AN IMPROVED NON-INTRUSIVE INTELLIGIBILITY METRIC FOR NOISY AND REVERBERANT SPEECH

João F. Santos, Mohammed Senoussaoui, Tiago H. Falk

INRS-EMT, University of Quebec, Montreal, QC, Canada

ABSTRACT

Non-intrusive speech intelligibility metrics are based solely on the corrupted speech information and a prior model of the speech signal in a given representation. As such, any sources of variability not taken into account by the model will affect the metric's performance. In this paper, we investigate two sources of variability in the auditory-inspired model used by the speech-to-reverberation modulation energy ratio (SRMR) metric, namely speech content and pitch, and propose two updates that aim to reduce the variability caused by these sources. First, we limited the dynamic range of the energies in the modulation spectrum bands in order to reduce the effect of speech content and speaker variability. Second, the range of the modulation filter bank was modified to reduce the variability due to pitch. Experimental results show that the updated metric presents higher performance and lower variability relative to the original SRMR when assessing speech intelligibility in noisy and reverberant environments, as well as outperforms several standard intrusive and non-intrusive benchmark metrics.

***Index Terms*—** Speech intelligibility, objective metrics, modulation spectrum

1. INTRODUCTION

Room acoustics, particularly reverberation and noise, severely degrade the performance of far-field speech technologies, such as speech recognition. For hearing aid and cochlear implant users, in turn, reverberation and noise alter speech temporal envelopes, thus reducing intelligibility to unacceptable levels. To overcome these issues, reverberation and noise suppression (i.e., speech enhancement) algorithms are often employed. In order to gauge the effects of speech enhancement on noisy and reverberant speech, algorithm developers typically rely on subjective listening tests where listeners either rate the quality or are asked to transcribe the speech signal heard. In the latter scenario, correctly identified words are commonly used as a measure of speech intelligibility.

Subjective listening tests, however, despite their accuracy and reliability if carefully crafted, are expensive, laborious and time-intensive, as well as unsuitable for real-time monitoring purposes. As such, objective metrics have been the focus of recent research with so-called intrusive and non-intrusive models being developed based on the need, or not, of a clean reference signal, respectively. While a number of methods and algorithms have been developed for telephone speech, only a handful have been proposed for reverberant speech. In [1, 2, 3], several metrics were tested assuming both normal and impaired listeners (e.g., cochlear implantees). In these tests, a non-intrusive measure called speech-to-reverberation modulation energy ratio (SRMR) stood out as a reliable candidate [4]. Variants of the SRMR measure have since been proposed for blind reverberation time and direct-to-reverberant energy ratio estimation [5].

Despite its high correlation with subjective quality and intelligibility ratings, the SRMR metric has been shown to be sensitive to inter- and intra-speaker variability [6, 7]. This limitation arises from the fact that the measure does not rely on a reference signal, thus is sensitive to prior assumptions made by the model. In this case, the main assumption is that room acoustics effects show up in higher speech envelope frequencies (termed modulation frequencies) whereas speech components show up in lower modulation frequency regions (≤ 16 Hz). In this paper, we investigate two sources of variability in the auditory-inspired SRMR model, namely pitch and speech content, and propose two updates to mitigate this variability. Experimental results show significant improvements with the updated metric relative to its original counterpart, both in terms of increased correlation with subjective ratings, as well as reduced variability in estimation errors.

The remainder of this paper is organized as follows. Section 2 describes the original SRMR metric, its limitations, as well as the proposed updates. Sections 3-5 present the experimental setup, results, and conclusions, respectively.

2. PROPOSED UPDATES TO THE SRMR METRIC

Here, we describe the original SRMR implementation, present two sources of model variability, and propose updates to reduce the variability of intelligibility prediction errors.

2.1. SRMR: Original implementation

It is well known that the modulation envelopes of a speech signal provide useful cues for objective speech quality and intelligibility estimation. The modulation energy of clean anechoic speech is typically concentrated in lower frequencies, from 2-20 Hz, with a spectral peak at 4 Hz [8]. Speech under the effects of reverberation and/or noise, in turn, will exhibit temporal envelopes with higher frequency components [5]. The SRMR metric explores this effect and relies on the ratio between the energy in higher to lower modulation frequencies to predict speech intelligibility [4].

