Non-Intrusive Intelligibility Prediction Using a Codebook-Based Approach

Charlotte Sørensen*†, Mathew S. Kavalekalam*, Angeliki Xenaki†, Jesper B. Boldt† and Mads G. Christensen*
*Audio Analysis Lab, AD:MT, Aalborg University, Denmark
{cs,msk,mgc}@create.aau.dk
†GN Hearing A/S, Lautrupbjerg 7, DK-2750, Ballerup, Denmark
{csoerensen,axenaki,jboldt}@gnresound.com

***Abstract*—It could be beneficial for users of hearing aids if these were able to automatically adjust the processing according to the speech intelligibility in the specific acoustic environment. Most speech intelligibility metrics are intrusive, i.e., they require a clean reference signal, which is rarely available in real-life applications. This paper proposes a method, which allows using an intrusive short-time objective intelligibility (STOI) metric without requiring access to a clean signal. The clean speech reference signal is replaced by the clean speech envelope spectrum estimated from the noisy signal. The spectral envelope has been shown to be an important cue for speech intelligibility and is used as the reference signal inside STOI. The spectral envelopes are estimated as a combination of predefined dictionaries, i.e., codebooks, that best fits the noisy speech signal. The simulations show a high correlation between the proposed non-intrusive codebook-based STOI (NIC-STOI) and the intrusive STOI indicating that NIC-STOI is a suitable metric for automatic classification of speech signals.**

I. Introduction

Speech is a fundamental tool for human communication. Understanding speech becomes a challenging task in adverse listening conditions such as ”the cocktail party scenario” especially for hearing impaired individuals [1], [2]. Speech enhancement algorithms aim to improve speech intelligibility for hearing aid users [3], [4], [5]. However, speech enhancement algorithms may be beneficial in some acoustic scenarios whereas the same algorithms can have a negative impact on quality and intelligibility in other conditions [5], [6]. Thus, it would be beneficial for HA users if speech enhancement algorithms are automatically limited to scenarios in which they provide an improvement in speech intelligibility [3], [4]. This could be facilitated by an objective speech intelligibility metric processed online in the HA.

Several methods can with an acceptable accuracy predict the speech intelligibility intrusively, i.e., they require access to a clean speech reference [4]. Some of the earliest intrusive metrics that predict the intelligibility well for a limited type of degradations, like linear filtering and additive noise, include the articulation index (AI) [7] and the speech transmission index (STI) [8]. Later, the short-time objective (STOI) metric [9] and the speech-based envelope power spectrum model (sEPSM) [10] were introduced for more complex distortion types and are reported to have an useful reliability [4]. However, the need for a clean speech signal would be a limitation for real-time prediction of speech intelligibility, since this is rarely available. More recently, a number of non-intrusive metrics not requiring access to a clean speech reference signal have been introduced, e.g., the speech-to-reverberation modulation energy ratio (SRMR) [11], the modulation spectrum area (ModA) [12]. These methods are, however, either limited to assessment of reverberated speech or still inferior to the intrusive metrics [4].

This paper proposes a non-intrusive intelligibility prediction method referred to as the non-intrusive codebook-based STOI (NIC-STOI). The method estimates the intelligibility of noisy speech non-intrusively by comparing relevant features of the clean speech with the features of the noisy speech inside a well-established intrusive intelligibility framework, STOI, similar to [13], [14]. The relevant features of the clean speech are based on the spectral envelope of the speech, which has been shown to be an important cue for speech intelligibility [15]. The spectral envelopes of the clean speech and the noise signal are estimated as the most suitable combination from a predefined speech and noise spectra dictionary, a codebook, which best fits the noisy speech signal using a codebook-based approach [16], [17]. These codebooks consist of filter coefficients that capture the overall structure of the spectral envelope.

