

TOWARDS OBJECTIVE MEASURES OF SPEECH INTELLIGIBILITY FOR COCHLEAR IMPLANT USERS IN REVERBERANT ENVIRONMENTS

Stefano Cosentino¹, Torsten Marquardt¹, David McAlpine¹ and Tiago H. Falk²

¹ Ear Institute, University College London (UCL), London, UK

² Institut National de la Recherche Scientifique, INRS-EMT, Montréal, Canada

ABSTRACT

This study validates a novel approach to predict speech intelligibility for Cochlear Implant users (CIs) in reverberant environments. More specifically, we explore the use of existing objective quality and intelligibility metrics, applied directly to vocoded speech degraded by room reverberation, here assessed at ten different reverberation time (RT_{60}) values: 0 s, 0.4 s – 1.0 s (0.1 s increments), 1.5 s and 2 s. Eight objective speech intelligibility predictors (SIPs) were investigated in this study. Of these, two were non-intrusive (i.e. did not require a reference signal) audio quality measures, four were intrusive, and two were intrusive speech intelligibility indexes. Three types of vocoders were implemented to examine how speech intelligibility predictions depended on the vocoder type. These were: noise-excited vocoder, tone-excited vocoder and a FFT-based N-of-M vocoder. Experimental results show that several intrusive quality and intelligibility measures were highly correlated with exponentially fit CI intelligibility data. On the other hand, only a recently - developed non-intrusive measure showed high correlations. These evaluations suggest that CI intelligibility may be accurately assessed via objective metrics applied to vocoded speech, thus may reduce the need for expensive and time-consuming listening tests.

Keywords: Vocoders, Reverberation, Cochlear Implants, Objective Measures, Speech Intelligibility;

1. INTRODUCTION

Reverberation produces temporal envelope smearing, low spectral contrast and flattening of formant transition as a result of self-masking effects. These signal alterations, especially in the signal envelope, have a dramatic effect on the speech intelligibility of cochlear implant (CI) user, as it has been shown both via simulations with vocoders on normal hearing (NH) listeners [1] [2] and via intelligibility tests on CI users [3]. Several dereverberation algorithms have been proposed and evaluated via one or both of the methods mentioned above. Subjective testing, however, is very time consuming, costly and often hindered by the high inter- and intra-subject variability. Vocoders are software which can simulate CI hearing, hence have the advantage of not requiring CI subjects; these tests are still quite time consuming. A third option is potentially available: the use of objective measures applied to the

vocoded signal. In [5], and more recently in [4], Chen and Loizou studied different objective measures as predictors for speech intelligibility. In [5] they compared quality scores produced by a standardised speech quality measurement algorithm termed PESQ [10] with the intelligibility scores of 20 NH subjects in vocoded speech at different Signal-to-Noise Ratio (SNR) scenarios (-5, 0 and 5dB); correlation values between 0.92 and 0.94 were obtained. In the subsequent work ([4]) they investigated the suitability of additional objective measures (e.g. NCM, CSII) together with a more systematic study on the effect that some parameters of the vocoder have on the overall correlation. Their investigation did not consider reverberation.

In a recent publication Kokkinakis *et al.* [3] showed that speech intelligibility in CI drops exponentially as RT_{60} linearly increases. The intelligibility scores measured in CI were exponentially fit as:

$$score(\%) = e^{(C_1 \cdot RT_{60} + C_2)}, \quad (1)$$

where $C_1 = 0.0014$, $C_2 = 4.528$ and the RT_{60} is a measure of the amount of reverberation in a room (here expressed in *ms*). Such fitting resulted in correlation with subjective CI scores as high as 0.996. The present study tests an objective approach to predict speech intelligibility in CI from vocoded speech when reverberation is the only form of noise. In order to do so, several speech quality and intelligibility measures are used as speech intelligibility predictors (SIPs), and their performances are compared in terms of Pearson's correlation and Spearman's correlation between the predicted value and the exponential fit of CI data, as provided in eq.(1). These measures are estimated after the signal is passed through a vocoder which simulates the CI listening. Three types of vocoders were used in this study to investigate the impact of the vocoder specifics on the reliability of the objective measure. Tone-excited and noise-excited vocoders were implemented with 6, 12 and 24 channels. A third type of vocoder was an FFT-analyzer 6-of-12 vocoder.

