinforcement algorithms based on optimization of a merit figure have been studied in the past [8, 3, 4], but these schemes do not use an abstract functional formulation (distortion measure d) and/or do only consider a single-zone scenario.

978-1-4799-0972-8/13/$31.00 ©2013IEEE

2. PRELIMINARIES

As motivated in Sec. 1, we consider a source-based speech enhancement scenario working across multiple zones, say $N \geq 1$ zones. We work with a point-to-point model, defining a zone as a pair of points in space, namely a loudspeaker (source) point and a microphone (receiver) point. Furthermore, we work with speech frames in the discrete Fourier transform (DFT) domain. Denote the f^{th} DFT frequency bin of the source (clean) speech of zone i by $s_i(f) \in \mathbb{C}$, $f \in \{0, 1, \ldots, L-1\}$, $i \in \{1, 2, \ldots, N\}$, where L is the DFT size. We will work with vectors consisting of clean speech stacked up for all zones at each frequency f, $\mathbf{s}(f) \in \mathbb{C}^N$, given by

$$\mathbf{s}(f) = [s_1(f), s_2(f), \ldots, s_N(f)]^{\mathrm{T}}, \tag{1}$$

where $(\cdot)^{\mathrm{T}}$ denotes matrix/vector transposition. Also, the global clean speech vector, $\mathbf{s} \in \mathbb{C}^{LN}$, will be defined by

$$\mathbf{s} = [\mathbf{s}(0)^{\mathrm{T}}, \mathbf{s}(1)^{\mathrm{T}}, \ldots, \mathbf{s}(L-1)^{\mathrm{T}}]^{\mathrm{T}}. \tag{2}$$

The notations defined for the clean speech signal extend to all related signals (*e.g.*, processed speech, noisy speech or noise).

The processed speech, which is played back at the loudspeaker points, is denoted by $\mathbf{s}' \in \mathbb{C}^{LN}$ and is, in general, a function of the clean speech, noise statistics, and acoustic channel parameters. We will model the source and processed speech to be deterministic along our analysis, since source realizations are directly available in speech reinforcement. We further assume a known deterministic acoustic channel transfer, which we describe via the channel matrix

$$\mathbf{H}(f) = \begin{bmatrix} h_{11}(f) & h_{12}(f) & \ldots & h_{1N}(f) \\ h_{21}(f) & h_{22}(f) & \ldots & h_{2N}(f) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N1}(f) & h_{N2}(f) & \ldots & h_{NN}(f) \end{bmatrix}, \tag{3}$$

where $h_{ij}(f)$ is the frequency response from the loudspeaker point of zone j to the microphone point of zone i at frequency bin f. Matrices will be denoted by upper-case bold letters, whereas vectors use lowercase bold. $\mathbf{H}(f)$ can be used to model reverberation in a zone through the diagonal components $h_{ii}(f)$, and cross-talk (leaked signals) from one zone to another through the off-diagonal components $h_{ij}(f), i \neq j$. At the reception (microphone points), there is additive noise inserted, denoted by vector $\mathbf{b}(f)$, which we model to be zero-mean stochastic. The received speech process $\mathbf{x}(f)$ in a multizone scenario is then written down as

$$\mathbf{x}(f) = \mathbf{H}(f)\,\mathbf{s}'(f) + \mathbf{b}(f). \tag{4}$$

We assume further that a globally operating distortion measure $d(\mathbf{s}, \mathbf{x})$ quantifies the dissimilarity between clean and distorted speech realizations, for all frequencies and zones, in the whole system. We assume this measure to be real-valued and continuously differentiable (class C^1) when viewed as a real-argument function in the $2LN$ real variables unraveled by $\mathbf{x}$. The problem we will be concerned with is finding the processed signal vector $\mathbf{s}'$ which minimizes the expected value of the distortion measure, *i.e.*, we want to find the minimizer

$$\bar{\mathbf{s}}' = \underset{\mathbf{s}' \in \mathbb{C}^{LN}}{\arg\min}\ \mathrm{E}[d(\mathbf{s}, \mathbf{g}(\mathbf{s}'))], \tag{5}$$

where $\mathrm{E}[\cdot]$ denotes the expectation operator and $\mathbf{g}(\mathbf{s}')$ represents the function which maps $\mathbf{s}'$ to $\mathbf{x}$ according to the signal model of Eq. (4), for all frequencies f.

3. OPTIMALITY CONDITIONS

In this section, we derive necessary conditions for the processed speech $\mathbf{s}'$ to solve the problem expressed in Eq. (5).