The auditory-inspired modulation spectrum representation used by SRMR is computed as follows. First, the input speech signal is decomposed in 23 acoustic channels by a gammatone filterbank with filter center frequencies ranging from 125 Hz to approximately half the sampling frequency, and with bandwidths characterized by the equivalent rectangular bandwidth, ERB [9]. Each acoustic channel has its temporal envelope extracted (via the Hilbert transform). In order to emulate frequency selectivity in the modulation domain [10], each envelope is decomposed into eight overlapping modulation bands with center frequencies logarithmically spaced between $4 - 128$ Hz. Modulation spectral energy is then computed for each of the filtered envelopes (corresponding to modulation band/acoustic band pairs) as the squared magnitude of its discrete Fourier transform for 256 ms frames with 32 ms overlap. Lastly, the SRMR value is computed as the ratio of the average modulation energy content (over all frames) in the first four modulation bands ($3-20$ Hz) to the average modulation energy content available in the last four modulation bands ($20-160$ Hz).

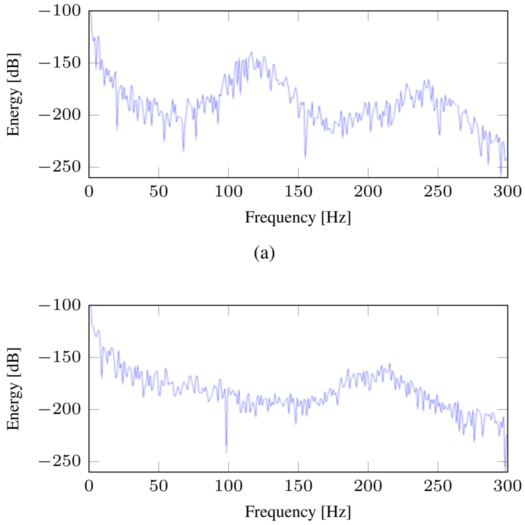
2.2. Investigating sources of SRMR variability: pitch and speech content

Previous studies have shown that the energy envelope of the speech signal exhibits a structure with periodicity equal to its fundamental frequency [11]. While these results were obtained using full-band speech envelopes, it is suspected that pitch effects are also present in the acoustic subband model used by the SRMR metric. Secondly, subband based modulation features have been used in the past for speech recognition [12], thus suggesting that the model used by the SRMR metric may be sensitive to speech content, such as different phonemes. While the SRMR metric utilizes significantly larger analysis windows (256 ms against 32 ms zero-padded windows in [12]) to reduce such speech-content sensitivity, it is suspected that some residual dependency remains.

To investigate the effect of speech content and pitch variability on the SRMR metric, we processed clean speech data from two different databases. The first database consisted of consonant-vowel pairs (CVs), and contained 1,728 samples [13]. Four talkers (2 males and 2 females) recorded 8 tokens for each of 18 consonants (fricatives: s, z, ʃ, ʒ, f, v, θ, ð; stops: p, t, k, b, d, g,; nasals m, n; affricates: ʧ, ʤ) in 3 vowel contexts (ɑ, i, and u). To evaluate the variability over sentences we used a second database, based on a subset of the TIMIT corpus consisting of 160 anechoic, noise-free sentence recordings from 8 male and 8 female native English speakers [14]. Each speaker recorded 10 phonetically rich sentences. Of the 160 samples used, 130 were recordings of different sentences. Both databases comprised of 16-bit single channel files sampled at 16 kHz, and were downsampled to 8 kHz prior to computation of the SRMR metric.

