

SEMI-NON-INTRUSIVE OBJECTIVE INTELLIGIBILITY MEASURE USING SPATIAL FILTERING IN HEARING AIDS

Charlotte Sørensen^{1,2}, Jesper B. Boldt², Fredrik Gran² and Mads G. Christensen¹

¹Audio Analysis Lab, AD:MT, Aalborg University, Denmark

²GN ReSound A/S, Lautrupbjerg 7, DK-2750, Ballerup, Denmark

{csoerensen, jboldt}@gnresound.com, {mgc}@create.aau.dk

ABSTRACT

Reliable non-intrusive online assessment of speech intelligibility can play a key role for the functioning of hearing aids, e.g. as guidance for adjusting the hearing aid settings to the environment. While existing intrusive metrics can provide a precise and reliable measure, the current non-intrusive metrics have not been able to achieve acceptable intelligibility predictions. This paper presents a new semi-non-intrusive intelligibility measure based on an existing intrusive measure, STOI, where an estimate of the clean speech is extracted using spatial filtering in the hearing aid. The results indicate that the STOI score obtained with the proposed method using an estimate of the clean speech correlates well with the STOI score having the original clean speech signal available.

Index Terms— Non-intrusive objective intelligibility measure, generalized sidelobe canceller, hearing aids

1. INTRODUCTION

For users of hearing aids speech intelligibility depends highly on the specific listening environment. One of the main issues is significantly decreased speech intelligibility in noisy multi-talker environments termed the "cocktail party problem" [1, 2]. Therefore, a lot of research has gone into the development of various speech enhancement algorithms (e.g., noise and echo suppression) to overcome this challenge. However, noise suppression techniques, such as adaptive directional filtering, can have a negative impact on localization performance of hearing aid users [3]. The fact that hearing aid users receive distorted localization cues can lead to decreased intelligibility due to losing a binaural advantage of 3-12 dB [4, 3]. As such, it is important to quantify, whether the gain from the noise suppression techniques are advantageous if localization cues are lost in return by assessing the intelligibility of the current environment. For the users of assistive listening devices it would be a great benefit, if the devices were able to automatically detect when advanced speech enhancement actually provides an improvement and adjust the

hearing aid settings accordingly. Generally, the remaining hearing of the hearing aid user should be relied on as much as possible such that speech enhancement processing is limited to when it provides a benefit and the proposed method could facilitate exactly this. Fast and robust online evaluation of the listening environment could assure that speech enhancement processing is only applied when necessary and selected without requiring an action of the hearing aid user [5, 6]. As such, the proposed method can be seen as an alternative to environment classification based on intelligibility rather than classifying the different environments [7].

Thus, it would be preferable if objective intelligibility measures could become a crucial part of the online processing of assistive listening devices. Intrusive objective measures (e.g., the short-time objective intelligibility (STOI) metric [8], the normalized covariance metric (NCM) [9]) with access to both the clean and noisy speech can generally provide a precise and reliable measure for the speech intelligibility [6]. However, online processing in a hearing aid requires a non-intrusive objective measure, since access to the clean speech is rarely available. Over the years a number of non-intrusive metrics have been developed (e.g., the modulation spectrum area (ModA) [10], the speech-to-reverberation modulation energy ratio (SRMR) [11]). However, according to a recent comprehensive review none of the tested the existing non-intrusive measures have achieved acceptable results [6].

This paper is concerned with a method in between the intrusive and non-intrusive technique that can be processed online in a hearing aid while taking advantage of the reliability of existing intrusive metrics. The approach is to extract an estimate of the clean speech with directional spatial filtering in the hearing aid and use this in existing intrusive objective intelligibility metrics. In other words, an estimate of the intelligibility is obtained by comparing the output of a beamformer at the direction of the desired talker with the output of an omnidirectional microphone using an existing objective measure such as STOI. The online processed intelligibility prediction of the specific environment can then be used to determine, whether the intelligibility is below a certain threshold and apply speech enhancement processing when it is beneficial.

This work was supported by the Innovation Fund Denmark.

