

Predicting the Bilateral Advantage in Cochlear Implantees using a Non-Intrusive Speech Intelligibility Measure

Stefano Cosentino¹, Tiago H. Falk² and David McAlpine¹

¹Ear Institute, University College London, United Kingdom

²INRS-EMT, University of Québec, Montréal, Canada

Abstract

A measure to predict speech intelligibility in unilateral and bilateral cochlear implant (CI) users is proposed that does not need *a priori* information (i.e. is *non-intrusive*), such as the room acoustics. Such measure, termed *BiSIM_{CI}*, combines an equalization-cancellation stage together with a modulation frequency estimation stage. Simulated and actual subjective data from CI users were used to validate the proposed measure. The actual CI subjective data consisted of speech reception thresholds (SRTs) collected in anechoic rooms with a total of 28 target/interferer spatial configurations. The simulated CI subjective data were generated by running the intrusive algorithm by Culling *et al.* [Ear & Hearing 33 (6), 673-682 (2012)] across 109 different target/interferer conditions from two environments, one anechoic and one highly reverberant room ($RT_{60} = 0.89s$). The experimental results indicate that the proposed non-intrusive measure provides reliable predictions when compared with both actual and simulated SRT; an average correlation of 0.94 was reported for these conditions, and an average correlation of 0.97 was obtained when some intrusive assumptions were made.

1. Introduction

Several studies have shown that bilateral, compared to unilateral, cochlear implantation (CI) improves the identification of speech-in-noise. This advantage has recently been studied in [1], where a model (henceforth referred to as “*CAR12_{CI}*”) was described that could make predictions of the improvement in signal-to-noise ratio (SNR) in CI users for various spatial configurations. Despite achieving very high correlations with actual CI speech reception thresholds (SRTs), the *CAR12_{CI}* model has a practical limitation in that it requires information not directly available to the listener, such as the room impulse responses. Such models depend on *a priori* information, and are referred as “intrusive”. In this study we develop and evaluate a non-intrusive measure of the bilateral advantage to speech intelligibility based on cochlear implant auditory-inspired insights. Cochlear implantees are believed to receive three main benefits when implanted with a second device (i.e. bilateral cochlear implant, BCI). The first bilateral benefit is the better ear effect (BE), which refers to the ability of the BCI user to attend the signal from the device with a better SNR. This is the largest component in the total bilateral advantage (BA), and ranges from 3 to 10 dB [1-10]. A second advantage is the binaural unmasking (BU), or binaural interaction, which results from subtracting the interferer components from the sound mixture, in order to enhance the SNR. This latter component has often been reported to be very small, between 0 and 2 dB [2-5, 9, 11]. Finally, a third reported bilateral advantage is the binaural summation

(BS), which refers to the improved target identification when this is co-located with the interferer at 0°; in this specific condition, the presentation of the stimuli over two devices instead of one has been reported to produce few decibel of advantage [6, 9, 12], although this is possibly due to differences in peripheral hearing between right and left ear in the same CI user [1, 13]. To date, the *CAR12_{CI}* is the only bilateral model reported in literature able to predict speech intelligibility in CI users, and this is achieved by using *a priori* information. A non-intrusive bilateral intelligibility measure, however, could find more practical applications, such as in bilateral intelligibility-aware speech enhancement algorithms for CI. Recently, a unilateral non-intrusive measure termed Speech to Reverberation Modulation Energy Ratio (*SRMR*) was developed that could predict the effect of reverberation on speech intelligibility for normal-hearing (NH) listeners [14]. The *SRMR* computes a ratio between the energy in the low frequency modulations, attributed mostly to spoken speech components, and the energy in the high frequency modulations, which are mostly attributed to room acoustic effects. In [15], a CI version of the *SRMR* (“*SRMR_{CI}*”) was described that could reliably predict the speech intelligibility in unilateral CI users for conditions involving noise and reverberation. In this study we propose a non-intrusive measure, referred to as “*BiSIM_{CI}*” (Blind Bilateral Speech Intelligibility Model for Cochlear Implant users), that can predict the bilateral advantage to speech intelligibility in CI users as combination of BE and BU components. The BE is estimated by using a bilateral extension of the *SRMR_{CI}* [15], whilst the BU is modelled by measuring binaural masking level differences. The performance of the developed measure is evaluated by comparing its BA predictions against subjective and simulated CI data under different listening conditions.