Figures 1 (a) and (b) depict the periodograms of the envelopes of the first acoustic subband from sentences uttered by a male and a female speaker, respectively. As can be seen, the envelope has a first peak (after 0 Hz) which coincides with the speakers pitch. Over the abovementioned dataset, we found that the correlation between this first peak frequency and the speakers fundamental frequency was greater than 0.8 for 13 of the 23 subbands and greater than 0.6 for 18 of the 23 subbands. Since the SRMR metric analyzes modulation spectral content up to 128 Hz with modulation filter bandwidths increasing with frequency, it is expected that fundamental frequency effects will show up in such higher modulation bands.

To show the effect of speech's phonetical content on SRMR, we grouped the samples from the CV pairs database into two different categories: by vowel and by manner of articulation. Figure 2 shows the SRMR means (dark grey) as the bar heights and standard deviations, as error bars, for each group. There is a large difference in the means for different vowels, especially between /u/ and the other two vowels. Nasals and affricates also tend to have lower SRMR scores than fricatives and stops. The intra-group relative standard deviation (RSD%, absolute standard deviation over mean in percentage) is between 66 and 100%.



(a)

(b)

Fig. 1. Periodograms of the envelopes of the first acoustic subband from sentences uttered by (a) a male and (b) a female speaker.

2.3. Proposed updates: Development of SRMR$_{\textbf{norm}}$

In order to reduce the effects of pitch and speech content on the SRMR metric, we propose two updates. First, to avoid the effect of pitch, we experimented reducing the modulation frequency ranges used by SRMR. We kept the center frequency of the first filter in the modulation filterbank at 4 Hz and varied the frequency of the last filter, starting at 30 Hz and going up to 100 Hz in 5 Hz increments. Center frequencies for the 6 filters in between were logarithmically spaced as in the original filterbank. In our experiments, we noticed that reducing the modulation frequency range from 4–128 Hz to 4–40 Hz resulted in a reduction of the correlation between SRMR and pitch from 0.76 to -0.05.

By observing the temporal progression of the energy in the modulation frequency bands of clean CVs and sentences, we confirmed that there were intra- and interspeaker differences in the ratio between lower and higher bands, as pointed by the high RSD% values found in the two databases. In order to reduce this difference, we propose as a second update to SRMR an energy thresholding method similar to the one proposed in [12]. The objective is to truncate extremely low energies, which lead to high ratios due to the division done in SRMR, and also limit the modulation energy dynamic range. In our modulation energy limitation scheme, we first compute the energy values for each of the acoustic and modulation frequencies in all frames, and then compute the average peak value, given by:

$$\bar{E}_{peak} = \max_{j,f_b}(\frac{1}{M}\sum_{m=1}^{M} E_j(m,f_b)) \tag{1}$$

where m corresponds to the frame index, f_b to the modulation frequency band index, j to the acoustic band index, and $E_j(m, f_b)$ is the energy in the j-th acoustic band and f_b-th modulation frequency band for the m-th frame. This average peak value is then used as an upper bound for the modulation energy in each band for all frames. Finally, we set the value $\bar{E}_{peak} - 30dB$ as the modulation energy lower bound, to truncate extremely low energies. In Figure 2, the

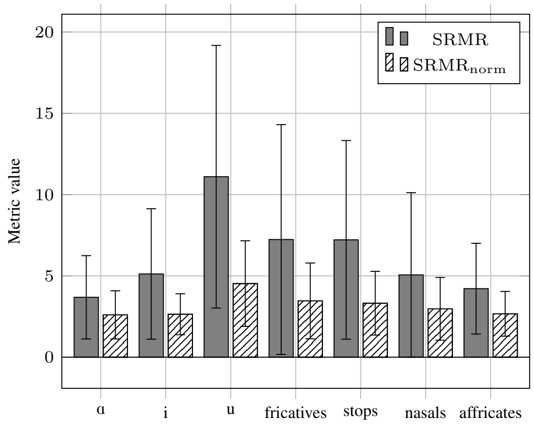


Fig. 2. Variability of the SRMR and SRMR$_{\mathrm{norm}}$ measures for the CV pairs.

hatched bars show the means and standard deviations of the SRMR metric after employing this energy limitation scheme. We can notice that the means of different CV groups are closer, and that the intra-group variability has also decreased to between 48% to 67% (against 66 to 100% in the original implementation).