II. The NIC-STOI measure

NIC-STOI allows predicting the intelligibility from the noisy signal only using an intrusive metric (STOI) without requiring access to the clean speech signal. The approach behind the method is to replace the clean reference signal with an estimate of the clean speech features obtained from the noisy signal. An estimate of the clean speech spectral envelope is used as the relevant features of speech intelligibility in the method. Then, NIC-STOI gives a non-intrusive intelligibility prediction by comparing the correlation of the estimated clean speech spectrum with the noisy spectrum with the intrusive STOI measure. The framework of the measure is illustrated by a block diagram in Fig. 1. The framework can be divided into three main steps: (1) The parameters needed to obtain the clean speech reference are estimated, (2) time-frequency-spectra of the clean and noisy speech signals are composed from the estimated parameters, and, (3) an intelligibility score is predicted with the intrusive STOI framework.

ISBN 978-0-9928626-7-1 © EURASIP 2017

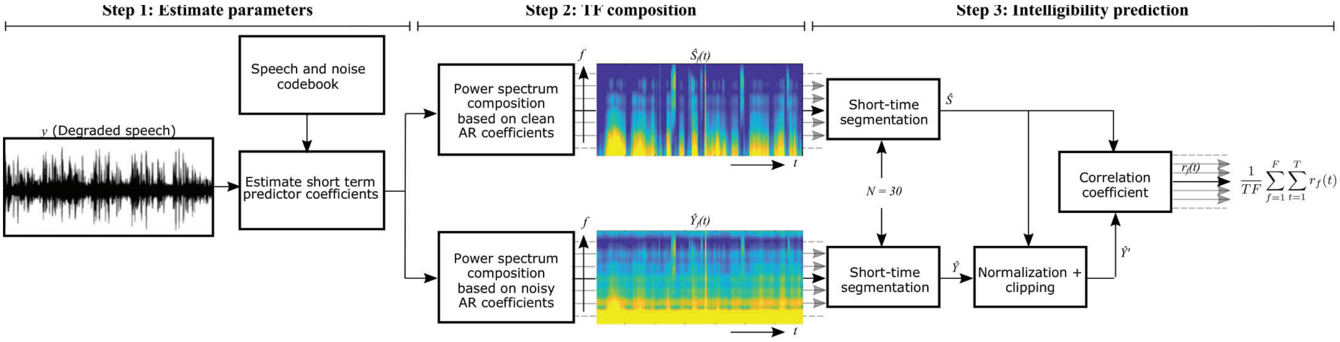
Step 1: Estimate parameters
Step 2: TF composition
Step 3: Intelligibility prediction
Speech and noise codebook
y (Degraded speech)
Estimate short term predictor coefficients
Power spectrum composition based on clean AR coefficients
Power spectrum composition based on noisy AR coefficients
f
t
$\hat{S}_f(t)$
$\hat{Y}_f(t)$
Short-time segmentation
Short-time segmentation
$N = 30$
$\hat{S}$
$\hat{Y}$
Normalization + clipping
$\hat{Y}'$
Correlation coefficient
$r_f(t)$
$\frac{1}{TF}\sum_{f=1}^{F}\sum_{t=1}^{T} r_f(t)$

Fig. 1. Block diagram illustrating the proposed non-intrusive codebook-based STOI metric in which the relevant features of the clean and noisy speech signals are composed as time-frequency power spectra using a codebook-based approach and utilized within the intrusive framework, STOI.

A. Signal model

The proposed method is based on an additive noise model assuming the speech and noise are statistically uncorrelated from [16], [17], i.e.,

$$y(n) = s(n) + w(n), \tag{1}$$

where $y(n)$, $s(n)$ and $w(n)$ represent the sampled noisy speech, clean speech and noise, respectively. The clean speech signal can be modeled as a stochastic autoregressive (AR) process

$$s(n) = \sum_{i=1}^{P} a_{s_i}(n)s(n-i) + u(n) = \mathbf{a}_s(n)^T \mathbf{s}(n-1) + u(n), \tag{2}$$

where $\mathbf{s}(n-1) = [s(n-1), \dots, s(n-P)]^T$ with the P past speech samples, $\mathbf{a}_s(n) = [a_{s_1}(n), a_{s_2}(n), \dots, a_{s_P}(n)]^T$ is a vector containing the speech linear prediction coefficients (LPC), and $u(n)$ is zero mean white Gaussian noise with excitation variance $\sigma_u^2(n)$. Similarly, the noise signal can be modeled as

$$w(n) = \sum_{i=1}^{Q} a_{w_i}(n)w(n-i) + v(n) = \mathbf{a}_w(n)^T \mathbf{w}(n-1) + v(n), \tag{3}$$

where $\mathbf{w}(n-1) = [w(n-1), \dots, w(n-Q)]^T$ with the Q past noise samples, $\mathbf{a}_w(n) = [a_{w_1}(n), a_{w_2}(n) \dots, a_{w_Q}(n)]^T$, and $v(n)$ is zero mean white Gaussian noise with excitation variance $\sigma_v^2(n)$.

