NON-INTRUSIVE OBJECTIVE SPEECH QUALITY AND INTELLIGIBILITY PREDICTION FOR HEARING INSTRUMENTS IN COMPLEX LISTENING ENVIRONMENTS

Tiago H. Falk[1], *Stefano Cosentino*[2], *João Santos*[1], *David Suelzle*[3], *and Vijay Parsa*[3]

[1]INRS-EMT, University of Quebec, Montreal, QC, Canada
[2]Ear Institute, University College London, London, UK
[3]University of Western Ontario, Electrical and Computer Eng., London, ON, Canada

ABSTRACT

A non-intrusive objective speech quality and intelligibility measure tailored to hearing restoration instruments is proposed and evaluated in complex listening environments. The measure builds upon the previously-proposed "speech-to-reverberation modulation energy ratio" (SRMR) by incorporating hearing impairment percepts, such as hearing loss thresholds and altered modulation frequency selectivity. Performance is assessed using speech data corrupted by additive noise, reverberation, and noise-plus-reverberation which were subjectively rated by cochlear implant and hearing aid users. Experimental results show that the developed measures outperform the original SRMR metric for hearing impaired listeners and achieve performance levels inline with existing intrusive quality and intelligibility metrics, but with the advantage of not requiring access to a clean reference signal. As such, the measure may be used to develop quality- or intelligibility-aware speech enhancement algorithms for advanced hearing restoration instruments.

***Index Terms*—** Cochlear implant devices, hearing aids, quality measurement, intelligibility prediction, reverberation

1. INTRODUCTION

According to 2005 estimates from the World Health Organization, 278 million people worldwide had moderate to profound hearing loss in one or both ears. Depending on the degree of hearing impairment, these subjects can become candidates for hearing aid (HA) or cochlear implant (CI) devices. Recently, a number of factors, such as aging population, enlargement of candidacy criteria, and technological advances have drawn great attention to HA and CI research and development. Ultimately, users of these hearing restoration instruments are interested in obtaining improved quality and intelligibility, particularly in noisy and reverberant environments. As such, current research has focused on the development of speech enhancement techniques (e.g., noise suppression, echo cancellation) to meet this demand [1, 2, 3, 4]. To assure that the developed algorithms are behaving as expected, quality and intelligibility monitoring has to be performed.

Traditionally, subjective tests have been used to assure that acceptable levels of speech quality and intelligibility are attained. For CI devices, two approaches are commonly taken. The first makes use of vocoded speech to simulate CI hearing and presents vocoded speech to normal hearing (NH) listeners for identification (e.g., [5]). The second approach is more direct and presents degraded (or enhanced) speech stimuli directly to hearing impaired (HI) CI users for analysis (e.g., [6]). For HA users, this latter approach has been commonly used to investigate the effects of various HA signal processing techniques, such as noise suppression and feedback cancellation on the perceived speech quality [3, 4].

Subjective testing, however, is laborious, time-consuming, and expensive. Speech enhancement algorithm developers, on the other hand, require a solution that is fast and inexpensive, such that different algorithmic parameters can be optimized throughout the development stage to improve speech quality/intelligibility. Automated, repeatable, fast, and cost-effective quality/intelligibility monitoring can only be obtained with objective metrics, which replace the listeners with an auditory-inspired computational algorithm. Objective metrics can be further classified as intrusive or non-intrusive depending on the need for a clean reference signal or not, respectively. While significant effort has been placed in developing objective measures for telephone bandwidth speech with NH listeners [7], little effort has been made to date to develop objective tools targeted towards CI/HA users. In this paper, one such non-intrusive tool is proposed which incorporates hearing loss and hearing instrument precepts directly into the metric. Experiments with noisy and reverberant speech data show that the performance of the proposed non-intrusive measure is inline with that obtained with existing state-of-the-art intrusive measures, but with the added benefit of not requiring a clean reference signal, which is often unavailable in practical everyday situations.

The remainder of this paper is organized as follows. Section II will describe previous objective metrics proposed for HI listeners. Section III describes the proposed algorithm, Sections IV and V present the experimental setup and results, respectively, and Section VI concludes the paper.

