INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Harmonic Beamformers for Non-Intrusive Speech Intelligibility Prediction

Charlotte Sørensen[1,2], *Jesper B. Boldt*[2], *Mads G. Christensen*[1]

[1]Audio Analysis Lab, CREATE, Aalborg University, Denmark
[2]GN Hearing A/S, Lautrupbjerg 7, DK-2750, Ballerup, Denmark
{csoerensen,jboldt}@gnresound.com, mgc@create.aau.dk

Abstract

In recent years, research into objective speech intelligibility measures has gained increased interest as a tool to optimize speech enhancement algorithms. While most intelligibility measures are intrusive, i.e., they require a clean reference signal, this is rarely available in real-time applications. This paper proposes two non-intrusive intelligibility measures, which allow using the intrusive short-time objective intelligibility (STOI) measure without requiring access to the clean signal. Instead, a reference signal is obtained from the degraded signal using either a fixed or an adaptive harmonic spatial filter. This reference signal is then used as input to STOI. The experimental results show a high correlation between both proposed non-intrusive speech intelligibility measures and the original intrusively computed STOI scores.

Index Terms: hearing aids, non-intrusive, speech intelligibility prediction, STOI

1. Introduction

Speech intelligibility is an important property to consider, when developing signal processing for a wide range of applications, e.g., telecommunications [1, 2], and hearing aids [3]. As such, research into using objective measures of speech intelligibility as a tool to optimize speech enhancement algorithms has gained increased interest in recent years. There exists numerous different measures to estimate speech intelligibility with acceptable accuracy. The articulation index (AI) [4] and the speech transmission index [1] are some of the earliest measures to predict speech intelligibility of limited types of degradations such as linear filtering and additive noise. The speech-based envelope power spectrum model (sEPSM) [5] and the short-time objective intelligibility (STOI) [6] measure were recently introduced in order to increase the prediction accuracy for more complex degradation types. All the aforementioned measures are intrusive, i.e., in addition to the degraded signal they require access to a clean reference signal, which is rarely available in real-time applications.

This limitation has led to the proposal of non-intrusive speech intelligibility prediction measures, which do not require access to a clean reference signal. The speech to reverberation modulation energy ratio (SRMR) [7] provides an intelligibility prediction of reverberated speech based on the ratio between the energy of the low and high modulation frequency content. Similarly, the average modulation-spectrum area (ModA) [8] measure provides an intelligibility prediction based on the area of modulation spectrum of the degraded signal. Both these measures have been shown to perform well for conditions such as reverberation and additive noise compared to the previous non-intrusive measures [7, 3, 8].

Another approach to predict the speech intelligibility non-intrusively is to exploit a well-established and reliable intrusive metric, e.g. STOI [6], and obtaining an estimate of the clean speech reference from degraded signal. Recently, different approaches to estimate the reference signal have been proposed using machine learning [9, 10], spectral codebooks [11, 12], principal component analysis [13] and neural network [14] methods. These approaches have been shown to outperform the existing non-intrusive speech intelligibility prediction measures and to have a comparable performance to the intrusive measures [9, 12, 13, 14]. However, since these methods are all single channel and non-intrusive, they have no way of determining which speech signal is the desired target if multiple speakers are present given that the model is not trained for the specific speaker.

Using a multi-channel approach such as spatial filtering, i.e. beamforming, offers the possibility to overcome this limitation with a non-intrusive approach given the direction of the desired speech signal as proposed in [15]. The advantage of this method is that it has a very low complexity such that it can run on applications with low computational power, e.g. a hearing aid. On the other hand, the performance deteriorates with increasing number of interferers and reverberation. Similarly, the pitch-based STOI (PB-STOI) [16] measure also exploits the spatial content but instead of a filtering approach it reconstructs the reference signal from estimates of the properties of the signal model of the clean signal. It is based on a spatio-temporal model, which assumes the desired signal to be a sum of sinusoids whose frequencies are integral multiples of the pitch. Combining the spatial and the temporal characteristics (i.e., the direction of the desired signal its pitch) makes it more robust to competing speakers and reverberation, since it is possible to follow the pitch of the desired speech signal. PB-STOI has been shown to have a high correlation with the intrusive STOI scores even under adverse conditions with multiple interferers. However, the method also requires more computational power than the beamforming-based approach.