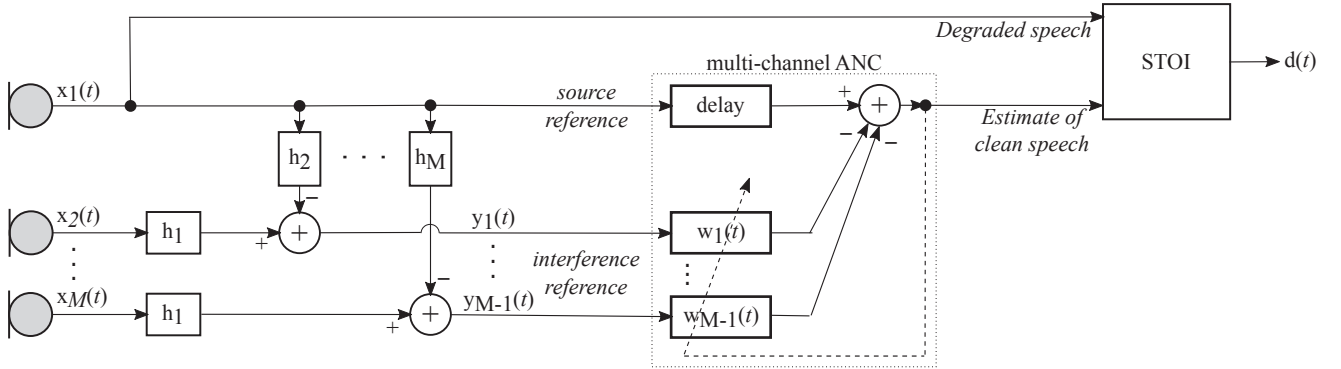


Fig. 1. Block diagram of the proposed semi-non-intrusive objective intelligibility measure in which an estimate of the clean speech is extracted with the GSC structure and compared with the output of an omnidirectional microphone using STOI.

2. METHOD

In this section the approach and method behind the proposed semi-non-intrusive objective intelligibility measure is presented. A block diagram incorporating the whole semi-non-intrusive objective intelligibility measure with both the beamformer and the existing intrusive intelligibility measure STOI is shown in Figure 1. The principles behind the beamforming structure and notation are explained in Section 2.1. The STOI metric gives a prediction, $d(t)$, of the speech intelligibility on a 0-1 scale by comparing the correlation of a clean and degraded version of the same speech signal [8]. As illustrated in the diagram the noisy signal from an omnidirectional microphone is both used as the degraded speech input to STOI as well as reference of the source to an adaptive noise cancellation (ANC) stage in the beamformer. The remaining microphone signals are used in a fixed spatial filtering stage in the beamformer to extract a reference of the interference.

2.1. Generalized sidelobe cancellation

An estimate of the clean speech is obtained using a widely applied beamformer for hearing aid applications based on the generalized sidelobe cancellation (GSC) structure [12, 13, 14]. The beamformer has four microphones by exploiting the front and rear microphone of two BTE hearing aids assuming a bilateral wireless link between them. The implemented GSC structure consists of a fixed spatial preprocessor and an ANC unit similar to the approach of [14] extended to four microphones as illustrated in Figure 1 with $M = 4$.

It is assumed that each microphone signal x_k , $k = 1, \dots, M$ is the desired source additively interfered with a number of interferers, N i.e.

$$x_k(t) = h_k * s(t) + \sum_{n=1}^N h_{k,n}^{\text{interf}} * s_n^{\text{interf}}(t) \quad (1)$$

where h_k and $h_{k,n}^{\text{interf}}$ are the acoustic impulse responses between the k th microphone and the desired source, $s(t)$, and

interferers, $s_n^{\text{interf}}(t)$, respectively and $*$ denotes convolution. Ambient noise can be created by adding up multiple interferers with reverberation included in the acoustic impulse responses.

During periods of interference-only, $s(t) = 0$, each microphone signal is the sum of the interferers convolved with the acoustic impulse response between each interferer and the k th microphone, i.e.

$$x_k(t) = \sum_{n=1}^N h_{k,n}^{\text{interf}} * s_n^{\text{interf}}(t) \quad (2)$$