978-1-4799-0356-6/13/$31.00 ©2013 IEEE

2. PREVIOUS WORK

As mentioned previously, limited work has been done to develop objective speech quality and intelligibility metrics optimized for hearing restoration instruments. For HA devices, a handful of intrusive objective quality metrics have been developed, such as the Hearing Aid Speech Quality Index (HASQI) [8, 9] and the so-called HA "auditory distance" parameter [10]. The HASQI uses a sensorineural hearing loss model to derive a set of perceptually-relevant features from HA-processed signal and its clean reference counterpart. HASQI has been validated under various *simulated* HA conditions [8]. The auditory-distance parameter, in turn, models the HA response using a time-varying ARMA system and the distance between the HA response and the model output is used to quantify the amount of distortion in the hearing aid. The model was validated with HI listeners, but not under complex listening situations with speech enhancement. Alternately, existing intrusive standardized algorithms originally developed for NH listeners (e.g., ITU-T PESQ) have also been modified to account for impaired listening, but the obtained results have not been satisfactory [11]. In terms of intelligibility, variants of conventional measures such as the articulation index have been used (e.g., [12]). To the best of the authors' knowledge, non-intrusive signal-based quality or intelligibility metrics do not exist for hearing aid users.

For CI devices, objective speech *quality* models have yet to be developed and recent studies with existing intrusive measures (e.g., PESQ) have resulted in poor correlations with subjective scores, both with normal hearing listeners using vocoded speech to simulate CI hearing and with noisy speech data presented directly to CI users [13, 14, 15]. On the other hand, *intelligibility* prediction for CI users in complex listening scenarios is a more mature research area and tools such as the normalized covariance metric (NCM) and the coherence-based speech intelligibility index (CSII) have shown to be fairly reliable indicators of intelligibility [14, 15]. As such, emphasis is placed here on intelligibility prediction for CI users. Interestingly, a recently-proposed *non-intrusive* speech quality and intelligibility metric termed SRMR (speech to reverberation modulation energy ratio) [16] showed promising results across noisy and reverberant conditions [14]. The measure was originally developed for NH listeners and is based on an auditory-inspired modulation spectral signal representation. The objective of the present study is to improve on the existing SRMR measure by incorporating hearing loss and instrument insights into the modulation spectral auditory model; this is performed for both CI and HA hearing models.

3. PROPOSED METRICS

3.1. Original SRMR Implementation

The SRMR non-intrusive metric was originally developed for reverberant and dereverberated speech and evaluated against subjective NH listener data [16]. Here, only a brief description of the measure is given and the interested reader is referred to [16] for complete details. The block diagram depicted in Figure 1 (within the dashed lines) shows the signal processing steps involved in the computation of the original SRMR metric. First, the input speech signal is filtered by a 23-channel 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 [17]. Temporal envelopes are then computed via the Hilbert transform for each of the 23 filterbank outputs and used to extract modulation spectral energy for each critical band via a discrete Fourier transform (using 256 ms frames, 32ms shifts). In order to emulate frequency selectivity in the modulation domain [18], modulation frequency bins are grouped into eight overlapping modulation bands with centre frequencies logarithmically spaced between $4-128$ Hz. Lastly, the SRMR value is computed as the ratio of the average modulation energy content available in the first four modulation bands ($3-20$ Hz, consistent with clean speech modulation content [19]) to the average modulation energy content available in the last four modulation bands ($20-160$ Hz), which were previously shown to convey room acoustics information [20].

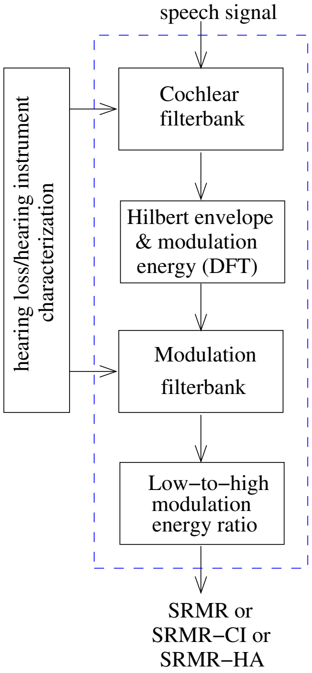
speech signal
hearing loss/hearing instrument characterization
Cochlear filterbank
Hilbert envelope & modulation energy (DFT)
Modulation filterbank
Low−to−high modulation energy ratio
SRMR or SRMR−CI or SRMR−HA

Fig. 1. Signal processing steps in the calculation of the original (within dashed lines) and proposed objective metrics.