2. BiSIM_{CI} Description

The *BiSIM_{CI}* computes the BA from the separate estimation of the BE and BU (both expressed in dB) as:

$$BA = BE + BU \quad (1)$$

2.1. BE - Better ear effect

The BE was computed by taking the maximum of a modified version of the *SRMR_{CI}* estimated in the right and left ear channels. Some of the CI-inspired modifications to the original *SRMR* are already described in [15], and are: (i) the frequency-analysis filterbank was chosen to match the one of a typical CI device, the Nucleus Freedom by Cochlear; (ii) the frequency-modulation filterbank was limited in the range [4-64] Hz, (the original *SRMR* was using up to 128 Hz), in order to mimic the altered modulation sensitivity observed in CI

Frequency-Analysis Filterbank [0.1 · kHz]																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
f_k	2.5	3.8	5	6.3	7.5	8.8	10	11.3	12.5	14.4	16.9	19.4	21.9	25	28.8	33.1	38.1	43.8	50	56.9	65	74.4
BW	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25	2.5	2.5	2.5	2.5	3.75	3.75	5	5	6.25	6.25	7.5	8.75	10

Table 1: Centre frequencies (f_k) and bandwidths (BW) for the frequency-analysis filterbank used in the *SRMR-CI* described in [15]

Frequency-Modulation Filterbank [Hz]								
	1	2	3	4	5	6	7	8
f_c	4.0	5.9	8.8	13.1	19.5	29.0	43.1	64.0
BW	2.0	3.0	4.5	6.6	9.8	14.5	21.5	32.0

Table 2: Centre frequencies (f_c) and bandwidths (BW) for the frequency-modulation filterbank used in the *SRMR-CI* described in [15]

users [16] (centre frequencies and bandwidths of the frequency-analysis and the frequency-modulation filterbanks are reported in Table 1 and 2, respectively). Additionally, a frequency weighting was applied to the *SRMR-CI* in this study. The purpose of this weighting is to emphasise the degradation occurring in those channels that are more important for speech intelligibility. In our implementation, a weighting based on NH studies ([17]) was followed, although in the future, a frequency weighting based on CI-specific parameters could be implemented.

The final BE value is therefore estimated as:

$$BE = \max(SRMR-CI_{left}^*, SRMR-CI_{right}^*), \quad (2)$$

where

$$SRMR-CI^* = \sum_k^{22} SRMR-CI(f_k) \cdot W(f_k) \quad (3)$$

W are the frequency weights described in [17]; $SRMR-CI(f_k)$ is the *SRMR* estimated in each of the 22 channels, filtered via gammatone filterbank, as described in greater details in [15].

A limitation with this approach is that the scale of the predicted BE is not in decibel, therefore not compatible with the BU measured in eq. (1). To overcome this limitation, the BE produced by the *BiSIMCI* is linearly scaled in the range of the expected SRTs. This normalisation is a limitation to the non-intrusiveness of the model and alternate scaling solutions have been explored, as will be discussed in Section 3.3.

2.2. BU - Binaural unmasking

The BU stage was estimated via the binaural masking level difference (BMLD) between the target and the interferer in each of the 22 frequency channels. This concept originates from the Equalisation-Cancellation theory [18] which describes the spatial release from masking in binaural NH listeners, later modelled in [19, 20] as:

$$BMLD(f) = 10 \cdot \log_{10} \left(\frac{h(f) - \cos(2\pi \cdot f \cdot (\tau_T - \tau_I))}{h(f) - \rho(f)} \right) \quad (4)$$

where $\tau_T - \tau_I$ is the difference between the interaural time delays (ITDs) of the target and the interferer, respectively, and ρ is the interaural coherence of the interferer; h is a frequency-dependent constant obtained as:

$$h(f) = (1 + \delta_\epsilon^2) \cdot \exp((2\pi \cdot f)^2 \cdot \delta_\sigma^2) \quad (5)$$

where $\delta_\epsilon = 0.25$ and $\delta_\sigma = 105\mu s$ as in [18]. As can be noted from eq. (4), sensitivity to ITD is essential to perform release from masking, and thus to benefit from BU. Cochlear implant users, however, have very low ITD sensitivity [21], and consequently small BU values [2-5, 9]. In order to mimic such reduced sensitivity, uncorrelated Gaussian noise was added to the envelopes of the right and left filtered signals, while fine-structure was entirely discarded, as is common in CI processing [22]. The noise power was empirically chosen in order to produce a 1 dB BU advantage for target and interferer at 0° and 30° , respectively, as shown in [5]. For this condition, a noise power of +12 SNR produced the smallest error between the BU predictions and 1 dB, thus in the *BiSIMCI* model Gaussian noise at SNR=+12dB was added to the envelopes of the right and left filtered signals. The $\tau_T - \tau_I$ and the ρ were estimated by using a blind direction-of-arrival (DOA) model developed by Dietz *et al.* in [23]. The DOA produced a pair of coherence (ρ) and ITD (τ_I) values per time sample, which were then sorted via Gaussian fit on the ITD histogram into separate interferers. For multiple interferers, multiple BMLDs were estimated and the smallest in each channel was selected, reflecting the limited ability of CI users to perform spatial release from masking when multiple interferers are present. Finally, the same frequency weighting described in Sec. 2.1 was used to integrate the BMLD across channels and obtain the BU as:

$$BU = \sum_k^{22} BMLD(f_k) \cdot W(f_k) \quad (6)$$

An assumption we introduced is that the target was always in the front, which implied $\tau_I = 0s$. This is somewhat necessary so that the algorithm knows which source is the target. This assumption will be relaxed in a BE-only version of the *BiSIMCI* described in Section 3.3.

3. Experimental Setup and Results

Two numerical experiments were designed to validate the proposed measure. In experiment I, the BA predictions from the *BiSIMCI* were compared against two databases of SRTs measured in actual unilateral and bilateral CI users for spatially distributed target and interferers. In a second experiment, simulated subjective data was created by using the *CAR12CI* model described in [1]. Binaural room impulse responses (BRIRs) were employed to simulate target/interferer spatial locations in a specific environment. The main features of the BRIRs used in this study are reported in Table 3.

3.1. Numerical Experiment I

3.1.1. Loizou *et al.* 2009 [5]

Loizou *et al.* (2009, [5]) collected SRTs in six bilateral CI users for eight different spatial configurations, each measured for three listening conditions: only the right implant on, only the left implant on, or with both devices on. The total number of conditions was therefore 24. The target sentences were from

Room Type:	A	B	C	D
RT60 [s]	0	0.32	0	0.89
DRR [dB]	∞	6.09	∞	6.12
No. of Conditions	24	4	72	37
Reference	[24]	[24]	[25]	[24]

Table 3: Parameters of the BRIRs used in this study throughout numerical experiments I and II. RT_{60} , reverberation time; DRR, direct-to-reverberation ratio.

the IEEE database [26] and were always at 0° , while the interferer (one or three speech modulated noise signals) was located in one or more of the following possible locations: 0° , -30° , 30° , 60° , 90° (negative angles refer to interferers located to the left of the subject). Since the SRT data was collected in anechoic room, BRIRs from an anechoic environment ([24], room A in Table 3), which are essentially head related transfer functions, were used to produce the stimuli used as input for the $BiSIM_{CI}$. The same type of target/interferer material used in [5] was also used in our simulation. The BA predicted from the $BiSIM_{CI}$ in each condition was averaged over 20 repetitions. For displaying purposes, the BA predictions from the $BiSIM_{CI}$ were converted into predicted SRTs by inverting the sign and subtracting the mean of the observed SRTs. The predicted and observed SRTs obtained a correlation coefficient of 0.93, as depicted by the scatter plot in Figure 1.