The AR model is used to model the speech and noise signals as well as training the codebook dictionaries.

B. Step 1: Estimate parameters

The spectra of the clean and noisy speech signals are estimated from the LPC and the excitation variances concatenated in the vector $\theta = [\mathbf{a}_s \; \mathbf{a}_w \; \sigma_u^2(n) \; \sigma_v^2(n)]$. These parameters are estimated using a priori information from a trained codebook about the speech and noise spectral shapes in the form of LPC based on the approach in [16], [18], [17], where more details on the derivation of this method can be found. Given the observed vector of noisy samples $\mathbf{y} = [\; y(0) \; y(1) \; \dots \; y(N-1) \;]$ for the current frame of length N, the MMSE (minimum mean square error) estimate of θ can be given as $\hat{\theta} = \mathrm{E}(\theta|\mathbf{y})$ for the support space of the parameters to be estimated, Θ, and using Bayes' theorem can be reformulated as

$$\hat{\theta} = \int_{\Theta} \theta p(\theta|\mathbf{y}) d\theta = \int_{\Theta} \theta \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})} d\theta. \tag{4}$$

The vector, $\theta_{ij} = [\mathbf{a}_{s_i} \; \mathbf{a}_{w_j} \; \sigma_{u,ij}^{2,\mathrm{ML}}(n) \; \sigma_{v,ij}^{2,\mathrm{ML}}(n)]$, is then defined for each i^{th} entry of the speech codebook and j^{th} entry of the noise codebook, respectively. The maximum likelihood (ML) estimates of the speech and noise excitation variances, $\sigma_{u,ij}^{2,\mathrm{ML}}$ and $\sigma_{v,ij}^{2,\mathrm{ML}}$, respectively, are then given by [18], [16]

$$\mathbf{C} \begin{bmatrix} \sigma_{u,ij}^{2,\mathrm{ML}} \\ \sigma_{v,ij}^{2,\mathrm{ML}} \end{bmatrix} = \mathbf{D}, \tag{5}$$

where

$$\mathbf{C} = \begin{bmatrix} \|\frac{1}{P_y^2(\omega)|A_s^i(\omega)|^4}\| & \|\frac{1}{P_y^2(\omega)|A_s^i(\omega)|^2|A_w^j(\omega)|^2}\| \\ \|\frac{1}{P_y^2(\omega)|A_s^i(\omega)|^2|A_w^j(\omega)|^2}\| & \|\frac{1}{P_y^2(\omega)|A_w^j(\omega)|^4}\| \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} \|\frac{1}{P_y^2(\omega)|A_s^i(\omega)|^2}\| \\ \|\frac{1}{P_y^2(\omega)|A_w^j(\omega)|^2}\| \end{bmatrix} \tag{6}$$

where A_s^i and A_w^j are the spectra of the i^{th} and j^{th} vector from the speech codebook and noise codebook, respectively, and with $\|f(\omega)\| = \int |f(\omega)| d\omega$. The spectral envelope of the speech codebook, the noise codebook and the noisy signal are given by $\frac{1}{|A_s^i(\omega)|^2}$, $\frac{1}{|A_w^j(\omega)|^2}$ and $P_y(\omega)$, respectively. In practice, the MMSE estimate of θ in Eq. 4 is evaluated as a weighted linear combination of θ_{ij} by

$$\hat{\theta} = \frac{1}{N_s N_w} \sum_{i=1}^{N_s} \sum_{j=1}^{N_w} \theta_{ij} \frac{p(\mathbf{y}|\theta_{ij}) p(\sigma_{u,ij}^{2,\mathrm{ML}}) p(\sigma_{v,ij}^{2,\mathrm{ML}})}{p(\mathbf{y})}, \tag{7}$$