The remainder of this paper is organized as follows: Section 2 describes the objective metrics which we used in our experiments. The experimental setup is presented in Section 3, whereas results and discussion are reported in the Section 4. Conclusions are provided in Section 5.

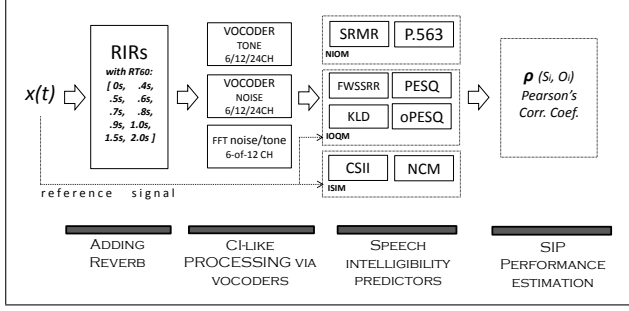


Fig. 1. Experimental steps used to assess the performance of eight objective quality and intelligibility metrics on CI data.

2. OBJECTIVE QUALITY AND INTELLIGIBILITY METRICS

This section describes the eight SIPs used in the study. These are divided into three groups: non-intrusive objective quality measures (NIOM), intrusive objective quality measures (IOQM) and intrusive speech intelligibility measures (ISIM). While quality metrics have not been developed for the purpose of intelligibility prediction, in many instances they can serve both purposes, as in [4] or [5]. For a full overview of the methods please see the diagram in Fig.1.

2.1. Non-Intrusive Objective Quality Measures (NIOM)

Two NIOMs were investigated: the Speech to Reverberation Modulation Energy Ratio (SRMR) and the ITU-P563 standard algorithm. The non-intrusiveness refers to the fact that these measures produce a quality prediction without the need for a reference signal.

2.1.1. ITU P563

The P563 is the only non-intrusive measure so far tested and approved by the ITU-T, which recommended it for several test factors, coding technologies and applications. It works by taking into account the effects of both transmission level distortions and signal related distortions (e.g. unnaturalness of speech, interruptions, robotic voice). As will be recalled in the conclusion of this paper, the P563 measure is not recommended for synthesised speech, although in our simulation bitrate requirements were met. For more information the interested reader is referred to [6].

2.1.2. SRMR

Falk *et al.* [7] proposed a non-intrusive objective measure of reverberation in speech based on estimation of spectral modulation energy shift, across frequency, as a result of late and early reflections. This measure - termed Speech to Reverberation Modulation Energy Ratio (SRMR) - was shown to outperform several standard measures designed for perceived speech intelligibility, coloration and reverb

estimation. The SRMR is estimated as:

$$SRMR = \frac{\sum_{m=1}^4 \sum_{k=1}^{23} |F\{e_k(m, n)\}|^2}{\sum_{m=5}^M \sum_{k=1}^{23} |F\{e_k(m, n)\}|^2} \quad (2)$$

where e_k is the envelope of the filtered signal in critical band k , F refers to the Fourier transform, and m is the index of the M total modulation frequency bands. A detailed description of the method is beyond the scope of this paper and the interested reader is referred to [7][8] for more details on the measure.

2.2. Intrusive Objective Quality Measures (IOQM)

Four intrusive objective quality measures were implemented: the Perceptual Evaluation of Speech Quality (PESQ), an optimised PESQ for reverberation (oPESQ), the Kullback-Leibler divergence (KLD) and the Frequency-Weighted Segmental Speech-to-Reverberation Ratio (FWSSRR). All the intrusive measures described in this and the following section require a reference signal. Given that we made the hypothesis of considering the vocoded speech as representative of CI hearing, we chose the vocoded dry signal ($RT_{60} = 0$) as reference. Nonetheless, it has been shown that using the clean unprocessed speech as reference leads to same patterns in the results [4].