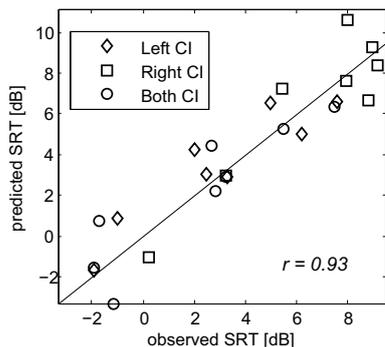


Figure 1: Comparison between observed ([5]) and predicted SRTs. Left and Right refer to unilateral listening.

3.1.2. Culling et al. 2012 [1]

Eight unilateral CI subjects were recruited in [1] to measure SRTs in four conditions. These conditions included a speech target and a single noise interferer presented in four possible arrangements [deg/deg]: $0/+90$, $0/-90$, $+60/-60$ and $-60/+60$. A fifth condition with target and interferer in 0° (i.e. $0/0$) was used as baseline to measure the BA. In order to simulate the dry room ($RT_{60} = 0.11s$) used in [1], we chose BRIRs from [24] with $RT_{60} = 0.32s$ (room B in Table 3) as the closest approximation. As in the original study, sentences from the IEEE corpus were used as speech target, and speech-shaped noise was used as noise interferer. The result from each condition was averaged over 20 repetitions. The BA predictions from the $BiSIM_{CI}$ and the actual BA data achieved a correlation coefficient of 0.97, and are plotted in Figure 2. In addition, the predictions and the data generated from normal hearing (NH) subjects

are also plotted, although a description of the $BiSIM_{NH}$ (the normal-hearing version of the $BiSIM_{CI}$, [27]) is beyond the scope of this article.

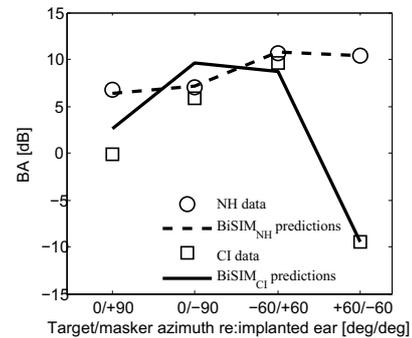


Figure 2: Comparison between observed (from [1]) and predicted BAs in cochlear implant (CI) and normal hearing (NH) subjects.

3.2. Numerical Experiment II

In order to validate further the proposed model, two additional settings were simulated: an anechoic room, simulated via BRIRs recorded for 0° elevation, and azimuths ranging from -180° to 180° in steps of 5° [25] (rooms C in Table 3); a reverberant room ($RT_{60} = 0.89s$), with BRIRs recorded for 0° elevation, and azimuths ranging from -90° to 90° in steps of 5° [24] (room D in Table 3). Both BRIR databases offered a large number of interferer azimuths that allowed the creation of spatial maps of the BA. Given the lack of suitable subjective CI data that is available in literature, real subject SRT data was replaced with BA predictions from the $CAR12_{CI}$ model developed in [1]. Such model showed high correlations with the same two databases we tested in Exp. I, ($r = 0.97$) and II ($r = 1.0$). Since the $CAR12_{CI}$ does not model the BU, only the BE as predicted by $BiSIM_{CI}$ was used. Figure 3 shows the percent BA in anechoic (panel a) and reverberant (b) environments predicted by the $BiSIM_{CI}$ and the $CAR12_{CI}$ models as a function of the interferer azimuth for target speech always at 0° . For the anechoic simulation of Figure 3a, both models predict a similar BA trend ($r = 0.98$), including the trough in $\pm 90^\circ$. This is due to the “bright spot” phenomenon, which is the sudden decrease in noise level at the contralateral ear originated by the constructive phase interference of the audio waves propagating around the head. This effect was experimentally shown in [28], and demonstrated in [29]. It can also be noted that, for this acoustic setting, the BAs for noise located in the front are larger than for interferer located in the back. These BA functions suggest that a bilateral CI user that attempts to follow a conversation with a target located straight ahead could expect a maximum advantage from the second implant when the interferer is located from about 60° and 70° away from the target. Different BA functions are obtained for the reverberant condition shown in Figure 3b, although even in this instance the $BiSIM_{CI}$ could accurately predict the expected BA values ($r = 0.99$). With respect to the anechoic condition, a more asymmetric pattern is found between the BA estimated for the right and left hemispheres. This is possibly due to a left/right-asymmetry of the BRIRs. More importantly, the bright spot at 90° is highly reduced (and indeed cancelled), due to the decor-