where N_s and N_w are the the number of entries in the speech and noise codebooks, respectively. The weight of the MMSE estimate, $p(\mathbf{y}|\theta_{ij})$, can be computed as

$$p(\mathbf{y}|\theta_{ij}) = e^{-d_{\mathrm{IS}}(P_y(\omega), \hat{P}_y^{ij}(\omega))} \tag{8}$$

$$\hat{P}_y^{ij}(\omega) = \frac{\sigma_{u,ij}^{2,\mathrm{ML}}}{|A_s^i(\omega)|^2} + \frac{\sigma_{v,ij}^{2,\mathrm{ML}}}{|A_w^j(\omega)|^2} \tag{9}$$

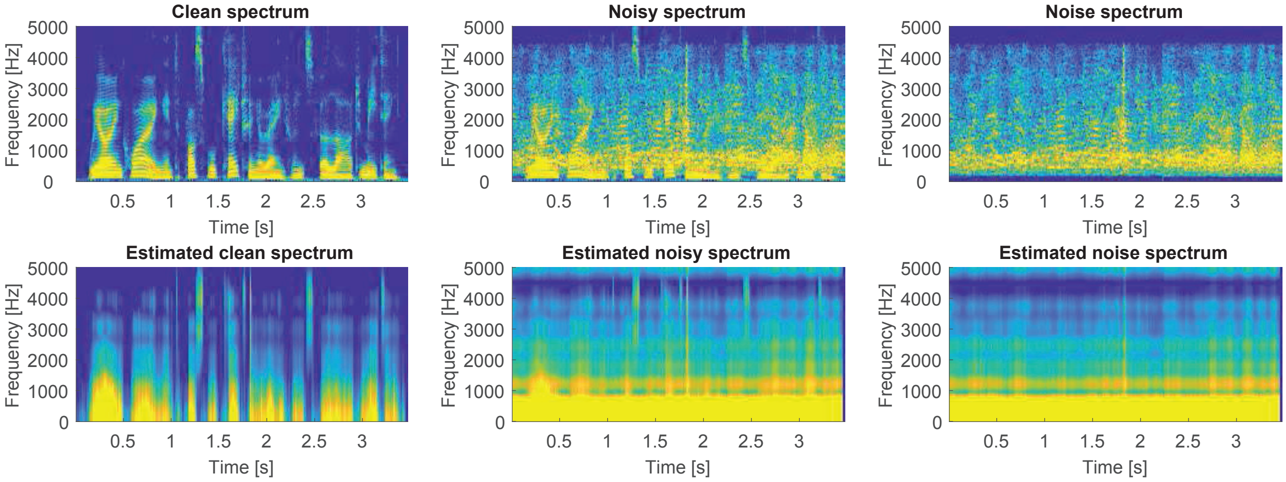

Fig. 2. Spectrograms of the original clean speech signal, noisy speech signal at 0 dB SNR and noise signal are depicted in the top panel from left to right, respectively, as well as their corresponding estimated power spectra from the codebook-based approach in the bottom panel.

$$p(\mathbf{y}) = \frac{1}{N_s N_w} \sum_{i=1}^{N_s} \sum_{j=1}^{N_w} p(\mathbf{y}|\theta_{ij}) p(\sigma^2_{u,ij}) p(\sigma^2_{v,ij}), \tag{10}$$

where the Itakura-Saito distortion between the noisy spectrum and the modeled noisy spectrum is given by $d_{\text{IS}}(P_y(\omega), \hat{P}_y^{ij}(\omega))$ [19], [17]. The weighted summation of the LPC should be performed in the line spectral frequency domain in order to insure stable inverse filters [16], [17].