We note that if $d(\mathbf{s}, \mathbf{x})$ is smooth when viewed as a real argument function of the real and imaginary components of $\mathbf{x}$ (class $C^1(\mathbb{R}^{2LN})$), then the expected distortion is smooth as a function of the real and imaginary components of $\mathbf{s}'$, since it is a (continuous) sum of smooth functions in $\mathbf{x}$, composed with a smooth signal model of Eq. (4). Furthermore, it is also known from real-field calculus that a necessary condition for the optimality (minimality or maximality) of $\mathrm{E}[d(\mathbf{s}, \mathbf{g}(\mathbf{s}'))]$ is that its gradients with respect to the real and imaginary components of $\mathbf{s}'$ should vanish. By introducing the complex differential operators [9, Ch. 13, Sec. 2] and noting that the distortion measure $d(\mathbf{s}, \mathbf{x})$ is a real-valued function, we arrive at the equivalent condition

$$\frac{\partial}{\partial \mathbf{s}'^{*}} \mathrm{E}[d(\mathbf{s}, \mathbf{g}(\mathbf{s}'))] = 0, \tag{6}$$

where $\frac{\partial}{\partial \mathbf{s}'^{*}} \equiv \frac{1}{2}\left(\frac{\partial}{\partial \mathbf{s}'_R} - \frac{1}{j}\frac{\partial}{\partial \mathbf{s}'_I}\right)$, $\mathbf{s}'_R$ and $\mathbf{s}'_I$ are the real and imaginary components of $\mathbf{s}'$, respectively, and $\partial/\partial \mathbf{v}$ denotes the transposed Jacobian matrix of a function with respect to argument $\mathbf{v}$.

We now compute the gradient at the left-hand side of Eq. (6) for the sub-variable $\mathbf{s}'(f)$. We thereby obtain

$$\frac{\partial}{\partial \mathbf{s}'(f)^{*}} \mathrm{E}[d(\mathbf{s}, \mathbf{g}(\mathbf{s}'))] = \mathrm{E}\left[\frac{\partial}{\partial \mathbf{s}'(f)^{*}} d(\mathbf{s}, \mathbf{g}(\mathbf{s}'))\right] \tag{7}$$

$$= \mathrm{E}\left[\sum_{\nu=0}^{L-1} \left(\frac{\partial \mathbf{g}(\nu)}{\partial \mathbf{s}'(f)}\right)^{*} \frac{\partial d(\mathbf{s}, \mathbf{x})}{\partial \mathbf{x}(\nu)^{*}}\right] \tag{8}$$

$$= \mathbf{H}(f)^{\mathrm{H}}\, \mathrm{E}\left[\frac{\partial d(\mathbf{s}, \mathbf{x})}{\partial \mathbf{x}(f)^{*}}\right], \tag{9}$$

where $(\cdot)^{*}$ and $(\cdot)^{\mathrm{H}}$ denote conjugation and conjugate transposition, respectively. For all that follows, we denote the probability density function (PDF) of a random variable z at value ζ by $\mathbb{P}_z(\zeta)$, and the ℓ_2 norm by $\|\cdot\|$. In the derivation of Eqs. (7)–(9), we assume that the tail of $\mathbb{P}_{\mathbf{b}}(\boldsymbol{\beta})$ converges to zero fast enough when $\|\boldsymbol{\beta}\| \to \infty$, so that we can exchange differentiation and integration (expectation) order. We also use properties of the complex differential operators in [9, Ch. 13, Sec. 2], the analyticity of the signal model in Eq. (4), computing the corresponding (transposed) Jacobian matrix, and the linearity of the expectation operator combined with a deterministic assumption for $\mathbf{H}(f)$.

If we compare Eq. (9) with Eq. (6), we find out that the necessary conditions for a processed speech vector $\mathbf{s}'$ to optimize the expected value of the distortion measure $d(\mathbf{s}, \mathbf{x})$ are

$$\mathbf{H}(f)^{\mathrm{H}}\, \mathrm{E}\left[\frac{\partial d(\mathbf{s}, \mathbf{x})}{\partial \mathbf{x}(f)^{*}}\right] = 0, \quad \forall f \in \{0, 1, \ldots, L-1\}. \tag{10}$$