3.2. SRMR Tailored to Hearing Instruments

In order to tailor the SRMR measure for CI and HA instruments, a few modifications were implemented, as depicted by Fig. 1. For CI devices, the 23-channel gammatone filterbank was replaced by the 22-channel filterbank (with mel-like spacing) available in the Nucleus research device which was used by the listeners in the subjective test. Moreover, the $4-128$ Hz range of the eight modulation filterbank centre frequencies of the original SRMR metric were reduced to $4-64$ Hz, following recent insights described in [13, 14]. The SRMR metric tailored to CI devices is henceforth refereed to as SRMR-CI. Similarly, to simulate HA hearing, the 23-channel gammatone filterbank was modified to take into account the listener's individual hearing loss thresholds obtained via an audiogram. More specifically, the Q-factor of each of the 23 filters were adjusted to simulate the hearing loss due to outer hair cell damage. In summary, as hearing loss increased, so did the filter bandwidths (i.e., Q-factors decreased). Since there is no previous literature on modulation domain frequency-selectivity for hearing aid users, the full $4-128$ Hz range of modulation filterbank centre frequencies was kept. The SRMR metric tailored to HA devices is henceforth refereed to as SRMR-HA.

4. EXPERIMENTAL SETUP

Two separate datasets were used in our experiments. Both comprised of anechoic speech data corrupted by recorded room impulse responses with varying reverberation times and by additive noise. One dataset was developed at the University of Texas at Dallas (USA) and presented to CI users in a controlled listening test, whereas the other was collected at the University of Western Ontario (Canada) and presented to HA users. A brief summary of the two datasets is given below; the interested reader is referred to [6, 21] for complete details on the subjective listening test conditions.

4.1. Database 1: CI Speech Intelligibility

The speech sentences presented to the CI users were taken from the well-known IEEE sentence corpus. Four recorded room impulse responses were convolved with the clean speech data to simulate reverberant speech with reverberation times (RT60) of 0.3, 0.6, 0.8, and 1 s. Speech-shaped noise was also added to the anechoic and the reverberant signals to generate noise-only and noise-plus-reverberation degradation conditions, respectively. Noise was added at a signal-to-noise-ratio (SNR) of -5, 0, 5 and 10 dB for the anechoic samples and 5 and 10 dB for the reverberant samples. For the noise-plus-reverberation condition, the reverberant signals served as reference for SNR computation.

Eleven adult CI users were recruited to participate in the subjective *intelligibility* experiments. The participants were all native speakers of American English with post-lingual deafness and had an average age of 64 years. All participants had a minimum of one year experience using their device routinely, with some being bilaterally implanted for over 6 years. For consistency, all participants were temporarily fitted with a SPEAR3 research processor with parameters matching the individual CI user's clinical settings. Participants were presented with 20 sentences randomly selected from the IEEE database, each corrupted by the above mentioned degradation conditions. Degraded stimuli were presented directly to the audio input of the research processor and the level was adjusted individually for comfort at the beginning of the experiment. Listeners were instructed to repeat all identifiable words and per-participant intelligibility scores were calculated as the ratio of the number of correctly identified words to the total number of presented words.

4.2. Database 2: HA Speech Quality

The speech material presented to the HA users consisted of the well-known HINT (hearing in noise test) sentences presented to participants in a double-walled sound booth (RT60=0.1 s) and in a reverberant chamber (RT60=0.9 s). Both clean speech data and data corrupted by additive noise (multi-talker babble and speech-shaped noise at 0 and 5 dB SNR) were presented to listeners via four loudspeakers placed within the room's critical distances at $0°, 90°, 180°$, and $270°$ azimuths, thus simulating noise-only, reverberation-only, and noise-plus-reverberation listening conditions.

Twenty-two adult HA users (average age of 71 years) were recruited to participate in the subjective *quality* experiments. Each of the participants were fitted bilaterally with the Unitron experimental behind-the-ear HA. Participants listened to the corrupted speech files four times, each time with a different HA setting, namely: omnidirectional microphone, adaptive directional microphone, partial strength speech enhancement enabled (directionality and noise reduction algorithms operating below their maximum strengths), and full strength speech enhancement (all enhancement algorithms operating at maximum strength). Subjects rated their perceived quality for each stimulus using the well-known $[0-100]$ MUSHRA quality scale, with 0 referring to poor quality and 100 to excellent. Lastly, the above mentioned data collection protocol was repeated but with a head and torso dummy equipped with the experimental HA programmed to the individual listener's hearing loss profile. This allowed for the actual enhanced signals heard by the listeners to be used by the benchmark and proposed SRMR-HA algorithms. Since data was presented binaurally, the left and right channel data were processed by the quality prediction algorithms and the average was used for performance comparisons.