2.2.1. PESQ

The PESQ measure is probably the most reliable and employed objective predictor of speech sound quality [10]. It is recommended by ITU-T P.862 for speech quality assessment of narrow-band handset telephony and narrow-band speech codecs. The quality prediction is based on a perceptual model which output is calculated as a linear combination of two factors: the disturbance value (D_{ind}) and the average asymmetrical disturbance values (A_{ind}). Both factors are estimated by comparing the clean and the processed signal and are integrated to form a final quality rating according to:

$$PESQ = a_0 + a_1 \cdot D_{ind} + a_2 \cdot A_{ind}, \quad (3)$$

$$\text{where } \begin{cases} a_0 = +4.5 \\ a_1 = -0.1 \\ a_2 = -0.0309 \end{cases} \quad (4)$$

2.2.2. oPESQ

The PESQ measure has been developed as a predictor of speech sound quality when in presence of noise, but not reverberation. As such, in [11] three different variations of the PESQ measure were presented which were optimised to correlate with reverberation perception of normal hearing in the form of speech coloration, reverberation tail effect and overall speech quality. In all three cases the optimised measure was obtained via multiple linear regression analysis to determine optimal a_0 , a_1 and a_2 parameters in eq.3. In our study we implemented the optimised PESQ for overall speech quality as a potential SIP, naming it oPESQ. The oPESQ is hence obtained from eq. (3)

by using the coefficients as in [11]:

$$\begin{cases} a_0 = +4.6876 \\ a_1 = -0.5678 \\ a_2 = +0.1024 \end{cases} \quad (5)$$

2.2.3. KLD

First proposed in [12], the KLD measure was then successfully implemented as a quality measure for reverberation in [11]. The KLD estimates the distance between the probability distribution functions (*pdf*) of two signals; in our case, the *pdf* of the clean and reverberant vocoded speech. As an effect of the spectral and temporal smearing produced by the reverberation, the *pdf* of the reverberant speech (p_R) will always be more flat compared with the *pdf* of the clean vocoded speech, p_C . Hence, the KLD measure is a non-negative measure which tends to zero when the distributions become similar (it is zero if $p_C = p_R$) and it is estimated as:

$$KLD = - \int p_C(t) \cdot \log_{10} \frac{p_C(t)}{p_R(t)} dt \quad (6)$$

where the integration is over the time variable t .

2.2.4. FWSSRR

The FWSSRR measure is obtained by estimating the signal-to-noise ratio for each time frame and for each critical band. These values are then weighted according to frequency weights published in (ANSI 1997 [9]) and averaged along the whole time length of the signal. In our study the FWSSRR was computed as:

FWSSRR =

$$\frac{10}{N} \sum_{n=1}^N \frac{\sum_{k=1}^K W(k) \cdot \log_{10} \frac{|C(n,k)|^2}{|C(n,k) - R(n,k)|^2}}{\sum_{k=1}^K W(n,k)} \quad (7)$$

where $C(n, k)$ and $R(n, k)$ are respectively the clean and reverberant vocoded speech signal in the time frame n , and for the channel frequency k ; $K = 25$ is the number of critical bands, N is the total number of time frames, $W(k)$ is the weighting function as derived for the articulation index (AI), and published in the standard ANSI 1997 [9].

2.3. Intrusive Speech Intelligibility Measures (ISIM)

The Normalised Covariance Metric (NCM) and the Coherence Speech Intelligibility Index (CSII) were also used. The reason for choosing these two measures is that they have been specifically developed to predict speech intelligibility, although they were fitted to NH listeners and in situations involving noise rather than reverberation. For both the NCM and the CSII measure we have used the vocoded dry speech as reference signal.

2.3.1. NCM

The NCM was implemented as in [13] and [4], where it was shown to correlate well with intelligibility scores for vocoded speech. The NCM is a Speech Transmission Index (STI, [14]) related measure, where the essential difference from the STI being the fact that the NCM uses the covariance of the envelope between clean and processed signal, whereas STI measures use the differences in their modulation transfer functions. For the NCM derivation the envelopes of clean and reverberant vocoded signals are first extracted via Hilbert transform for each of the 25 channels, after filterbank analysis. The normalised correlation coefficients between the respective envelopes produce a local SNR which is then limited in the range $[-15, 15]$ dB, and linearly mapped in the range $[0, 1]$. These values are then weighted in each channel according to the articulation index (AI) weights published in the ANSI (1997, [9]), and the average is taken as the final NCM value. The formula to estimate the NCM is the following:

$$NCM = \frac{10}{N} \sum_{n=1}^N \frac{\sum_{k=1}^{K=25} W(f_k) \cdot [\log_{10} \frac{r_{ch}^2}{1-r_{ch}^2}]_{[0,1]}}{\sum_{k=1}^{K=25} W(n, f_k)} \quad (8)$$

where r_{ch} is the correlation coefficient between the clean and reverberant envelopes estimated in each channel; the $[\cdot]_{[0,1]}$ operator refers to process of limiting and mapping into $[0, 1]$ range. For more detailed information about how this measure is calculated please refer to [4] or [15].