The present paper proposes new solutions to non-intrusive speech intelligibility prediction using, respectively, a fixed and an adaptive harmonic spatial filter based on a combination of the principles in [15, 16]. More specifically, the reference signal to be used as input to the intrusive framework STOI is obtained using model-based harmonic beamforming that resembles a filterbank designed for the given spatial and spectral characteristics of the desired signal. The rationale behind this approach is that the most energetic spectro-temporal regions, i.e. glimpses, occur during the voiced, i.e. harmonic, parts of speech. According to the glimpses model, intelligibility is related to the presence of such glimpses in which the most energetic regions are most important for speech intelligibility [17]. It is shown that the number of such glimpses correlates well with measured intelligibility and, thus, is a promising predictor for speech intelligibility [17].

http://dx.doi.org/10.21437/Interspeech.2019-2929

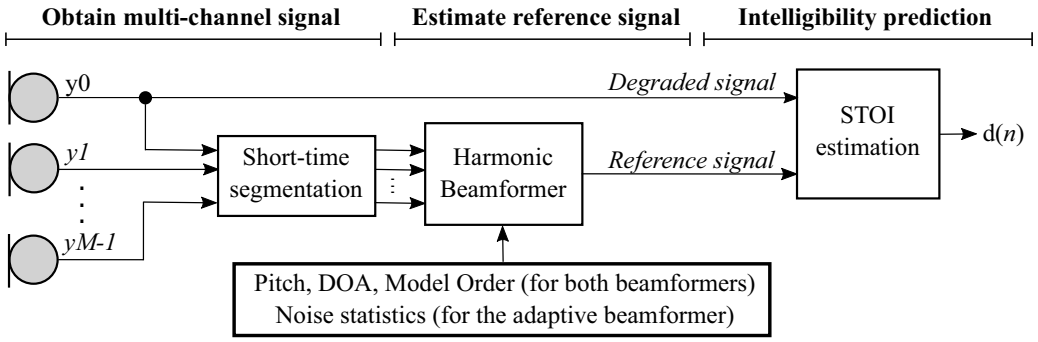

Figure 1: *Block diagram of the proposed harmonic beamforming-based non-intrusive intelligibility measures in which a reference signal is obtained using a harmonic spatial filter and compared with the output of an omnidirectional microphone with the intrusive STOI measure.*

2. Methods

This section presents the principles behind the proposed non-intrusive speech intelligibility predictions measures based on a fixed harmonic beamformer, dubbed the harmonic delay-and-sum beamformer-based STOI (HDSB-STOI), and an adaptive harmonic beamformer, dubbed the harmonic wiener beamformer-based STOI (HWB-STOI). Both HDSB-STOI and HWB-STOI allows predicting the speech intelligibility non-intrusively, i.e., without requiring access to the clean reference signal, by obtaining a reference signal from the degraded signal using a harmonic spatial filter and use this as input to STOI. Figure 1 depicts the general structure of both methods, which consists of three main steps: 1) Obtain a multi-channel signal using a microphone array, 2) Estimate a reference signal with the harmonic spatial filters and 3) Predict the speech intelligibility within the STOI framework.

2.1. Fundamentals

In the following, the signal model and the associated assumptions of the proposed methods are presented based on [18] in which a more thorough description of the theory behind the harmonic beamformers is available. In the proposed methods it is assumed that a uniform linear array (ULA) consisting of M microphones obtains the desired speech signal added to a mixture of interfering background noise and reverberation such that the samples of the mth microphone observations in a vector of frame length L is given by:

$$\mathbf{y}_m(t) = \mathbf{x}_m(t) + \mathbf{v}_m(t), \tag{1}$$

where $\mathbf{x}_m(t)$ and $\mathbf{v}_m(t)$ are vectors containing samples of the desired signal and the inference at the mth microphone, respectively.

The desired speech signal, $x_m(t)$, is modeled as a sum of sinusoids, i.e., a harmonic signal model, which is a good model for the voiced speech segments. Furthermore, using the harmonic model to obtain the desired signal does not only reduce the interfering sources but also reverberation, since spectral and temporal smearing of the signal source due to reverberation is not included in the harmonic model. As such, the desired speech signal is modeled as [18, 19]:

$$\mathbf{x}_m(t) = \mathbf{D}_{m,N}(\theta, \omega_0)\mathbf{a}(t, \omega_0), \tag{2}$$

where $\mathbf{D}_{m,N}(\theta, \omega_0)$ is a $L \times 2N$ matrix with the nth column being a vector of length L given by