Furthermore, if we assume the distortion measure to be additive in frequency, *i.e.*, if we take distortion measures of the form $d(\mathbf{s}, \mathbf{x}) = \sum_{\nu=0}^{L-1} d'(\mathbf{s}(\nu), \mathbf{x}(\nu))$ for some intermediate distortion measure d' operating independently for each frequency, Eq. (10) simplifies to

$$\mathbf{H}(f)^{\mathrm{H}}\, \mathrm{E}\left[\frac{\partial d'(\mathbf{s}(f), \mathbf{x}(f))}{\partial \mathbf{x}(f)^{*}}\right] = 0, \quad \forall f. \tag{11}$$

4. APPLICATION EXAMPLE

We now apply the optimality condition derived in Section 3 to the second-order magnitude distortion measure defined by

$$d(\mathbf{s},\mathbf{x}) = \| |\mathbf{x}|^2 - |\mathbf{s}|^2 \|^2 \,, \tag{12}$$

where the magnitude $|\cdot|$ is taken component-wise. This kind of distortion measures is common in speech enhancement (*e.g.*, [10]), as they measure the distance between magnitude spectra. For simplicity, we restrict ourselves to a non-singular channel matrix $\mathbf{H}(f)$.

It is easy to see that the distortion measure is additive in frequency, so that we can take the sub-distortion measure

$$d'(\mathbf{s}(f),\mathbf{x}(f)) = \| |\mathbf{x}(f)|^2 - |\mathbf{s}(f)|^2 \|^2 \tag{13}$$

for each frequency f. By application of the differential operator [9, Ch. 13, Sec. 2] and its properties, and substituting the resulting derivative in Eq. (11), we arrive at

$$2\mathbf{H}(f)^{\mathrm{H}} \,\mathrm{E}\left[(|\mathbf{x}(f)|^2 - |\mathbf{s}(f)|^2) \otimes \mathbf{x}(f) \right] = 0 \,, \tag{14}$$

where $\otimes$ denotes component-wise multiplication. Under the non-sigularity assumption of $\mathbf{H}(f)$, the unique solution of Eq. (14) corresponds to making the expected value vanish. For working out the expected value, we use the signal model of Eq. (4), properties of the expectation operator, the assumption of zero-mean noise DFT coefficients and an additional assumption that the noise DFT coefficients are circular symmetric, *i.e.*, that their PDF, when expressed in polar coordinates, is of the form $\mathbb{P}_{b_i(f)}(re^{j\theta}) = \frac{1}{2\pi}\mathbb{P}_{|b_i(f)|}(r)$, for some marginal distribution $\mathbb{P}_{|b_i(f)|}(r)$ of the noise DFT magnitudes, for all zones i and frequencies f. This last assumption is needed to make cross terms vanish and is ubiquitously satisfied. We arrive at the result

$$\left(|\mathbf{p}(f)|^2 + 2\,\mathrm{E}[|\mathbf{b}(f)|^2] - |\mathbf{s}(f)|^2 \right) \otimes \mathbf{p}(f) = 0 \,, \tag{15}$$

where $\mathbf{p}(f) = \mathbf{H}(f)\,\mathbf{s}'(f)$ is the processed received speech (without noise). Note that, due to the component-wise nature of the expectation argument in Eq. (14), the noise coefficients do not need to be uncorrelated.

Discarding the uninteresting case where the processed signal suppresses the input speech, the optimal processed received speech should satisfy

$$|\bar{\mathbf{p}}(f)|^2 = |\mathbf{s}(f)|^2 - 2\,\mathrm{E}[|\mathbf{b}(f)|^2] \,, \tag{16}$$

i.e., a per-zone spectral subtraction scheme with over-subtraction factor equal to 2 should be applied. Besides spectral subtraction, channel inversion is also applied; indeed, by definition of $\mathbf{p}$, the optimally processed signal at the source $\mathbf{s}'$ is given by

$$\bar{\mathbf{s}}'(f) = \mathbf{H}(f)^{-1}\bar{\mathbf{p}}(f) \,. \tag{17}$$

Eqs. (16), (17) introduce a speech reinforcement scheme which is derived from an expected distortion optimization procedure on the distortion measure (12). The scheme exploits both the knowledge about noise statistics and acoustic channel information. Note also the relation of Eq. (17) with crosstalk cancellation approaches [7]. There, the challenge is to design a pre-mixing linear matrix filter, so that crosstalk vanishes. In contrast, in our framework, we concentrate on optimizing a distortion measure operating on source and received signals, potentially resulting in a non-linear pre-processing scheme. Due to the simplicity of the distortion measure eventually chosen and the deterministic channel assumption, the concrete approach in Eq. (17) boils down to the ideal linear operation desired in crosstalk cancellation.