A reference of the interference is created by steering a zero towards the direction of the desired speaker. The location of the desired speaker is assumed to be in the front of the listener at zero degrees but can easily be relaxed to other positions. The desired source is canceled using spatial filters, which give an estimate of interference-only at the k th microphone for $k = 2, \dots, M$, where h_1 is the acoustic impulse response between the desired source at 0° and the first microphone:

$$\begin{aligned} y_{k-1}(t) &= x_k(t) * h_1 - x_1(t) * h_k \\ &= h_k * s(t) * h_1 + \sum_{n=1}^N h_{k,n}^{\text{interf}} * s_n^{\text{interf}}(t) * h_1 \\ &\quad - (h_1 * s(t) * h_k + \sum_{n=1}^N h_{1,n}^{\text{interf}} * s_n^{\text{interf}}(t) * h_k) \\ &= \sum_{n=1}^N h_{k,n}^{\text{interf}} * s_n^{\text{interf}}(t) * h_1 - \sum_{n=1}^N h_{1,n}^{\text{interf}} * s_n^{\text{interf}}(t) * h_k \end{aligned} \quad (3)$$

where y_{k-1} , $k = 2, \dots, M$ is the interference reference at the k th microphone. It can be seen that the filters block out $s(t)$ in the derivation of y_{k-1} . The coefficients of the blocking filters have been determined based on the impulse responses between the source at 0° and the k th microphone measured on a KEMAR artificial head and torso as described in Section 3.

The ANC unit attenuates the interference in the desired source reference that is correlated with the interference reference using the filters $\mathbf{w}_k(t) = [w_{k,1}(t), w_{k,2}(t), \dots, w_{k,L}(t)]$, where L is the length of the filter. The ANC unit is updated with a least squares (LS) approach but can in online processing easily be implemented as a least mean square (LMS) algorithm.

The incorporation of the fixed spatial filter in the preprocessor reduces the amount of speech leakage into the interference reference but cannot completely prevent it [15, 13]. Therefore, the ANC is adapted during periods of interference-only in order to avoid possible cancellation of the desired speech source. For this purpose a robust speech detector is assumed available in this paper.

3. EXPERIMENTAL METHODOLOGY

The acoustic impulse responses have been measured using the front and rear microphones on a GN ReSound Alera 312 BTE hearing aid on a KEMAR artificial head and torso in an anechoic room with a maximum length sequence (MLS) with a code length of 11 and averaged over 30 repetitions. The KEMAR artificial head and torso was rotated in the horizontal plane with a resolution of 2 degrees using a Brüel & Kjær Turntable system type 9640.

The speech samples of both the desired source and the interferers were taken from the EUROM_1 database as 5 second recordings of the English sentence corpus [16]. The level of the interferers were varied according to the level of the desired speech source as the source-to-interference ratio (SIR) [17]. The clean speech of the desired source was convolved with the acoustic impulse responses from 0° to each microphone and the interfering speech sources were convolved with the impulse responses from 140° , 270° , 50° and 300° for one, two, three or four speakers, respectively. Compared to current state-of-the-art studies four interferers can be considered a relatively complex scenario with speech-on-speech masking being a difficult task [1, 2, 6].

4. RESULTS

The performance of the proposed semi-non-intrusive objective intelligibility measure is evaluated by comparing the STOI score of the noisy speech obtained using the estimate of the clean speech as reference with the STOI score obtained using the original clean speech as reference. Figure 2-5 show the STOI scores as function of SIR for one, two, three and four interferers, respectively. For one interferer located at 140° (Figure 2) it can be seen that the STOI score obtained using the output from the implemented GSC beamformer as reference (dashed line) correlates well with the STOI scores obtained with access to the original clean speech signal (solid line) for all SIRs.

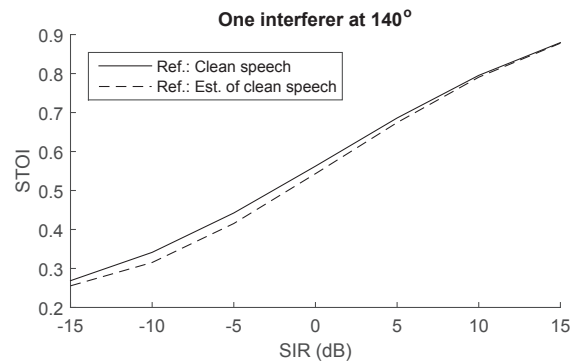


Fig. 2. STOI score as function of SIR with one interferer at 140° using the clean speech signal (solid line) and the estimate of the clean speech extracted with the implemented 4 microphone GSC beamformer (dashed line) as reference.