4.3. Benchmark Algorithms

In order to gauge the benefits of the proposed non-intrusive measures, existing intrusive parameters were used as benchmarks. For HA conditions, the recently-proposed and validated HASQI speech *quality* prediction parameter was used.

The interested reader is referred to [8, 9] for its complete description. For CI conditions, in turn, two intrusive metrics were used, namely the normalized covariance metric (NCM) and the coherence-based speech intelligibility index (CSII). These parameters have been shown to be reliable *intelligibility* indicators for CI hearing in noise and reverberation [14, 15, 13]. The NCM parameter estimates speech intelligibility based on the covariance between the envelopes of the clean and degraded speech signals, whereas the CSII parameter estimates speech intelligibility based on the spectral coherence between the two signals (e.g., [22, 23, 24]).

4.4. Performance Metrics

In order to assess the performance of the tested algorithms, two performance metrics were used, namely Pearson correlation (R) and the standard deviation of the prediction error (ϵ). As suggested in the literature, performance values are reported on a per-condition basis, where condition-averaged objective performance ratings and condition-averaged subjective intelligibility/quality ratings are used in order to reduce intra- and inter-subject variability [7]. In the CI experiment, thirteen conditions were available: clean, four noise-only conditions (-5 to 10 dB SNR with 5 dB increments), four reverberation-only conditions (RT60 = 0.3, 0.6, 0.8, 1.0 s), and four noise-plus-reverberation conditions (RT60 = 0.6 and 0.8 s with SNR of 5 and 10 dB). In the HA experiment, 40 conditions were available: 2 RT60 levels × (4 HA settings × 2 noise types × 2 SNRs + 4 HA settings in quiet).

5. RESULTS

Table 1 presents the performance metrics (R, ϵ) obtained by the proposed and benchmark algorithms for the two datasets. As can be seen, the promising results obtained with the original SRMR measure provide further evidence of the importance of temporal envelope cues for speech intelligibility prediction in cochlear implants. Notwithstanding, further improvements were obtained once CI precepts were incorporated into the metric. More specifically, gains of 3.2% and 11.7% were obtained in R and ϵ, respectively with the SRMR-CI parameter relative to the original SRMR. Overall, the proposed SRMR-CI metric performed inline with the existing intrusive metrics, but with the benefit of not requiring access to a clean reference. Further investigation into the SRMR-CI estimates showed that the performance remained stable across the noise-only, reverberation-only, and noise-plus-reverberation conditions, thus signalling its usability over a wide range of complex listening conditions.

For the HA database, in turn, the proposed SRMR-HA measure achieved somewhat lower performance than the benchmark HASQI algorithm. This drop in performance may be compensated by the fact that the proposed metric does not require access to a clean reference signal, thus is better suited for real-time applications. Further investigation into the SRMR-HA estimates showed that the performance also remained stable across different conditions. Moreover, it can also be observed that by incorporating hearing loss information into the metric, significant improvements can be obtained relative to the original SRMR measure; more specifically, gains of 15% and 20% were obtained in R and ϵ, respectively. Our ongoing work focuses on investigating the effects of unwanted HA speech enhancement artefacts on the modulation spectrum; these insights may lead to further improvements in speech quality measurement performance.

Table 1. Overall per-condition performance metrics of proposed and benchmark algorithms on the CI (Dataset 1) and HA (Dataset 2) datasets

	Database 1		Database 2	
Metric	R	ϵ	R	ϵ
NCM	0.96	12.4	–	–
CSII	0.93	10.6	–	–
HASQI	–	–	0.95	5.4
SRMR	0.93	12.8	0.73	11.5
SRMR-CI	0.96	11.3	–	–
SRMR-HA	–	–	0.84	9.2

6. CONCLUSIONS

This paper proposed a new non-intrusive speech quality and intelligibility metric tailored towards hearing instruments for complex listening environments. More specifically, measures were developed for cochlear implant and hearing aid devices. The so-called SRMR-CI and SRMR-HA metrics, respectively, were shown to accurately estimate speech quality/intelligibility across noise-only, reverberation-only and noise-plus-reverberation listening conditions. The obtained performance results were inline with those obtained with state-of-the-art intrusive measures, but with the added benefit of not requiring access to a clean reference signal. Ultimately, access to a reliable non-intrusive speech intelligibility/quality metric may open doors to intelligibility- and/or quality-aware speech enhancement algorithms to improve speech-in-noise recognition for hearing instrument users.