C. Step 2: TF composition

Time-frequency (TF) power spectrum of the estimated reference signal, $\hat{S}$, are composed from the estimated AR filter coefficients of the clean speech signal $\hat{\mathbf{a}}_s$ for each time frame:

$$\hat{S}(\omega) = \frac{\hat{\sigma}^2_u}{|\hat{A}_s(\omega)|^2}, \tag{11}$$

where $\hat{A}_s(\omega) = \sum_{k=0}^{P} \hat{a}_{s_k} e^{-j\omega k}$. In the same manner, the estimated noise AR filter coefficients, $\hat{\mathbf{a}}_w$, are used to compose a TF spectrum of the noise:

$$\hat{W}(\omega) = \frac{\hat{\sigma}^2_v}{|\hat{A}_w(\omega)|^2}, \tag{12}$$

where $\hat{A}_w(\omega) = \sum_{k=0}^{Q} \hat{a}_{w_k} e^{-j\omega k}$. The LPC, i.e. $\hat{\mathbf{a}}_s$ and $\hat{\mathbf{a}}_w$, determine the shape of the envelope of the corresponding signals $\hat{S}(\omega)$ and $\hat{W}(\omega)$, respectively. The excitation variances, $\hat{\sigma}_u$ and $\hat{\sigma}_v$, determine the overall signal magnitude. Finally, the noisy spectrum is composed as the combined sum of the clean and the noise power spectra:

$$\hat{Y}(\omega) = \hat{S}(\omega) + \hat{W}(\omega). \tag{13}$$

These time-frequency spectra replace the discrete Fourier transform of the clean reference signal and the noisy signal in the original STOI measure [9].

D. Step 3: Intelligibility Prediction

In the final step, the intelligibility prediction is carried out in exactly the same manner as for the STOI measure [9]. The power spectra of the noisy speech, $\hat{Y}$, are further clipped by a normalisation procedure expressed in Eq. 14 in order to de-emphasize the impact of region in which noise dominates the spectrum:

$$\hat{Y}' = \max(\min(\lambda \cdot \hat{Y}, (1 + 10^{-\beta/20}) \cdot \hat{S}), (1 - 10^{-\beta/20}) \cdot \hat{S}), \tag{14}$$

where $\hat{S}$ is the power spectrum of the estimated reference signal, $\lambda = \sqrt{\sum \hat{S}^2 / \sum \hat{Y}^2}$ is a scale factor for normalizing the noisy TF bins and $\beta = -15$ dB is the lower signal-to-distortion ratio. Given the local correlation coefficient, $r_f(t)$, between $\hat{Y}$ and $\hat{S}$ at frequency f and time t, the NIC-STOI prediction is given by averaging across all bands and frames:

$$\text{NIC-STOI} = \frac{1}{TF} \sum_{f=1}^{F} \sum_{t=1}^{T} r_f(t). \tag{15}$$

III. Simulation Methodology

The proposed metric NIC-STOI is evaluated on speech samples from of 5 male and 5 female speakers from the EUROM_1 database of the English sentence corpus [20]. The interfering additive noise signal is simulated in the range of -30 to 30 dB SNR as multi-talker babble from the NOIZEUS database [6]. The LPC and variances of both the speech and noise signal are estimated from 25.6 ms frames with sampling frequency 10 kHz. The speech and, thus, the STP parameters are assumed to be stationary over these very short frames. The AR model order P and Q of both the speech and noise, respectively, is set to 14 according to literature [16], [18], [17]. The speech codebook is generated on a training sample of 15 minutes of speech from multiple speakers in the EUROM_1 database in order to assure a generic speech model using the

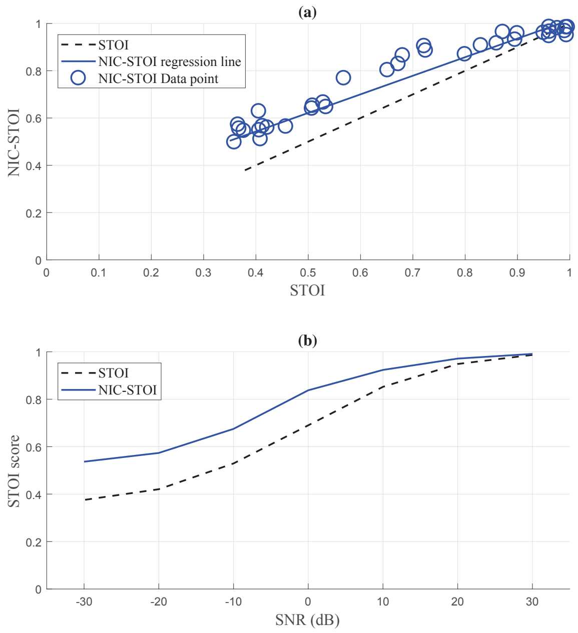

Fig. 3. (a) Scatter plot of the non-intrusive codebook-based STOI (NIC-STOI) metric versus the intrusive STOI metric and (b) STOI and NIC-STOI as a function of SNR.