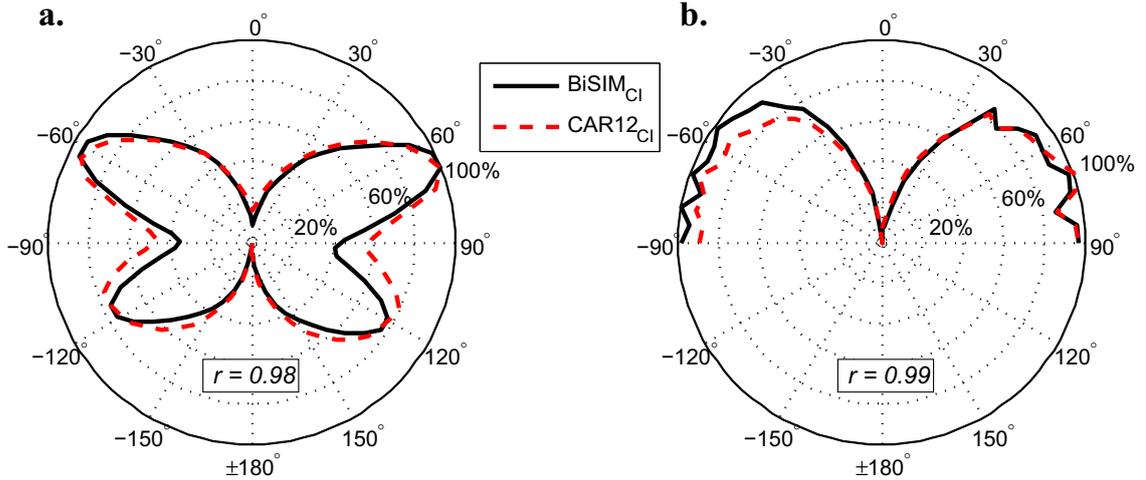


Figure 3: Percent bilateral advantage (BA) as a function of the interferer azimuth for target straight ahead. Correlation (r) is estimated between the BA predicted by the $BiSIM_{CI}$ and the $CAR12_{CI}$ models (note that $BA = BE$ for both).

a.: anechoic room ($RT_{60} = 0s$); **b.**: reverberant room ($RT_{60} = 0.89s$).

relation of the signals reaching the contralateral ear via different paths. In environments more realistic than anechoic ones, such as the reverberant room of Figure 3.b, the bright spot is only marginal, and the bilateral cochlear implant listener can be expected to obtain maximum release from masker when this is located more than 40° - 50° away from the sagittal plane.

3.3. BE-scaling effects

The first row of Table 4 reports the correlations between actual (Exp. I) or simulated (Exp. II) data from CI users with the predictions from the $BiSIM_{CI}$. As can be noted, the proposed non-intrusive model performed well across the two numerical experiments, with an average correlation of 0.97. However, as discussed in Section 2.1, the intrusive scaling of the BE component within the range of expected SRT values affects the non-intrusiveness of the algorithm, and presumably increases the correlations reported in Table 4. In order to investigate the impact of the scaling on the performance of the $BiSIM_{CI}$, we correlated the output of the BE stage before scaling directly to the SRT/BA observed values. By this approach, the BU as in eq. (1) is discarded (and so is the assumption made in Sec. 2.2 of target located straight ahead), which is a simplification often supported in the literature [4, 5], and no scaling is necessary for the BE. The correlations estimated with this second approach did not change (see second row of Table 4), except for the data from [5], reported as scatter plot in Figure 4. Future studies will have to address this aspect in order to develop a quantitative, not only qualitative, estimation of the BA on a dB scale.