$$\mathbf{d}_{m,n}(\theta, \omega_0) = e^{-\jmath n\omega_0 f_s \tau_m(\theta)} \times \left[\, 1 \;\; e^{-\jmath n\omega_0} \;\; \cdots \;\; e^{-\jmath n\omega_0 (L-1)} \,\right]^T, \tag{3}$$

where the superscript ^{T} is the transpose operator, N is the model order, $\jmath = \sqrt{-1}$ is the imaginary unit, ω_0 is the pitch or fundamental frequency, f_s is the sampling frequency and $\tau_m(\theta)$ is the relative delay of the desired source on the ULA. Furthermore, the complex amplitude $\mathbf{a}(t, \omega_0)$ is a vector of length $2N$ given by:

$$\mathbf{a}(t, \omega_0) = [\, a_{-N} e^{-\jmath N\omega_0 t} \;\; a_{-N+1} e^{-\jmath (N-1)\omega_0 t} \;\; \cdots \;\; a_N e^{\jmath N\omega_0 t} \,]^T, \tag{4}$$

and the correlation matrix of $\mathbf{a}$ (of size $2N \times 2N$) is

$$\mathbf{R}_{\mathbf{a}} = \mathrm{diag}\left(E\left[|a_{-N}|^2\right], E\left[|a_{-N+1}|^2\right], \ldots, E\left[|a_N|^2\right]\right), \tag{5}$$

and

$$\mathbf{R}_{\underline{\mathbf{v}}} = E\left[\underline{\mathbf{v}}(t)\underline{\mathbf{v}}^H(t)\right], \tag{6}$$

where $E[\cdot]$ is the mathematical expectation, and the superscript ^{H} is the conjugate-transpose operator.

Concatenating all the microphone signal vectors gives the vector of lenth ML:

$$\underline{\mathbf{y}}(t) = \underline{\mathbf{D}}_N(\theta, \omega_0)\mathbf{a}(t, \omega_0) + \underline{\mathbf{v}}(t), \tag{7}$$

where $\underline{\mathbf{y}}(t) = \left[\, \mathbf{y}_1^T(t) \;\; \mathbf{y}_2^T(t) \;\; \cdots \;\; \mathbf{y}_M^T(t) \,\right]^T$, $\underline{\mathbf{v}}(t) = \left[\, \mathbf{v}_1^T(t) \;\; \mathbf{v}_2^T(t) \;\; \cdots \;\; \mathbf{v}_M^T(t) \,\right]^T$ and

$$\underline{\mathbf{D}}_N(\theta, \omega_0) = \begin{bmatrix} \mathbf{D}_{1,N}(\theta, \omega_0) \\ \mathbf{D}_{2,N}(\theta, \omega_0) \\ \vdots \\ \mathbf{D}_{M,N}(\theta, \omega_0) \end{bmatrix}. \tag{8}$$

2.2. Harmonic delay-and-sum beamformer-based STOI (HDSB-STOI)

The harmonic delay-and-sum beamformer (DSB) is a fixed beamformer, which cannot adjust to the spatial characteristics of the interfering noise. It is advantageous for applications such

as hearing aids, since it only requires low computational power and does not require estimates of the noise statistics but, at least in theory, comes at a cost in performance [18]. The DSB can be deduced by maximizing the white noise gain subject to the distortionless constraint:

$$\min_{\underline{\mathbf{h}}} \underline{\mathbf{h}}^H \underline{\mathbf{h}} \quad \text{subject to} \quad \underline{\mathbf{h}}^H \underline{\mathbf{D}}_N(\theta_0, \omega_0) = \mathbf{1}_{2N}^T, \tag{9}$$

where $\mathbf{1}_{2N} = [\,1\ \ 1\ \cdots\ 1\,]^T$ is a vector of length $2N$.

Then, the DSB is derived as the optimal solution given by:

$$\underline{\mathbf{h}}_{\text{HDSB}} = \underline{\mathbf{D}}_N(\theta_0, \omega_0) \left[\underline{\mathbf{D}}_N^H(\theta_0, \omega_0) \underline{\mathbf{D}}_N(\theta_0, \omega_0) \right]^{-1} \mathbf{1}_{2N}. \tag{10}$$

2.3. Harmonic Wiener beamformer-based STOI (HWB-STOI)

The harmonic Wiener beamformer is an adaptive beamformer that can adapt to the spatial characteristics of the interfering noise, which in theory should give a better performance than fixed beamformers. However, it also needs access to the noise statistics and requires more computational power than the fixed beamformer.