In the next session, we show that the updated metric, termed SRMR$_{\mathrm{norm}}$, leads to improved speech intelligibility prediction and exhibits lower variability when compared to the original SRMR implementation.

3. EXPERIMENTAL SETUP

Here, we describe the database used to test the proposed updates, the performance measures used, as well as the benchmark algorithms.

3.1. Speech intelligibility database

In order to evaluate the effect of the proposed updates on objective speech intelligibility prediction, subjective intelligibility tests were performed with a small group of normal hearing participants. The sentence stimuli were based on the IEEE sentence corpus [15]. The corpus contains sentences with 7-12 words, organized in 72 lists of 10 sentences each. The sentences were produced by a male speaker and recorded in anechoic conditions. A total of 18 distortion conditions were used. The reverberant stimuli were generated by convolving recorded room impulse responses (RIR) obtained experimentally in three different rooms. For reverberation time (RT) values of 0.3 s, 0.6 s and 0.8 s, RIRs were obtained on a rectangular reverberant room (length 10.06 m, width 6.65 m, height 3.4 m) which had its reverberation characteristics varied by hanging absorptive panels on the walls [16]. For RT = 1.0 s, a RIR obtained on a 5.5 m × 4.5 m × 3.1 m room was used [17]. Finally, the RIR used for RT = 1.4 s was recorded in a 40 m^3 chamber (Tyndall Bruce Monument stimuli from the Open Acoustic Impulse response library [18]). Speech-shaped noise (SSN) and babble noise (from the NOISEX-92 database [19]) were added to the anechoic signals at −5 dB, 0 dB, 5 dB and 10 dB SNR levels to generate the noisy conditions. Additionally, SSN at 5 dB and 10 dB was added also to reverberant signals to generate the reverberation + noise stimuli. For the reverberation + noise condition, the reference signal used for the SNR computation was the reverberant signal, as in [20].

Ten participants, aged 19 to 39 years (average 24.7 years, standard deviation 5.83 years), were recruited. They were either native or fluent English speakers, had no history of hearing loss or hearing disorders, and had never been exposed to the sentences in the database. All participants were paid for their participation. Tests were performed in a quiet room. Sentences were presented via headphones (Sennheiser HD 600, connected to a Focusrite Saffire Pro 14 sound card) to a single ear, chosen at random. Volume was set to a comfortable level by the participant during a 10-sentence training period which comprised material different from testing.

Sentence lists were presented at a random order, and sentence order within each list was also randomized. Participants were asked to listen to the sentences, one at a time, and type in the words they were able to understand. Each sentence could only be listened to once; once presented, the participant was asked by the software to type in the words he understood (not necessarily in the same order as he heard), and could then press a key to start the next sentence. Participants were allowed to take a 10 minute break after listening to the first 9 sentence lists. Intelligibility was measured as the percentage of correct words typed for each sentence list. Simple misspellings (cases where there was no ambiguity between the typed word and another word) were corrected by hand before computing intelligibility ratings.

3.2. Performance assessment

The objective intelligibility estimates were compared to the per-condition averages of the intelligibility scores found during subjective listening tests. Averages were computed for all 20 sentences for each participant in each condition, and then averaged over all participants. A total of 18 conditions were considered: one clean, 8 noise-only (4 SNR levels, 2 noise types), 5 reverberation-only, and 4 noise-plus-reverberation. As performance criteria, we considered Pearson's linear correlation (ρ_p), Spearman rank correlation (ρ_{sp}), and Pearson's correlation after a sigmoidal mapping between predicted and true scores (ρ_{sig}, motivated by Plomp's work [21]). Additionally, based on this sigmoidal mapping, we also compute the root mean square error (RMSE) between predicted and true scores for all the tested conditions.