5. SIMULATIONS

To assess the performance of the proposed second-order magnitude distortion optimal algorithm of Sec. 4, we apply it in a multizone speech reinforcement context, with $N = 4$ zones. The zones are laid out in a square constellation, with square side $d = 20$ m. For simplicity, the channel matrix $\mathbf{H}(f)$ of Eq. (17) is considered to be independent of frequency. This corresponds to including only attenuations (gains) in the channel model, discarding reverberation, delay and other channel response related effects. The gain transfer from zone j to zone i is modeled using a standard free-field law, $h_{ij} = 1/d_{ij}$, where d_{ij} is the distance between loudspeaker in zone j and microphone in zone i. For different zones $i \neq j$, this is approximated by the distance between zones, according to the constellation described above, and within a zone, the distance is chosen to be 10% of the square side length d.

Per zone, one hundred sentences were used (different for each zone), randomly chosen out of the TIMIT database, each sentence having a duration of at least two seconds. The sentences were silence-trimmed at the extremities, processed and passed through the signal model according to Eq. (4). In total, 19.1 minutes of speech were used. Speech-shaped noise DFT coefficients were generated for the noise term $\mathbf{b}(f)$, modelling thereby a diffuse noise field in the environment. We varied the signal-to-noise ratio (SNR) with respect to the source signal from -20 to 60 dB in steps of 5 dB. We evaluate speech processed using the proposed algorithm, using a single-zone variant where the proposed algorithm for $N = 1$ is applied independently for each zone (a 1-by-1 channel $\mathbf{H}(f) = h_{ii}(f)$ is used in each zone i) and, for reference, we include unprocessed speech as well. This results in a total number of $17 \times 3 = 51$ conditions. The signal-to-distortion ratio (SDR)

$$\mathrm{SDR} = 10\log_{10}\left(\frac{\mathrm{E}[d(\mathbf{s},\mathbf{0})]}{\mathrm{E}[d(\mathbf{s},\mathbf{g}(\mathbf{s}'))]} \right) \,, \tag{18}$$

being $d(\cdot,\cdot)$ the distortion measure of Eq. (12), and the Short-Time Objective Intelligibility measure (STOI) [11] are computed for each condition. The SDR essentially reflects the proposed distortion measure in a merit figure, while STOI is an objective intelligibility predictor designed for time-frequency weighted speech. Note that although STOI was not designed for our framework, but for traditional (far-end) noise suppression instead, we still aim to gain some (limited) insight on speech intelligibility.

All signals are sampled at a sample frequency $f_s = 16$ kHz. For implementing the processing function $\mathbf{s} \mapsto \mathbf{s}'$, we use a weighted overlap-add (WOLA) mechanism with 32 ms frame size and 50% overlap square-root Hann windows. The spectral subtraction scheme of Eq. (16) is implemented trimming negative values to zero and re-using the phase of the source speech [1]. Also, an ideal noise tracker is used which has access to the noise realizations and smoothens periodograms along time. Finally, expected distortions in Eq. (18) are estimated by frame-wise distortion computation and averaging along time. A Hann analysis window is used for this, with the remaining short-time DFT parameters as above.

The results are shown in Fig. 2, which plots the SDR and STOI for the three conditions as a function of the SNR. We observe that for both merit figures, multizone processed speech outperforms the other scenarios, followed by single-zone processing. Furthermore, processed speech achieves higher performance gains for higher SNR values. As far as the SDR is concerned, we observe that, in contrast to the multizone case, in non-multizone conditions there is a plateau in performance gain with increasing SNR. This

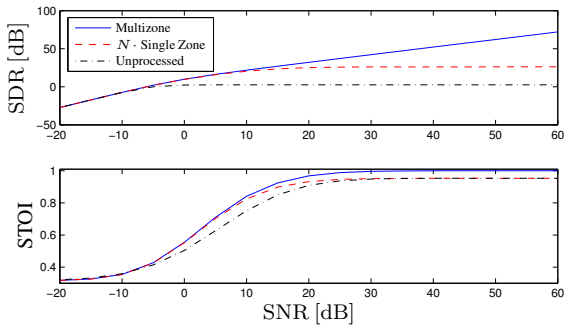

Figure 2: Simulated SDR and STOI for the second-order magnitude distortion as a function of SNR.

can be motivated by the fact that multizone processing performs perfect cross-talk cancellation and that under high SNR conditions, the additive noise term vanishes, yielding thereby zero distortion. In the single zone case, this is not possible since each zone only takes its own speech into account and only compensates for its own channel gain, leaving a residual signal over from the other zones, even for high SNR. The STOI simulation confirms this reasoning, as perfect intelligibility is obtained for the multizone case, in contrast to the other scenarios.