The harmonic Wiener beamformer can be derived using the mean square error (MSE), which is given by [18]:

$$\begin{aligned} J(\underline{\mathbf{h}}) &= E\left[|e(t)|^2\right] && (11)\\ &= \mathbf{1}_{2N}^T \mathbf{R}_{\mathbf{a}} \mathbf{1}_{2N} && (12)\\ &\quad + \underline{\mathbf{h}}^H \underline{\mathbf{D}}_N(\theta_0, \omega_0) \mathbf{R}_{\mathbf{a}} \underline{\mathbf{D}}_N^H(\theta_0, \omega_0) \underline{\mathbf{h}} \\ &\quad - \underline{\mathbf{h}}^H \underline{\mathbf{D}}_N(\theta_0, \omega_0) \mathbf{R}_{\mathbf{a}} \mathbf{1}_{2N} \\ &\quad - \mathbf{1}_{2N}^T \mathbf{R}_{\mathbf{a}} \underline{\mathbf{D}}_N^H(\theta_0, \omega_0) \underline{\mathbf{h}} + \underline{\mathbf{h}}^H \mathbf{R}_{\underline{\mathbf{v}}} \underline{\mathbf{h}}, \end{aligned}$$

where the error signal between the estimated and desired signal, $e(t) = \left[\underline{\mathbf{h}}^H \underline{\mathbf{D}}_N(\theta_0, \omega_0) - \mathbf{1}_{2N}^T\right] \mathbf{a}(t, \omega_0) + v_{\text{rn}}(t)$, is the sum of the signal distortion and the residual noise.

Finally, the optimal solution for the harmonic Wiener beamformer can be found by differentiating the MSE, $J(\underline{\mathbf{h}})$ [eq. (11)], with respect to $\underline{\mathbf{h}}$ and setting the result equal to zero:

$$\underline{\mathbf{h}}_{\text{HWB}} = \left[\underline{\mathbf{D}}_N(\theta_0, \omega_0) \mathbf{R}_{\mathbf{a}} \underline{\mathbf{D}}_N^H(\theta_0, \omega_0) + \mathbf{R}_{\underline{\mathbf{v}}} \right]^{-1} \times \underline{\mathbf{D}}_N(\theta_0, \omega_0) \mathbf{R}_{\mathbf{a}} \mathbf{1}_{2N}. \tag{13}$$

2.4. Pitch-based STOI (PB-STOI)

The results of the proposed HDSB-STOI and HWB-STOI are compared with the non-intrusive PB-STOI measure proposed in [16], where a more detailed description is available. Similar to the proposed methods in this paper, PB-STOI is based on a harmonic model that takes the spatial input into account:

$$\mathbf{y}_m = \beta_m \mathbf{Z} \mathbf{D}(m) \boldsymbol{\alpha} + \mathbf{v}_m, \tag{14}$$

where β_m is the attenuation of the desired source at the m'th microphone, $\mathbf{Z} = [\mathbf{z}(\omega_0) \ldots \mathbf{z}(L\omega_0)]$, $\mathbf{z}(l\omega_0) = [1 \ldots e^{jl\omega_0(N-1)}]$, $\mathbf{D}(m) = \text{diag}([e^{-j\omega_0 f_s \tau_k} \ldots e^{-jL\omega_0 f_s \tau_m}])$ for $l = 1, \ldots, L$ with all other entries equal to zero and $\mathbf{v}_m$ is the sum of the recorded noise and interference.

Based on the signal model, the attenuation factor, the complex amplitude, the variance and the pitch is estimated in an iterative manner. These parameters are then used to directly to reconstruct a reference signal from the signal model. Finally, the reconstructed reference signal is applied as input to STOI instead of the clean signal.