2.3.2. CSII

As opposed to the NCM, which is a temporal-based measure, the CSII is a spectral-based speech intelligibility measure [15]. It is calculated by multiplying coherence-based weights to the processed speech in the frequency domain. In order to do so, the signal is divided into N windowed segments (30ms Hanning window with 75% overlap) and its Fourier transform is computed. Each time-frequency segment is weighted by the Magnitude Squared Coherence (MSC) between the clean and reverberant signals estimated across the entire signal length. The mathematical derivation of the CSII is as follows:

$$CSII = \frac{10}{N} \sum_{n=1}^N [\log_{10} \cdot \frac{\sum_{k=1}^{K=25} G(f_k) \cdot MSC(f_k) \cdot |R(n, f_k)|^2}{\sum_{k=1}^{K=25} G(f_k) \cdot (1 - MSC(f_k)) \cdot |R(n, f_k)|^2}]_{[0,1]} \quad (9)$$

where:

$$MSC(f_k) = \frac{|\sum_{n=1}^N C(n, f_k) R^*(n, f_k)|^2}{\sum_{n=1}^N |C(n, f_k)|^2 \sum_{n=1}^N |R(n, f_k)|^2} \quad (10)$$

and $G(f_k)$ is the frequency response of the f_k^{th} critical pass-band filter (each filter has center frequency f_k).

$C(n, f_k)$ and $R(n, f_k)$ are the FFT spectra of clean and reverberant signal, respectively, estimated in the time frame n and channel k . As for the NCM, the $[\cdot]_{[0,1]}$ operator refers to process of limiting the argument in the range $[-15, 15]$ dB and successive mapping in the range $[0, 1]$.

Table 1. Vocoder Filterbank Specifics
CF = center frequency (Hz); BW = bandwidth(Hz)

6	CF	180	446	885	1609	2803	4773
Ch	BW	201	331	546	901	1487	2453
12	CF	124	224	353	519	731	1005
Ch		1355	1806	2385	3128	4084	5310
	BW	88	113	145	186	239	307
		395	507	651	836	1074	1379
24	CF	101	145	194	251	315	387
Ch		469	562	668	787	923	1077
		1251	1448	1671	1925	2212	2538
		2906	3324	3798	4335	4944	5634
	BW	41	47	53	60	68	77
		87	99	112	127	144	163
		185	210	238	269	305	346
		392	444	503	571	647	733

3. EXPERIMENTAL SETUP

3.1. Data Preparation

A subset of speech sentences from the IEEE sentence list [16] was convolved with room impulse responses (RIRs) to obtain reverberated signals. The RIRs were artificially created by using a RIR generator based on the image-method [17]. This software allows a controlled manipulation of the reverberation condition by changing the RT_{60} value. In our study we set the room dimension to be the same as in [3] (4.5m x 5.5m x 3.10m) and we investigated 9 RT_{60} conditions (0.4s, 0.5s, 0.6s, 0.7s, 0.8s, 0.9, 1.0s, 1.5s and 2.0s) plus the dry condition ($RT_{60} = 0$). Thirty sentences were used for each RT_{60} condition. The reverberant speech was then input to the three vocoders.

3.2. Vocoders Implementation

Vocoders have been widely used to simulate CI hearing. Many studies have reported high correlations between the intelligibility scores of CI users and NH with vocoded speech when presented with the same speech sentences (e.g. [18], [19]). In our study three different types of vocoders were chosen to investigate their impact on the ability of the different measures to predict speech intelligibility. The first type of vocoder is the tone vocoder (TV) as implemented in [18]. The signal processing in this vocoder consists of a first band-pass filtering via a 6th order Butterworth *Ch*-filterbank (analysis filterbank), with *Ch* being the number of channels. The filterbank spans the frequency range [80 – 6000] Hz, and each filter has bandwidth as measured on an ERB scale. Following, the envelope is extracted in each channel via half-wave rectification and a low-pass filtering (2nd order Butterworth, 300Hz cut-off frequency or half the analysis bandwidth, whichever the smallest). The envelope is then used to modulate a sinusoidal carrier with frequency equal to the

centre frequency of the pass-band filter for that channel. The outputs from each channel are then summed to produce the re-synthesised signal.