3.3. Benchmark metrics

As benchmark metrics, we used 9 different speech quality and intelligibility metrics, five of them being intrusive and four non-intrusive. A complete description of the benchmark algorithms is beyond the scope of this paper and the interested readers are referred to [22] and to the references given hereafter for more details. The Perceptual Evaluation of Speech Quality (PESQ) [23] and Perceptual Objective Listening Quality Assessment (POLQA) [24] are both ITU-T standards for intrusive speech quality measurement. oPESQ [1], in turn, is an adaptation of PESQ for use with reverberant speech. NCM [25] and STOI [26] are intrusive speech intelligibility metrics. ModA [27] is a non-intrusive speech intelligibility metric originally developed for cochlear implant users, and is also based on the modulation spectrum of a speech signal. ANIQUE+ [28], an ANSI standard, is a non-intrusive speech quality measure whose internal model also considers modulation spectrum features. Finally, P.563 [29] is a non-intrusive objective speech quality metric for narrow-band telephony.

Table 1. Performance results for the SRMR-based and benchmark metrics.

Metric	ρ_p	ρ_{sp}	ρ_{sig}	RSD%	RMSE
SRMR-based metrics					
SRMR	0.68	0.86	0.78	0.22	15.45
SRMR$_{\text{norm}}$	0.77	0.93	0.92	0.09	9.48
Benchmark metrics					
POLQA	0.68	0.94	0.94	0.09	7.81
NCM	0.57	0.72	0.53	0.15	22.98
CSII	0.51	0.71	0.46	0.26	23.80
STOI	0.44	0.77	0.36	0.08	23.24
PESQ	0.64	0.90	0.92	0.08	10.05
oPESQ	0.89	0.88	0.92	0.09	10.12
ANIQUE+	0.81	0.88	0.91	0.32	11.68
ModA	0.81	0.86	0.86	0.15	15.95
P.563	0.38	0.33	0.34	0.24	28.14

4. EXPERIMENTAL RESULTS

The performance results for SRMR, SRMR$_{\text{norm}}$, and the benchmark metrics are shown in Table 1. As expected from previous experiments, the original SRMR showed high variability, especially in cases with high intelligibility (which are the cases with lower variability between listeners). This is probably related to the speech content variability as shown in our CV experiments.

The value reported for SRMR$_{\text{norm}}$ corresponds to the modulation frequency range 4-40 Hz (which is the one that yielded best results). The updated metric shows a significant improvement when compared to SRMR, increasing all the correlations while decreasing the observed variability. Results are in line with POLQA, a state-of-the-art intrusive metric, even if we compare per-condition variability. The non-intrusive metric ANIQUE+, on the other hand, showed higher variability than both.

NCM, CSII, and STOI, which are specifically designed as speech intelligibility metrics, showed poorer performance than other intrusive metrics, even though they have been shown to perform well with noisy speech. As expected, oPESQ showed an improvement in linear correlation when compared to the standard PESQ measure. P.563 showed the poorest performance of all metrics. Figures 3 (a) and (b) depict the subjective versus objective scatter plots of the SRMR and SRMR$_{\text{norm}}$ outputs, respectively, for each of the noise-only, reverberation-only, and noise-plus-reverberation conditions, where error bars in the x-axis are the per-condition standard deviations, and y-axis error bars (shown only in Figure 3 (a)) the subjective standard deviations. Values were normalized between 0 and 1, where 1 corresponds to the maximum value of the metric obtained in the experiments, corresponding to the clean case. As can be seen, the variability of the SRMR$_{\text{norm}}$ outputs are significantly lower than that of SRMR.

5. CONCLUSIONS

In this paper, we proposed two updates to the SRMR metric aiming to reduce its variability related to pitch and speech content: a modulation energy thresholding scheme was employed to reduce speech content variability, while the modulation filterbank bandwidth was reduced to mitigate the effect of pitch. We assessed its performance by comparing its predictions to subjective speech intelligibility scores under noisy and reverberant conditions. The updated metric, termed SRMR$_{\text{norm}}$, was shown to have performance in line with state-of-the-art speech quality and intelligibility metrics, including intrusive ones, with correlations as high as 0.92.