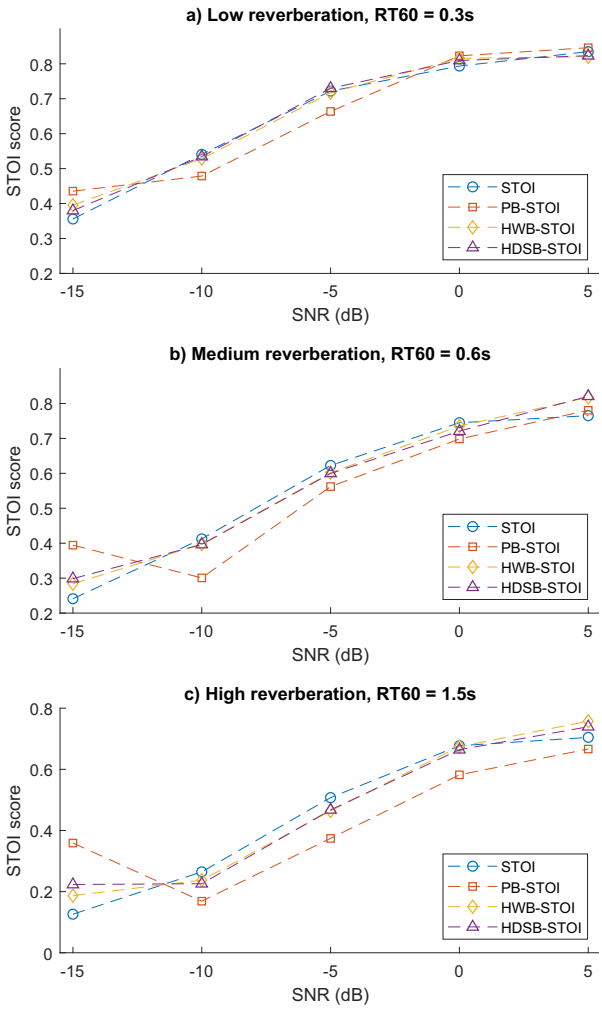

Figure 2: *Performance shown in terms of estimated STOI score as function of the SNR in dB for a) Low reverberation with RT60 = 0.3 s, b) Medium reverberation with RT60 = 0.6 s and c) High reverberation with RT60 = 1.5 s. The results of STOI, PB-STOI, HWB-STOI and HDSB-STOI is given by the blue circles, red squares, yellow diamonds and purple triangles, respectively.*

3. Experimental results

The proposed measures are evaluated using a broadside ULA setup consisting of $M = 10$ omnidirectional microphones with a microphone spacing of $d = c/f_s$, where the speech of sound in air is $c = 343$ m/s and the sampling frequency is $f_s = 8$ kHz. The direction of arrival (DOA) of the desired source was $\theta = 0^\circ$ resulting in a $\tau_m = 0$. The pitch is efficiently estimated using the multi-channel maximum likelihood pitch estimator proposed in [20] as the sum over the squared magnitude of the FFT of $y_m(t)$, denoted $Y_m(\omega_0)$, evaluated at a set of candidate harmonics, Ω_0, which is given by $\hat{\omega}_0 = \arg\max_{\omega_0 \in \Omega_0} \sum_{l=1}^{L} \sum_{m=1}^{M} |Y_m(\omega_0 l)|^2$ when assuming that the DOA is coming from the front, the noise variance is known and the same for all channels. The pitch is evaluated in the range $\Omega_0 = 80 - 400$ Hz and the model order, $L = 10$. In the experimental evaluation, a set of 50 English sentences (both male and female) from the EUROM_1 database [21] is used for

both the desired source and interfering speakers. The sentences contain both voiced and unvoiced segments. The signals are 5.0 s long and are processed in segments of 20 ms with 50 % overlap. The toolbox McRoomSim [22] is used to create the simulations of a complex multi-talker scenario with 8 interfering speakers in a room with dimensions of 10x6x4 m similar to the evaluation setup in [16]. The simulations are carried out at three different levels of reverberation ranging from low to high (RT60 = 0.3 s, RT60 = 0.6 s and RT60 = 1.5 s) at signal-to-noise ratios (SNRs) ranging from -15 to 5 dB. A white Gaussian noise is added to each microphone channel at a SNR of 20 dB.

The performances of the proposed non-intrusive intelligibility measures are evaluated against the original intrusively computed STOI scores as the ground truth. The results are shown in Figures 2(a), 2(b) and 2(c) for the low, medium and high reverberation scenarios, respectively. The results of the PB-STOI (red squares), HWB-STOI (yellow diamonds) and HDSB-STOI (purple triangles) are plotted together with the intrusively computed STOI scores indicated by the blue circles. At low reverberation all of the three non-intrusive measures show a good performance even though the harmonic beamforming-based non-intrusive speech intelligibility measures both outperform the PB-STOI measure. As reverberation increases, the performance of PB-STOI deteriorates and especially at low SNRs it predicts an increase in speech intelligibility rather than a decrease in intelligibility as predicted by the intrusive STOI measure. It is obvious that both of the proposed HDSB-STOI and HWB-STOI measures yield more suppression of interfering speakers and reverberation compared with the PB-STOI measure even though there is a slight decrease in performance at low SNRs with increasing reverberation levels for both measures. While the PB-STOI measure is also based on a harmonic model and should, thus, also perform well in reverberation, the difference in performance is likely due to PB-STOI being more sensitive to errors in the pitch estimate, since it is based on a reconstruction of the reference signal rather than a filtering approach.