A second type of vocoder is the noise vocoder (NV). It includes the same processing steps as the TV, except that narrow-band noise (white noise filtered with the same analysis filterbank) is used as carrier instead of sinusoids. For both vocoders three different channel values were tested: 6, 12 and 24 channels; this produced 6 conditions labelled as TV6, TV12, TV24, NV6, NV12 and NV24 (see Table 1 for both centre frequencies and channel bandwidths).

Lastly, a third type of vocoder implemented is a FFT-based vocoder with N-of-M channel selection. This vocoder follows similar processing stages as implemented in the Digital Speech Processor (DSP) of the Neurelec Digisonic SP (for more information see the implementation in [20]). The N-of-M is a peak selection criterion routinely implemented in most CI, but rarely modeled in vocoders; nonetheless, it has been shown that energy based N-of-M channel selections can be detrimental for CI in reverberant conditions [3]. This last vocoder was implemented as noise and tone vocoder, with a 6-of-12 channel selection (labels: TV6/12 and NV6/12).

4. EXPERIMENTAL RESULTS

4.1. Performance Metric

The performance of each SIP is evaluated in a *per-condition* basis (i.e., all scores under the same RT_{60} condition are averaged) using the Pearson's correlation coefficient (ρ). The measures estimate the sample correlation between the speech intelligibility scores predicted by the objective measure, and the expected speech intelligibility scores in CI as measured in [3] (see eq.(1) of this document). This performance metric is given by:

$$\rho = \frac{\sum(S_i - \mu_{S_i}) \cdot \sum(\hat{O}_i - \mu_{O_i})}{\sqrt{\sum(S_i - \mu_{S_i})^2} \cdot \sqrt{\sum(\hat{O}_i - \mu_{O_i})^2}} \quad (11)$$

where O_i is a 10-value array containing the speech intelligibility values predicted by the objective measure i per each RT_{60} , and S_i is the array containing the speech intelligibility values expected from (1); μ_{O_i} and μ_{S_i} are the respective means.

4.2. Results and Discussion

Table 2 reports the Pearson's correlation values averaged per single vocoder condition (30 sentences per condition; 10 RT_{60} values in each condition per each measure per each vocoder type), together with a visual matrix. These results are indicative of how suitable to predict CI data each measure is when coupled to a specific vocoder type or number of channels.

From the results it can be observed that the ITU P563 performed the worst, with an average absolute correlation of 0.667; in addition, correlations were instable across different vocoder conditions. This result is consistent with what

Table 2. Per-Vocoder Pearson’s (ρ) correlations.

TV=tone-excited vocoder; NV = noise-excited vocoder.

Vocoder:	TV	TV	TV	NV	NV	NV	TV	NV	Correlation Matrix
Ch:	6	12	24	6	12	24	6/12	6/12	ρ
P563	0.697	-0.474	-0.658	0.859	0.910	0.884	-0.547	-0.320	
SRMR	0.988	0.990	0.995	0.989	0.992	0.986	0.996	0.998	
PESQ	0.946	0.939	0.954	0.884	0.891	0.899	0.927	0.879	
oPESQ	0.983	0.991	0.988	0.987	0.987	0.991	0.999	0.991	
FWSSRR	0.992	0.991	0.995	0.927	0.935	0.950	0.975	0.942	
KLD	0.909	0.914	0.917	0.856	0.861	0.864	0.884	0.848	
NCM	0.999	0.999	0.999	0.964	0.965	0.964	0.997	0.977	
CSII	0.987	0.989	0.995	0.836	0.843	0.851	0.962	0.894	

was found in [21], where the P563 was shown to perform poorly with synthesized speech. The poor performance of the ITU P563 is also shown in Fig.2, where the \hat{O}_i values of three SIPs and the expected CI scores are reported. For this plot each O_i array has been scaled in the range $[0, S_i(0)]$, where $S_i(0)$ is the value in eq. (1) for the anechoic condition, which is 92.57%.