	Exp. I		Exp. II		abs. mean
	Loizou et al. 09	Culling et al. 12	$CAR12_{CI}$ in Room C	$CAR12_{CI}$ in Room D	
$BiSIM_{CI}$	0.93	0.97	0.98	0.99	0.97
BE-only	-0.81	0.97	0.98	0.99	0.94

Table 4: Correlations between predicted and observed/expected values for experiments I and II.

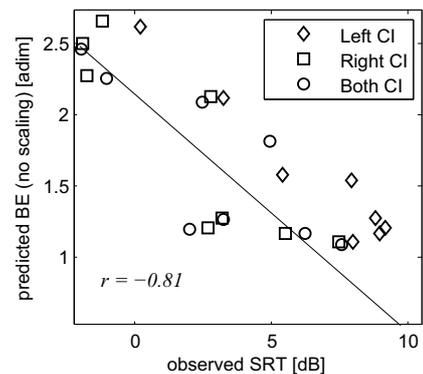


Figure 4: Scatter plot of observed SRTs (from [5]) and the not-scaled BE predictions. The negative correlation is due to the inverse relation between the BE and the SRT.

4. Conclusions

In this paper, a novel non-intrusive bilateral speech intelligibility measure was developed and tested. To the authors knowledge, this is the first measure that predicts the bilateral advantage for cochlear implant users that works without *a priori* information. This is achieved by estimating two components: the advantages accrued by the better ear effect and the binaural unmasking. The measure was tested across several environments, and with actual and simulated cochlear implant subjective measurements. An average correlation of 0.97 was obtained with expected values for a version of the measure that used some *a priori* information, and correlation of 0.94 was obtained when this *a priori* information was not used.

5. Acknowledgements

SC is funded by a PhD scholarship from UCL and Neurelec. THF acknowledges funding from NSERC Canada.