7. ACKNOWLEDGEMENTS

The authors would like to thank the Natural Sciences and Engineering Research Council of Canada and the Ontario Research Fund for the financial support, Dr. Oldooz Hazrati for sharing the CI database and Prof. Philipos Loizou (in memoriam) for his expert advise during earlier stages of this work.

8. REFERENCES

[1] K. Kokkinakis, O. Hazrati, and P. C. Loizou, "A channel-selection criterion for supressing reverberation in cochlear implants," *Journal of the Acoustical Society of America*, vol. 129, no. 5, pp. 3221–3232, 2011.

[2] L. P. Yang and Q. J. Fu, "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise," *The Journal of the Acoustical Society of America*, vol. 117, pp. 1001, 2005.

[3] I. Merks, S. Banerjee, and T. Trine, "Assessing the effectiveness of feedback cancellers in hearing aids," *Hearing Review*, vol. 13, no. 4, pp. 53–57, 2006.

[4] T. Rohdenburg, S. Goetze, V. Hohmann, K.D. Kammeyer, and B. Kollmeier, "Combined source tracking and noise reduction for application in hearing aids," in *ITG Conf Voice Communications*, 2008, pp. 1–4.

[5] M. Qin and A. Oxenham, "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," *Journal Acoustical Society of America*, vol. 114, no. 1, pp. 446–454, 2003.

[6] O. Hazrati and P. C. Loizou, "The combined effects of reverberation and noise on speech intelligibility by cochlear implant listeners," *International Journal of audiology*, Feb. 2012, PMID: 22356300.

[7] 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.

[8] J.M. Kates and K.H. Arehart, "The hearing-aid speech quality index (HASQI)," *Journal of the AES*, vol. 58, no. 5, pp. 363–381, May 2010.

[9] A.A. Kressner, D.V. Anderson, and C.J. Rozell, "Robustness of the hearing aid speech quality index (HASQI)," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2011, pp. 209–212.

[10] V. Parsa and D.G. Jamieson, "Hearing aid distortion measurement using the auditory distance parameter," in *111th AES Convention*, 2001.

[11] J. Beerends, K. Eneman, R. Huber, J. Krebber, and H. Luts, "Speech quality measurement for the hearing impaired on the basis of PESQ," in *124th AES Convention*, 2008.

[12] C.V. Pavlovic, G.A. Studebaker, and R.L. Sherbecoe, "An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals," *The Journal of the Acoustical Society of America*, vol. 80, pp. 50–57, 1986.

[13] Fei Chen and Philipos C. Loizou, "Predicting the intelligibility of vocoded speech," *Ear and Hearing*, vol. 32, no. 3, pp. 331–338, 2011.

[14] S. Cosentino, T. Marquardt, D. McAlpine, and TH Falk, "Towards objective measures of speech intelligibility for cochlear implant users in reverberant environments," in *Intl Conf Information Science, Signal Process and Applications*, 2012, pp. 4710–4713.

[15] J. Santos, S. Cosentino, O. Hazrati, P. C. Loizou, and T. H. Falk, "Performance comparison of intrusive objective speech intelligibility and quality metrics for cochlear implant users," in *InterSpeech*, 2012.

[16] 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 Lang Process*, vol. 18, no. 7, pp. 1766–1774, Sept. 2010.

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

[18] 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.

[19] T. Arai, M. Pavel, H. Hermansky, and C. Avendano, "Intelligibility of speech with filtered time trajectories of spectral envelopes," in *Intl Conf Spoken Language*, 1996, vol. 4, pp. 2490 –2493.

[20] T.H. Falk and W.Y. Chan, "Temporal dynamics for blind measurement of room acoustical parameters," *IEEE Trans Instr Meas*, vol. 59, no. 4, pp. 978–989, 2010.

[21] A. Keymanesh, P. Folkeard, S. Scollie, V. Parsa, D. Hayes, L. Cornelisse, and P. Allen, "Perceptual evaluation of hearing aid digital signal processing strategies across noisy and reverberant environments," *Intl Journal of Audiology*, 2012.

[22] I. Holube and B. Kollmeier, "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model," *Journal Acoustical Society America*, vol. 100, no. 3, pp. 1703–1716, 1996.

[23] J.M. Kates and K.H. Arehart, "A model of speech intelligibility and quality in hearing aids," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. 2005, pp. 53–56, IEEE.

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