Except the P563 measure, all the other SIPs showed to predict well the degradation introduced by reverberation, with an average correlation coefficient of 0.94. The two intrusive speech intelligibility indices (NCM and CSII) performed very well for tone vocoders (both 0.99 on average), while showing somewhat poorer performance with noise vocoders. Same trend is observed with the intrusive quality objective measures KLD, FWSSRR and PESQ. The reason for this is probably due to the fact that noise vocoders introduce their own noise-envelope: these random fluctuations reduce the similarity between the reference and the processed signal.

In contrast, oPESQ and SRMR seem to be the most robust measures, showing high correlation across all vocoder con-

ditions. The SRMR (average $\rho=0.992$) is a particularly interesting metric, as it does not depend on a reference signal, thus can be used for real-time quality and intelligibility monitoring applications (e.g., in a quality-aware enhancement algorithm). The good performance of the SRMR can be explained by the fact that it estimates the reverberation content only by using the envelope of the filter-bank output, regardless of the value obtained for the quiet condition, and discards time information (i.e. the fine structure contained in the original signal is not used). This processing is very similar to what is performed in a CI. The NCM metric works in a similar manner (but with a reference signal) and also results in reliable scores across the different conditions. Moreover, the oPESQ measure performed better than the PESQ in conditions involving reverberation (mean correlation values of 0.915 and 0.989 respectively), thus suggesting that an optimized mapping is indeed needed to estimate the intelligibility of reverberant speech for CI users.

With respect to the vocoder specifics, the N-of-M processing had negligible impact on the SIPs performance, whereas the tone/noise carrier selection had the largest impact on all but SRMR and oPESQ measures. Within the noise/tone vocoder groups, however, the number of channel did not show noticeable effect on the SIPs’ outcome. Since this is a preliminary study, we made two assumptions: the first assumption is about the existence of a measure of speech intelligibility that, as being objective, would be independent of subject-specific cognitive factors; therefore we should be careful in considering these measures as estimators of absolute performance. In fact, they are rather predictors of relative performance degradation across different RT_{60} values.

The second assumption is on the generalization of our results to reverberant environments. The estimated SIP performances may be dependent on how we define reverberation. In our study and in the study where CI data was collected ([3]), the RT_{60} was the only factor defining the reverberation. However, other factors such as the Signal to Reverb Ratio (SRR) or the exact room dimensions might also affect the speech intelligibility, thus further study is required to determine the impact of additional parameters.

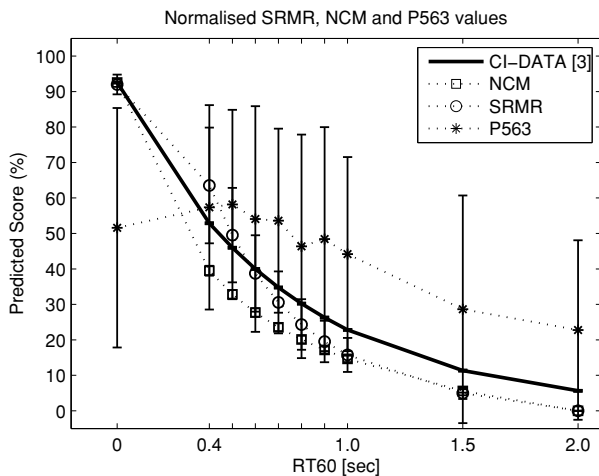


Fig. 2. Comparison between the predicted outputs of three different measures (\hat{O}_i , with $i = \text{SRMR, NCM or P563}$) and the expected values in CI (continuous line). Results are obtained by scaling and averaging all TV conditions.

5. CONCLUSION

This study has tested an objective procedure for predicting speech intelligibility for cochlear implant users (CI) in reverberant environments and evaluated several objective measures for this task. Amongst all, two measures stood out, namely oPESQ and SRMR, which showed stable and high correlation scores regardless of the condition tested. It was found that while vocoder type played a role in the performance of the tested metrics, the number of vocoder channels had little effect on the majority of the measures (except P.563). Further study is needed to assess the robustness of these results across other intelligibility-impairing conditions (e.g., room dimensions, speaker - listener distance, reverberation-plus-noise).