6. References

- [1] J. F. Culling *et al.*, "The Benefit of Bilateral Versus Unilateral Cochlear Implantation to Speech Intelligibility in Noise," *Ear & Hearing*, vol. 30, pp. 673-682, 2012.
- [2] J. Muller *et al.*, "Speech understanding in quiet and noise in bilateral users of the MED-EL COMBI 40/40+ cochlear implant system," *Ear & Hearing*, vol. 23, pp. 198-206, 2002.
- [3] R. S. Tyler, "Three-month results with bilateral cochlear implants," *Ear & Hearing*, vol. 23, pp. 80-89, 2002.
- [4] R. J. M. van Hoesel, "Speech perception, localization, and lateralization with bilateral cochlear implants," *The Journal of the Acoustical Society of America*, vol. 113, pp. 1617-1630, 2003.
- [5] P. C. Loizou *et al.*, "Speech recognition by bilateral cochlear implant users in a cocktail-party setting," *The Journal of the Acoustical Society of America*, vol. 125, pp. 372-383, 2009.
- [6] C. C. Dunn, "Comparison of speech recognition and localization performance in bilateral and unilateral cochlear implant users matched on duration of deafness and age at implantation," *Ear & Hearing*, vol. 29, pp. 352-359, 2008.
- [7] R. Laszig, "Benefits of bilateral electrical stimulation with the nucleus cochlear implant in adults: 6-month postoperative results," *Otology & Neurotology*, vol. 25, pp. 958-968, 2004.
- [8] R. Y. Litovsky *et al.*, "Spatial Hearing and Speech Intelligibility in Bilateral Cochlear Implant Users," *Ear & Hearing*, vol. 30, pp. 419-431, 2009.
- [9] P. Schleich *et al.*, "Head shadow, squelch, and summation effects in bilateral users of the MED-EL COMBI 40/40+ cochlear implant," *Ear & Hearing*, vol. 25, pp. 197-204, 2004.
- [10] F. Schon *et al.*, "Results of bilateral cochlear implantation," *European archives of oto-rhino-laryngology*, vol. 156, pp. 106-110, 1999.
- [11] R. Litovsky, *et al.*, "Simultaneous bilateral cochlear implantation in adults: a multicenter clinical study," *Ear & Hearing*, vol. 27, pp. 714-731, 2006.
- [12] N. V. G. Verhaert *et al.*, "Speech performance and sound localization abilities in Neurelec Digisonic SP Binaural cochlear implant users," *Otology & Neurotology*, vol. 17, pp. 256-266, 2012.
- [13] R. J. M. van Hoesel, "Exploring the Benefits of Bilateral Cochlear Implants," *Audiology and Neurotology*, vol. 9, pp. 234-246, 2004.
- [14] T. H. Falk *et al.*, "A Non-Intrusive Quality and Intelligibility Measure of Reverberant and Dereverberated Speech," *Audio, Speech, and Language Processing*, *IEEE Transactions on*, vol. 18, pp. 1766-1774, 2010.
- [15] T. H. Falk *et al.*, "Non-Intrusive Objective Speech Quality and Intelligibility Prediction for Hearing Instruments in Complex Listening Environments", *IEEE International Conference on Acoustics, Science and Signal Processing*, Vancouver, Canada, 2013, *in press*.
- [16] R. V. Shannon, "Temporal modulation transfer functions in patients with cochlear implants," *The Journal of the Acoustical Society of America*, vol. 91, pp. 2156-2164, 1992.
- [17] ANSI, "Methods for Calculation of the Speech Intelligibility Index," ANSI - American National Standards Institute, New York, vol. S3, 1997.
- [18] N. I. Durlach, "Equalization and Cancellation Theory of Binaural Masking-Level Differences," *The Journal of the Acoustical Society of America*, vol. 35, pp. 1206-1218, 1963.
- [19] J. F. Culling *et al.*, "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," *The Journal of the Acoustical Society of America*, vol. 116, pp. 1057-1065, 2004.
- [20] J. F. Culling *et al.*, "Erratum: The role head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources [J. Acoust. Soc. Am. 116, 1057 (2004)]," *The Journal of the Acoustical Society of America*, vol. 118, p. 552, 2005.
- [21] R. van Hoesel *et al.*, "Interaural Time-Delay Sensitivity in Bilateral Cochlear Implant Users: Effects of Pulse Rate, Modulation Rate, and Place of Stimulation," *The Journal of the Association for Research in Otolaryngology*, vol. 10, pp. 557-567, 2009.
- [22] B. S. Wilson and M. F. Dorman, "Cochlear implants: A remarkable past and a brilliant future," *Hearing Research*, vol. 242, pp. 3-21, 2008.
- [23] M. Dietz *et al.*, "Auditory model based direction estimation of concurrent speakers from binaural signals," *Speech Communication*, vol. 53, pp. 592-605, 2010.
- [24] C. Hummersone *et al.*, "Dynamic Precedence Effect Modeling for Source Separation in Reverberant Environments," *Audio, Speech, and Language Processing*, *IEEE Transactions on*, vol. 18, pp. 1867-1871, 2010.
- [25] B. Gardner and K. Martin, "HRFT Measurements of a KEMAR Dummy-head Microphone," *MIT Media Lab Perceptual Computing Technical Report*, 1994.
- [26] IEEE, "IEEE recommended practice speech quality measurements," *Transactions on Audio Electroacoustic*, vol. AU17, pp. 225-246, 1969.
- [27] S. Cosentino *et al.*, "A Blind Binaural Speech Intelligibility Prediction Model," *36th MidWinter Meeting of the Association for Research in Otolaryngology*, Baltimore, *abstract*, 2013.
- [28] J. Peissig and B. Kollmeier, "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," *The Journal of the Acoustical Society of America*, vol. 101, pp. 1660-1670, 1997.
- [29] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *The Journal of the Acoustical Society of America*, vol. 104, pp. 3048-3058, 1998.