generalized Lloyd algorithm (GLA) [16], [21]. The speech codebook training sample does not include speech samples from the speakers used in the test set. The noise codebook is trained on 2 minutes of babble talk. The sizes of the speech and noise codebooks are $N_s = 64$ and $N_w = 8$, respectively. The performance of the metric is evaluated using three performance criteria common for assessment of objective intelligibility metrics [4], [14]; Pearson's correlation (ρ) which characterizes the linear relationship, Spearman's rank (ρ_{spear}) and Kendall's tau (τ) which both quantify the ranking capability.

IV. Results and Discussion

The spectra of an example speech signal in the test set is shown in Fig. 2 for the original clean speech signal, the noisy speech signal at 0 dB SNR and the noise signal in the top panel from left to right, respectively. In the bottom panel the corresponding estimated power spectra of relevant signal features are composed using trained codebooks of speech and noise spectral shapes parametrized as LPC to model the a priori information in a Bayesian MMSE scheme.

It can be observed that the method only captures the overall envelope structure and not the fine structure of speech, since it is based on an AR model [19], [17]. Only modeling the overall envelope structure is assumed to be sufficient for depicting the essential features of clean speech, since the envelope structure has long been identified as an important cue for speech intelligibility used within other intrusive intelligibility prediction frameworks, i.e., STI and EPSM [15], [8], [10]. This viewpoint can also be supported by extensive vocoder simulations, where it has been shown that envelope cues from only four spectral bands are sufficient to yield a high intelligibility of speech perception in quiet [15]. As such, it seems to be a reasonable assumption that only depicting the overall envelope structure can be a good predictor for speech intelligibility.

TABLE I
Performance of the proposed metric in terms of Pearson's correlation (ρ), the Spearman rank (ρ_{SPEAR}) and Kendall's tau (τ) between NIC-STOI and STOI as well as the linear regression line.

Metric	ρ	ρ_{spear}	τ	Regression line
NIC-STOI	0.972	0.961	0.8521	$0.730 \cdot \text{STOI} + 0.285$

The performance of the NIC-STOI metric is evaluated in relation to the corresponding original STOI scores. In Fig. 3a there is a clear monotonic correspondence between the NIC-STOI score (blue solid line) and the intrusive STOI measure (black dashed line), such that a higher NIC-STOI score also corresponds to a higher STOI score. Furthermore, a strong linear trend can be observed between the NIC-STOI and STOI measures. This observation is also supported by the performance criteria given in Table I, where Pearson's correlation and the Spearman Rank is close to one implying a high correlation. This indicates that the proposed non-intrusive version of STOI can offer a comparable performance to the original intrusive STOI. In Fig. 3b the STOI measure (black dashed line) and the NIC-STOI measure (blue solid line) are depicted as function of SNR. There is a clear monotonic correspondence between NIC-STOI and STOI, such that a higher STOI measure results in a higher NIC-STOI score. Furthermore, the NIC-STOI scores also increase with increasing SNRs. The offset between the two graphs can be accounted for by the linear trend described in Table I, which gives the translation between NIC-STOI and STOI scores.

In future work, it would be interesting to investigate how the method performs with different noise types and environments as well as unseen noise conditions. Additionally, the objective results could be tested against subjective listening experiments for further validation in future work .

V. Conclusion

This paper proposes a method for objective prediction of speech intelligibility. The proposed method, NIC-STOI, allows using an intrusive intelligibility metric (STOI) without requiring access to the clean speech signal. Hence, NIC-STOI is essentially a non-intrusive metric. In principle, the method predicts the speech intelligibility by replacing the clean reference signal with an estimate of its spectrum. The features of the clean speech signal are estimated using a codebook-based approach, where the spectral shape of the speech is trained and parametrized using LPC. The proposed NIC-STOI metric shows a high correlation with the intrusive original STOI score and, hence, seems promising for predicting speech intelligibility non-intrusively using an intrusive intelligibility metric.