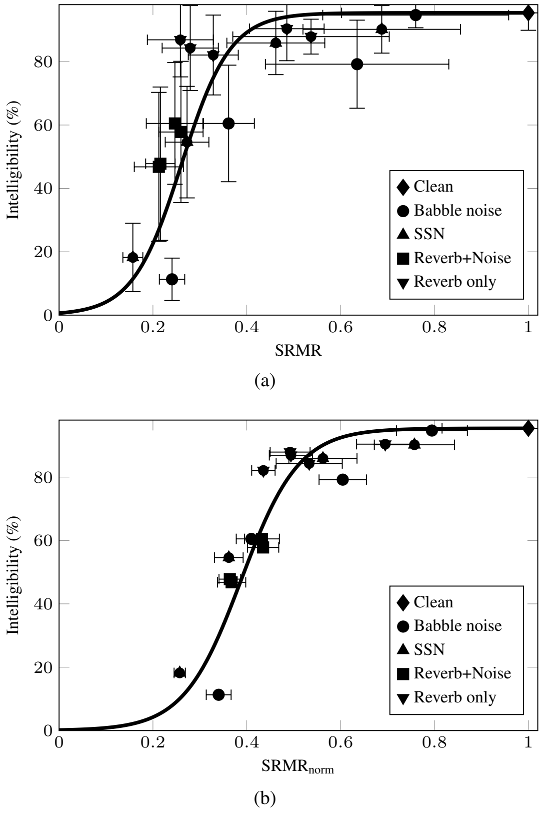


Fig. 3. Scatterplots for SRMR (a) and SRMR$_{\text{norm}}$ (b).

6. ACKNOWLEDGEMENTS

The authors would like to thank the Natural Sciences and Engineering Research Council of Canada for the financial support, Dr. Oldooz Hazrati for sharing part of the speech database used in the listening tests and the ModA MATLAB code, and Prof. Philipos Loizou (in memoriam) for his expert advice during earlier stages of this work.

7. REFERENCES

[1] K. Kokkinakis and P. C Loizou, “Evaluation of objective measures for quality assessment of reverberant speech,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2011, p. 2420–2423, IEEE.

[2] S. Cosentino, T. Marquardt, D. McAlpine, and Tiago H. Falk, “Towards objective measures of speech intelligibility for cochlear implant users in reverberant environments,” in *Proc. Intl Conf Information Science, Signal Process and Applications*, Montreal, Canada, 2012, pp. 4710–4713.

[3] João Felipe Santos, Stefano Cosentino, Oldooz Hazrati, Philipos C. Loizou, and T. H. Falk, “Performance comparison of

intrusive objective speech intelligibility and quality metrics for cochlear implant users," in *Proceedings of InterSpeech 2012*, Portland, Oregon, USA, 2012.

[4] Tiago H. Falk, Chenxi Zheng, and Wai-Yip Chan, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1766–1774, Sept. 2010.

[5] Tiago H Falk and Wai-Yip Chan, "Temporal dynamics for blind measurement of room acoustical parameters," *Instrumentation and Measurement, IEEE Transactions on*, vol. 59, no. 4, pp. 978–989, 2010.

[6] Nikolay D Gaubitch, Heinrich W Loellmann, Marco Jeub, Tiago H Falk, Patrick A Naylor, Peter Vary, and Mike Brookes, "Performance comparison of algorithms for blind reverberation time estimation from speech," in *Acoustic Signal Enhancement; Proceedings of IWAENC 2012; International Workshop on*. VDE, 2012, pp. 1–4.

[7] Jens Schröder, Thomas Rohdenburg, Volker Hohmann, and Stephan D Ewert, "Classification of reverberant acoustic situations," in *Proceedings of the International Conference on Acoustics NAG/DAGA*, 2009, pp. 606–609.

[8] T. Arai, M. Pavel, H. Hermansky, and C. Avendano, "Intelligibility of speech with filtered time trajectories of spectral envelopes," in *Fourth International Conference on Spoken Language (ICSLP)*, 1996, vol. 4, pp. 2490 –2493 vol.4.