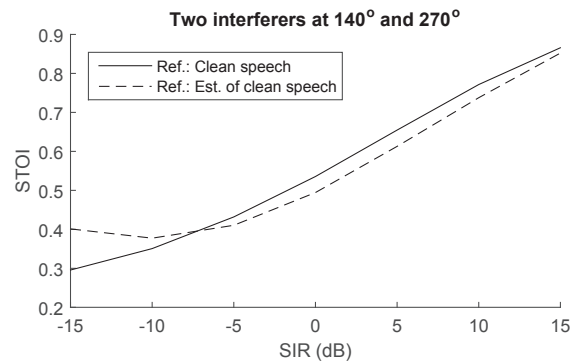


Fig. 3. STOI score as function of SIR with two interferers at 140° and 270° using the clean speech signal (solid line) and the estimate of the clean speech extracted with the implemented 4 microphone GSC beamformer (dashed line) as reference.

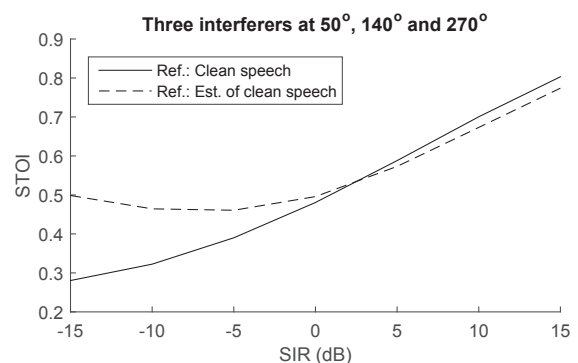


Fig. 4. STOI score as function of SIR with three interferers at 50° , 140° and 270° using the clean speech signal (solid line) and the estimate of the clean speech extracted with the implemented 4 microphone GSC beamformer (dashed line) as reference.

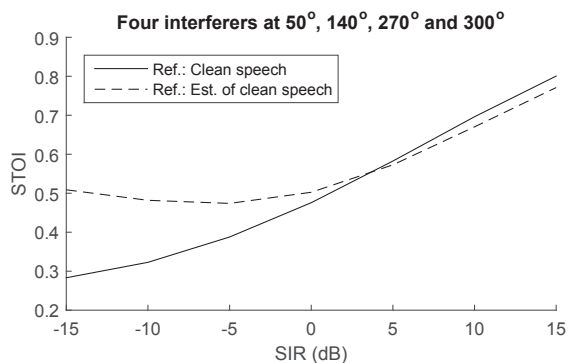


Fig. 5. STOI score as function of SIR with four interferers at 50° , 140° , 270° and 300° using the clean speech signal (solid line) and the estimate of the clean speech extracted with the GSC beamformer (dashed line) as reference.

In the case of two interferers located at 140° and 270° (Figure 3) the STOI score using the beamformed estimate of the clean speech as reference correlates well with the intrusive STOI score having access to the clean speech signal for STOI scores and SIRs higher than 0.4 and -5 dB, respectively. However, below this level the output from STOI using the estimate of the clean speech as reference starts to deviate from the STOI score obtained using the original clean speech as reference. In the cases of three interferers at 50° , 140° and 270° (Figure 4) and four interferers at 50° , 140° , 270° and 300° (Figure 5) the STOI scores with the estimate of the clean speech and the original clean speech as references, respectively, correlates well for STOI scores above 0.5 and SIRs above 5 dB but deviates below these levels. Noteworthy, the performance of the proposed method does not decrease substantially going from the case with three interferers to four interferers.

5. DISCUSSION

A reliable objective intelligibility measure in the online processing of hearing aids could be of great advantage to predict whether speech enhancement would provide a benefit for the user and adjust the hearing aid settings accordingly. Online processing would require a non-intrusive metric and even though a number of promising non-intrusive measures have been developed over the recent years none of them have achieved sufficient results for the purpose [6, 11, 10]. Previous studies have shown that STOI scores correlate well with subjective intelligibility scores and thus gives a reliable estimate for the speech intelligibility [8, 6]. As such, a non-intrusive measure performing similarly to STOI could yield a promising method for online processing of speech intelligibility in hearing aids. The intelligibility scores obtained with the proposed semi-non-intrusive technique correlates well with the intrusive STOI scores obtained with access to the

clean speech for STOI scores above 0.5 but deviates for lower scores. This may or may not be a problem for the intended purpose provided it reflects so little speech intelligibility that it conforms to the threshold for applying speech enhancement anyway. A STOI score below 0.6 may correspond to very low speech intelligibility depending on the speech material and the psychometric function relating STOI scores to subjective scores [6]. Furthermore, the proposed method could easily be implemented in today's hearing aids. The acoustic impulse responses used for the spatial filter design in the blocking matrix could either be the standard acoustic impulse responses measured on KEMAR or personalized acoustic impulse responses measured during adjustment of the hearing aid.