Notably, the measures based on the fixed and adaptive approach perform almost equally well. The performance of the adaptive Wiener beamformer is only slightly better at low SNRs at high reverberation levels. Even though the adaptive beamformer in theory should have a better performance this is not necessarily the case in practical performance, since it relies on estimates of the noise statistics. This is also supported by the findings in [18], where the adaptive beamformers provide a slightly lower SNR gain compared to the harmonic DSB. As such, due to being computational efficient and simple, i.e. not requiring noise statistics, the HDSB-STOI measure might be the best choice depending on the applications, e.g. hearing aids, given the comparable performance to the HWB-STOI measure.

4. Conclusions

This paper proposes two approaches, the harmonic delay-and-sum beamformer-based STOI (HDSB-STOI) and the harmonic wiener beamformer-based STOI (HWB-STOI), for non-intrusive prediction of speech intelligibility. The HDSB-STOI measure and the HWB-STOI measure estimate a reference signal from the degraded signal using a fixed and an adaptive harmonic spatial filter, respectively. The estimated reference signal is then used as input to the established and thoroughly evaluated intrusive measure STOI, which requires a clean reference signal. Both of the proposed non-intrusive measures have a high correlation with the original intrusively computed STOI scores.

5. Acknowledgements

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

6. References

[1] 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.

[2] S. Jørgensen, J. Cubick, and T. Dau, "Speech intelligibility evaluation for mobile phones." *Acustica United with Acta Acustica*, vol. 101, p. 1016 1025, 2015.

[3] 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.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

[10] D. Sharma, Y. Wang, P. A. Naylor, and M. Brookes, "A data-driven non-intrusive measure of speech quality and intelligibility," *Speech Communication*, vol. 80, pp. 84–94, 2016.

[11] C. Sørensen, M. S. Kavalekalam, A. Xenaki, J. B. Boldt, and M. G. Christensen, "Non-intrusive intelligibility prediction using a codebook-based approach," in *EUSIPCO*, 2017, pp. 216–220.

[12] ——, "Non-intrusive codebook-based intelligibility prediction," *Speech Communication*, vol. 101, pp. 85 – 93, 2018.

[13] A. H. Andersen, J. M. de Haan, Z. Tan, and J. Jensen, "A non-intrusive short-time objective intelligibility measure," in *ICASSP*, March 2017, pp. 5085–5089.

[14] ——, "Nonintrusive speech intelligibility prediction using convolutional neural networks," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 26, no. 10, pp. 1925–1939, 2018.

[15] C. Soerensen, J. B. Boldt, F. Gran, and M. G. Christensen, "Semi-non-intrusive objective intelligibility measure using spatial filtering in hearing aids," in *EUSIPCO*, August 2016, pp. 1358–1362.

[16] C. Sørensen, A. X. J. B. Boldt, and M. G. Christensen, "Pitch-based non-intrusive objective intelligibility prediction," in *ICASSP*, March 2017, pp. 386–390.

[17] M. Cooke, "A glimpsing model of speech perception in noise," *The Journal of the Acoustical Society of America*, vol. 119, no. 3, pp. 1562–1573, 2006.

[18] J. R. Jensen, S. Karimian-Azari, M. G. Christensen, and J. Benesty, "Harmonic beamformers for speech enhancement and dereverberation in the time domain," *Submitted to Speech Communication*, 2018.

[19] M. G. Christensen, "Accurate estimation of low fundamental frequencies from real-valued measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 21, no. 10, pp. 2042–2056, 2013.

[20] M. G. Christensen, "Multi-channel maximum likelihood pitch estimation," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 409–412.

[21] D. Chan, A. Fourcin, B. G. 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*, vol. 1, 18-21 September 1995, pp. 867–870.

[22] A. Wabnitz, N. Epain, C. Jin, and A. Van Schaik, "Room acoustics simulation for multichannel microphone arrays," in *Proceedings of the International Symposium on Room Acoustics*, 2010, pp. 1–6.