Finally, we would like to note that, as described earlier, negative power spectrum values were trimmed to zero while performing spectral subtraction in Eq. (16). Due to this post-processing step, the algorithm which is obtained in practice is sub-optimal. To assess how far we are from the optimal case, we evaluate the ratio of trimmed DFT bins with respect to the total amount of DFT bins. This ratio is shown in Fig. 3. We observe that from about 35 dB

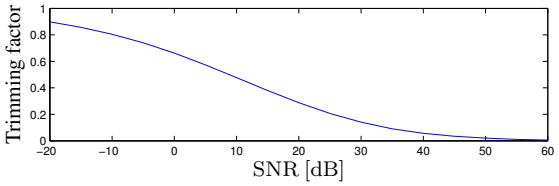

Figure 3: Trimming factor for the spectral subtraction step, as a function of SNR.

SNR on, the trimming factor reaches a negligible amount smaller than 10%, which corresponds to the performance region in Fig. 2 where the algorithm performs best. We thus conclude that for high SNR ratios, the practical implementation performs close to optimal.

6. CONCLUSIONS

In Sec. 2, we introduced an optimization framework for multizone speech reinforcement. This framework was used in Sec. 3 to derive necessary optimality conditions in the optimization of the expected value of *any* given smooth distortion measure. The conditions were applied in Sec. 4 for the second-order magnitude distortion measure. From this measure, an algorithm was derived which enhances speech by means of spectral subtraction and channel inversion. To the best of the authors' knowledge, this is the first approach considering the existence of multiple zones in speech reinforcement, and using a framework including noise, reverberation, *and* interzone crosstalk simultaneously. Also, a formal optimization approach was taken with the derivation of abstract (re-usable) optimality conditions.

The main limitation of the proposed algorithm is that it relies on a limited model of channel transfer. For example, no effects of time delay between zones can be taken into account, neither any effects related to reverberation. Note, nevertheless, that the abstract framework itself *does* include support for reverberation and convolutive effects. Furthermore, the deterministic model assumed for the channel limits its applicability due to the large errors induced on the channel estimate upon source-receiver displacement. Future work should thus concentrate on increasing robustness by including stochastic channel descriptions in the developed framework, as is done in crosstalk cancellation [7]. Also, given the re-usability of the optimality conditions, one could as well think of applying this work to other analytically defined measures.

7. REFERENCES

[1] R. C. Hendriks, T. Gerkmann, and J. Jensen, *DFT-Domain Based Single-Microphone Noise Reduction for Speech Enhancement – A Survey of the State-of-the-Art*. Morgan & Claypool Publishers, 2013.

[2] B. Sauert and P. Vary, "Near end listening enhancement: Speech intelligibility improvement in noisy environments," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. I, May 2006, pp. 493–496.

[3] ——, "Recursive closed-form optimization of spectral audio power allocation for near end listening enhancement," in *ITG-Fachtagung Sprachkommunikation*, vol. Paper 8, Oct. 2010.

[4] C. H. Taal, R. C. Hendriks, and R. Heusdens, "A speech preprocessing strategy for intelligibility improvement in noise based on a perceptual distortion measure," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, March 2012, pp. 4061–4064.

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

[6] M. D. Skowronski and J. G. Harris, "Applied principles of clear and Lombard speech for automated intelligibility enhancement in noisy environments," *Speech Communication*, vol. 48, pp. 549–558, 2006.

[7] D. B. Ward and G. W. Elko, "Virtual sound using loudspeakers: robust acosutic crosstalk cancellation," in *Acoust. Signal Proc. for Telecom*, S. L. Gay and J. Benesty, Eds. Boston, MA: Kluwer Academic, 2000, ch. 14.

[8] J. D. Griffiths, "Optimum linear filter for speech transmission," *J. Acoust. Soc. Am.*, vol. 43, no. 1, pp. 81–86, 1968.

[9] J. B. Conway, *Functions of one complex variable II*, 1st ed. Springer-Verlag, May 1995.

[10] C. You, S. Koh, and S. Rahardja, "Adaptive beta-order mmse estimation for speech enhancement," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 1, april 2003, pp. 900–903.

[11] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, Sept. 2011.