In future work it could be interesting to test the proposed method with added reverberation as this is known to affect the performance of the GSC beamformer [13, 14]. In order to properly simulate reverberation 3 dimensional acoustic room impulses would be required. Additionally, the objective intelligibility scores obtained with the proposed semi-non-intrusive technique could be tested against subjective listening tests in future work. In a similar manner to using the proposed method for prediction of the speech intelligibility the same approach could be used to evaluate speech quality with e.g. the perceptual evaluation of speech quality (PESQ [18]) by using the estimate of the clean speech to evaluate the speech quality before and after speech processing in the hearing aid. Furthermore, the proposed method could also be extended to include personalized hearing losses in the speech intelligibility prediction similarly to the technique in e.g. the hearing-aid speech perception index (HASPI) [19].

Recently, binaural speech intelligibility methods have with limited success attempted to predict the speech intelligibility by including the effects of spatial masking [20]. The proposed technique in this paper does not take advantage of the multiple channels used in the beamformer to predict the effects of spatial masking on the speech intelligibility. In future work this could be an interesting extension of the proposed technique.

6. CONCLUSION

This paper has presented a new feasible technique for online processing of speech intelligibility in hearing aids. The technique is based on an existing intrusive objective metric, where an estimate of the clean speech to be used as reference is obtained using a GSC structure with spatial filters as blocking matrix. The GSC structure is implemented using the front and rear microphones on two wirelessly linked BTE hearing aids. The results indicate that the obtained STOI scores using the estimate of the clean speech as reference correlate well with the intrusive STOI having access to the original clean speech for STOI scores above 0.5. Thus, the proposed method yields a promising and feasible technique for online processing of speech intelligibility in hearing aids.

7. 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] T. Van den Bogaert, T.J. Klaseen, M. Moonen, L. Van Deun, and J. Wouters, "Horizontal localization with bilateral hearing aids: Without is better than with," *J. Acoust. Soc. Am.*, vol. 119, no. 1, pp. 515–526, 2006.
- [4] M.L. Hawley, R.Y. Litovsky, and J.F. Culling, "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," *J. Acoust. Soc. Am.*, vol. 115, no. 2, pp. 833–843, 2006.
- [5] 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.
- [6] 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.
- [7] Luc Lamarche, Christian Gigure, Wail Gueaieb, Tyseer Aboulnasr, and Hisham Othman, "Adaptive environment classification system for hearing aids," *The Journal of the Acoustical Society of America*, vol. 127, no. 5, pp. 3124–3135, 2010.
- [8] 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.
- [9] R.L. Goldsworthy and J.E. Greenberg, "Analysis of speech-based speech transmission index methods with implications for nonlinear operations," *J. Acoust. Soc. Am.*, vol. 116, no. 6, pp. 3679–3689, 2004.
- [10] 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.
- [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] L.J. Griffiths and C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. 30, no. 1, pp. 27–34, 1982.
- [13] A. Spriet, M. Moonen, and J. Wouters, "Robustness analysis of multichannel wiener filtering and generalized sidelobe cancellation for multimicrophone noise reduction in hearing aid applications," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 4, pp. 487–503, 2005.
- [14] A. Spriet, L. Van Deun, K. Eftaxiadis, J. Laneau, M. Moonen, B. van Dijk, A. van Wieringen, and J. Wouters, "Speech understanding in background noise with the two-microphone adaptive beamformer beam in the nucleus freedom cochlear implant system," *Ear & Hearing*, vol. 28, pp. 62–71, 2007.
- [15] S. Nordebo, I. Claesson, and S. Nordholm, "Adaptive beamforming: Spatial filter designed blocking matrix," *IEEE J. Ocean. Eng.*, vol. 19, no. 14, pp. 583–590, 1994.
- [16] 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.
- [17] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.
- [18] 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 *Proc IEEE Int Conf Acoust Speech Signal Process*, 2001, vol. 2, pp. 749–752.
- [19] J.M. Kates and K.H. Arehart, "The hearing-aid speech perception index (haspi)," *Speech Commun.*, vol. 65, pp. 75–93, 2014.
- [20] R. Beutelmann, T. Brand, and B. Kollmeier, "Revision, extension, and evaluation of a binaural speech intelligibility model," *J. Acoust. Soc. Am.*, vol. 127, no. 4, pp. 2479–2497, 2010.