6. ACKNOWLEDGEMENTS

SC is funded by a UCL impact PhD award supported by Neurelec. THF is funded by the Natural Sciences and Engineering Research Council of Canada.

7. REFERENCES

- [1] Poissant, S.F., Whitmal, N.A., and Freyman, R.L., "Effects of reverberation and masking on speech intelligibility in cochlear implant simulations", J. Acoust. Soc. A., 2006. 119(3): p. 1606-1615.
- [2] Drgas, S., and Blaszkak, M.A., "Perception of speech in reverberant conditions using AM-FM cochlear implant simulation", Hearing Research, 2010. 269(1-2): p. 162-168.
- [3] Kokkinakis, K., Hazrati, O., and Loizou, P., "A channel-selection criterion for suppressing reverberation in cochlear implants", J. Acoust. Soc. A., 2011. 129(5): p. 3221-3232.
- [4] Chen, F., and Loizou, P., "Predicting the Intelligibility of Vcoded Speech", Ear & Hearing, 2011. 32(3): p. 331-338.
- [5] Chen, F., and Loizou, P., "Contribution of consonant landmarks to speech recognition in simulated acoustic-electric hearing", Ear & Hearing, 2010. 31(2): p. 259-267.
- [6] Malfait, L., Berger, J., and Kastner, M., "P.563 - The ITU-T Standard for Single-Ended Speech Quality Assessment", IEEE Trans. on Audio, Speech, and Language Processing, 2006. 14(6): p. 1924-1934.
- [7] Falk, T.H., Chenxi, Z. and Wai-Yip, C., "A Non-Intrusive Quality and Intelligibility Measure of Reverberant and Dereverberated Speech", IEEE Trans. on Audio, Speech, and Language Processing, 2010. 18(7): p. 1766-1774.
- [8] Falk, T.H. and Wai-Yip, C., "Temporal Dynamics for Blind Measurement of Room Acoustical Parameters", IEEE Trans. on Instrumentation and Measurement, 2010. 59(4): p. 978-989.
- [9] ANSI, "Methods for Calculation of the Speech Intelligibility Index", ANSI - American National Standards Institute, New York. 1997, S3(5).
- [10] ITU-T P.862, "Perceptual evaluation of speech quality: An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech coders", 2001, ITU-T.
- [11] Kokkinakis, K. and Loizou, P., "Evaluation of Objective Measures for Quality Assessment of Reverberant Speech", 2011. ICASSP, IEEE.
- [12] Kullback, S. and Leibler, R.A., "On Information and Sufficiency. The Annals of Mathematical Statistics", 1951. 22(1): p. 79-86.
- [13] Holube, I., and Kollmeier, B., "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model", J. Acoust. Soc. A., 1996. 100(3): p. 1703-1716.
- [14] Steeneken, H.J.M., and Houtgast, T., "A physical method for measuring speech-transmission quality", J. Acoust. Soc. A., 1980. 67(1): p. 318-326.
- [15] Ma, J., Hu, Y., and Loizou, P., "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions", J. Acoust. Soc. A., 2009. 125(5): p. 3387-3405.
- [16] IEEE, "IEEE recommended practice speech quality measurements", IEEE Trans. on Audio Electroacoust., 1969. AU17: p. 225246.
- [17] Allen, J.B., and Berkley, D.A., "Image method for efficiently simulating small-room acoustics", J. Acoust. Soc. A., 1976. 60(S1): p. S9.
- [18] Qin, M.K., and Oxenham, A.J., "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers", J. Acoust. Soc. A., 2003. 114(1): p. 446-454.
- [19] Litvak, L.M., Spahr, A.J., Saoji, A.A., Fridman, G.Y., "Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners", J. Acoust. Soc. Am., 2007. 122(2): p. 982-991.
- [20] Kallel, F., Frikha, M., Ghorbel, M., Hamida, A.B., Berger-Vachon, C., "Dual-channel spectral subtraction algorithms based speech enhancement dedicated to a bilateral cochlear implant", Applied Acoustics, 2011.
- [21] Moller, S., Kim, D.S., Malfait, L., "Estimating the Quality of Synthesized and Natural Speech Transmitted Through Telephone Networks Using Single-ended Prediction Models", Acta Acustica united with Acustica, 2008. 94(1): p. 21-31.