Acknowledgment

This work was supported by the Innovation Fund Denmark, Grant No. 99-2014-1.

References

[1] R.W. Peters, B.C.J. Moore, and T. Baer, "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," *J. Acoust. Soc. Am.*, vol. 103, no. 1, pp. 577–587, 1998.

[2] J.M. Festen and R. Plomp, "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *J. Acoust. Soc. Am.*, vol. 88, no. 4, pp. 1725–1736, 1990.

[3] V. Hamacher, J. Chalupper, E. Eggers, U. Kornagel, H. Puder, and U. Rass, "Signal processing in high-end hearing aids: State of the art, challenges, and future trends," *EURASIP J. Applied Signal Process.*, vol. 18, pp. 2915–2929, 2005.

[4] T.H. Falk, V. Parsa, J.F. Santos, K. Arehart, O. Hazrati, R. Huber, J.M. Kates, and S. Scollie, "Objective quality and intelligibility prediction for users of assistive listening devices: Advantages and limitations of existing tools," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 114–124, 2015.

[5] P.C. Loizou, *Speech Enhancement: Theory and Practice*, Signal processing and communications. Taylor & Francis, 2007.

[6] Y. Hu and P.C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *Speech Communication*, vol. 49, no. 78, pp. 588 – 601, 2007.

[7] N. French and J. Steinberg, "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Am.*, vol. 19, no. 1, pp. 90–119, 1947.

[8] H.J. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality," *J. Acoust. Soc. Am.*, vol. 67, no. 1, pp. 318–326, 1980.

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

[10] S. Jørgensen and T. Dau, "Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing," *J. Acoust. Soc. Am.*, vol. 130, no. 3, pp. 1475–1487, 1980.

[11] T.H. Falk, C. Zheng, and W.-Y. Chan, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 18, no. 7, pp. 1766–1774, 2010.

[12] F. Chen, O. Hazrati, and P.C. Loizou, "Predicting the intelligibility of reverberant speech for cochlear implant listeners with a non-intrusive intelligibility measure," *Biomedical Signal Processing and Control*, vol. 8, no. 3, pp. 311 – 314, 2013.

[13] C. Sørensen, A. Xenaki J.B. Boldt, and M.G. Christensen, "Pitch-based non-intrusive objective intelligibility prediction," in *ICASSP (to appear)*, March 2017.

[14] M. Karbasi, A.H. Abdelaziz, and D. Kolossa, "Twin-hmm-based non-intrusive speech intelligibility prediction," in *ICASSP*, March 2016, pp. 624–628.

[15] R.V. Shannon, F.G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, "Speech recognition with primarily temporal cues," *Science*, vol. 270, no. 5234, pp. 303–304, 1995.

[16] M.S. Kavalekalam, M.G. Christensen, F. Gran, and J.B. Boldt, "Kalman filter for speech enhancement in cocktail party scenarios using a codebook-based approach," in *ICASSP*, March 2016, pp. 191–195.

[17] S. Srinivasan, J. Samuelsson, and W.B. Kleijn, "Codebook-based bayesian speech enhancement for nonstationary environments," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 15, no. 2, pp. 441–452, 2007.

[18] S. Srinivasan, J. Samuelsson, and W.B. Kleijn, "Codebook driven short-term predictor parameter estimation for speech enhancement," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 14, no. 1, pp. 163–176, 2006.

[19] K.K. Paliwal and W.B. Kleijn, "Quantization of LPC parameters," in *Speech Coding and Synthesis*. 1995, pp. 433–468, Elsevier Science.

[20] D. Chan, A. Fourcin, B. Granstrom D. Gibbon, M. Huckvale, G. Kokkinakis, K. Kvale, L. Lamel, B. Lindberg, A. Moreno, J. Mouropoulos, F. Senia, I. Trancoso, C. Veld, and J. Zeiliger, "EUROM - a spoken language resource for the EU," in *Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Speech Technology*, 18-21 September 1995, vol. 1, pp. 867–870.

[21] Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Communications*, vol. 28, no. 1, pp. 84–95, 1980.