[9] Brian R Glasberg and Brian C.J Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, vol. 47, no. 1–2, pp. 103–138, 1990.

[10] S. D. Ewert and T. Dau, "Characterizing frequency selectivity for envelope fluctuations," *The Journal of the Acoustical Society of America*, vol. 108, pp. 1181, 2000.

[11] Holger Quast, Olaf Schreiner, and Manfred R Schroeder, "Robust pitch tracking in the car environment," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*. IEEE, 2002, vol. 1, pp. I–353.

[12] Brian ED Kingsbury, Nelson Morgan, and Steven Greenberg, "Robust speech recognition using the modulation spectrogram," *Speech communication*, vol. 25, no. 1, pp. 117–132, 1998.

[13] UCLA Speech Processing and Auditory Perception Laboratory, "Consonant vowel tokens CV database," .

[14] John S Garofolo, "Timit: acoustic-phonetic continuous speech corpus," 1993.

[15] E H Rothauser, W D Chapman, N. Guttman, H R Silbiger, M H L Hecker, G E Urbanek, K S Nordby, and M. Weinstock, "IEEE recommended practice for speech quality measurements," *IEEE Transactions on Audio and Electroacoustics*, vol. 17, no. 3, pp. 225–246, 1969.

[16] Arlene C. Neuman, Marcin Wroblewski, Joshua Hajicek, and Adrienne Rubinstein, "Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults," *Ear and Hearing*, vol. 31, no. 3, pp. 336–344, June 2010.

[17] Tim Van Den Bogaert, Simon Doclo, Jan Wouters, and Marc Moonen, "Speech enhancement with multichannel wiener filter techniques in multimicrophone binaural hearing aids.," *Journal of the Acoustical Society of America*, vol. 125, no. 1, pp. 360–371, 2009.

[18] Damian T. Murphy and Simon Shelley, "OpenAIR: An Interactive Auralization Web Resource and Database," in *Audio Engineering Society Convention 129*, 11 2010.

[19] Andrew Varga and Herman JM Steeneken, "Assessment for automatic speech recognition: Ii. noisex-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, no. 3, pp. 247–251, 1993.

[20] Oldooz Hazrati and Philipos C Loizou, "The combined effects of reverberation and noise on speech intelligibility by cochlear implant listeners," *International journal of audiology*, vol. 51, no. 6, pp. 437–443, Feb. 2012.

[21] R. Plomp, "A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired," *Journal of Speech and Hearing Research*, vol. 29, no. 2, pp. 146, 1986.

[22] S. Moller, W. Y. Chan, N. Cote, T. H. Falk, A. Raake, and M. Waltermann, "Speech quality estimation: Models and trends," *Signal Processing Magazine, IEEE*, vol. 28, no. 6, pp. 18–28, 2011.

[23] ITU-T P.862, "Perceptual evaluation of speech quality: An objective method for end-to-end speech quality assessment of narrow-band telephone network and speech coders," Tech. Rep., ITU Telecommunication Standardization Sector (ITU-T), 2001.

[24] ITU-T P. 863, "Perceptual Objective Listening Quality Assessment (POLQA)," Tech. Rep., ITU Telecommunication Standardization Sector (ITU-T), 2011.

[25] J. Ma, Y. Hu, and P. C. Loizou, "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions," *Journal of the Acoustical Society of America*, vol. 125, no. 5, pp. 3387–3405, 2009.

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

[27] Fei Chen, Oldooz Hazrati, and Philipos 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, May 2013.

[28] Doh-Suk Kim and Ahmed Tarraf, "ANIQUE+: a new american national standard for non-intrusive estimation of narrowband speech quality," *Bell Labs Technical Journal*, vol. 12, no. 1, pp. 221–236, 2007.

[29] ITU-T P. 563, "Single-ended method for objective speech quality assessment in narrow-band telephony applications," Tech. Rep., ITU Telecommunication Standardization Sector (ITU-T), 2004.