INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Effects of Cochlear Hearing Loss on the Benefits of Ideal Binary Masking

Vahid Montazeri, Shaikat Hossain, Peter F. Assmann

University of Texas at Dallas, Richardson, TX, USA
{vahid.montazeri, shaikat, assmann}@utdallas.edu

Abstract

Ideal Binary Masking (IdBM) is considered as the primary goal of computational auditory scene analysis. This binary masking criterion provides a time-frequency representation of noisy speech and retains regions where the speech dominates the noise while discarding regions where the noise is dominant. Several studies have shown the benefits of IdBM for normal hearing and hearing-impaired listeners as well as cochlear implant recipients. In this study, we evaluate the effects of simulated moderate and severe hearing loss on the masking release resulting from IdBM. Speech-shaped noise was added to IEEE sentences; the stimuli were processed using a tone-vocoder with 32 bandpass filters. The bandwidths of the filters were adjusted to account for impaired frequency selectivity observed in individuals with moderate and severe hearing loss. Following envelope extraction, the IdBM processing was then applied to the envelopes. The processed stimuli were presented to nineteen normal hearing listeners and their intelligibility scores were measured. Statistical analysis indicated that participants' benefit from IdBM was significantly reduced with impaired frequency selectivity (spectral smearing). Results show that the masking release obtained from IdBM is highly dependent on the listeners' hearing loss.

Index Terms: speech perception, computational auditory scene analysis, ideal binary masking, speech enhancement, hearing loss.

1. Introduction

Listeners with normal hearing are reasonably successful at understanding speech in the presence of one or more competing voices. The ability to segregate the target voice from the masker has been modeled by two distinct stages. First, the auditory periphery decomposes the input mixture into individual time-frequency (T-F) bins. The acoustic energy in each T-F bin is detected, provided that the signal-to-noise ratio (SNR) in that bin is above a certain level. In the second stage, based on *a priori* information about the target and other cues, the listener examines all the T–F units in the mixture, segregates the T–F units of the target, and integrates them into a single auditory image of the target signal. Cues such as periodicity, voice offsets and onsets, and amplitude /frequency modulations, are believed to be utilized by human listeners [1].

Several computational auditory scene analysis (CASA) techniques have been proposed in the literature implementing the above two-stage process [2]. The ideal binary masking (IdBM) technique, proposed by Wang [3], is believed to be the primary goal of CASA. With a two-dimensional time-frequency (T-F) representation of the mixture (target degraded with masker), an ideal binary masker is defined as a binary criterion within which a value of 1 denotes that the target energy in the corresponding unit exceeds the masker by a predefined threshold and is set to 0 otherwise. The threshold is called the local SNR criterion (LC), measured in decibels. More specifically, IdBM at the time instant t and frequency bin f is defined as [4]:

$$IdBM(t,f) = \begin{cases} 1 & if\ s(t,f) - m(t,f) > LC \\ 0 & otherwise \end{cases} \quad (1)$$

where $s(t,f)$ and $m(t,f)$ denote the target (speech) and masker (noise) energy values (in decibels) within the unit of time *t* and frequency *f*, respectively.

Several signal processing strategies have been proposed in the literature based on IdBM. These include criteria for noise suppression [5], practical algorithms for de-reverberation [6], noise suppression and new stimulation strategies for cochlear implants [7]. In particular, the algorithms proposed in [8], [9], and [10] show improvement of speech intelligibility in normal hearing, hearing-impaired, and cochlear implant listeners, respectively. These methods are based on training machine learning algorithms according to one or more speech features. For example, in [8], speech signals taken from the IEEE corpus [11] were synthetically added to babble noise. Assuming the knowledge of the SNR at each T-F bin, two separate GMMs, one for target-dominated bins and one for masker-dominated bins, were trained based on the amplitude modulation spectrograms of that bin. In the testing stage, the degraded (noisy) signals were decomposed into T-F bins. In each bin a Bayesian inference procedure was used to estimate the likelihood of that bin being speech or noise dominated. Healy et al. in [9] proposed the use of AMS features as well as relative spectral transform and perceptual linear prediction (RASTA-PLP) and mel-frequency cepstral coefficient (MFCC) features to train a deep neural network (DNN). The trained DNN was used to segregate speech from noise. Their proposed algorithm was able to provide benefit for hearing impaired listeners in terms of speech intelligibility improvement. An ongoing challenge in these types of methods is their inability to segregate speech from an unseen noise, i.e. noise which deviates from the noise that was used during training [12]. Generalizability of these algorithms to unseen noise is an ongoing challenge.

Li and Loizou [13] evaluated the benefits of IdBM when the speech signals were processed with a voice-encoder (vocoder). Their results showed that with low spectral resolution (fewer than 12 channels) and degraded temporal-fine structure (TFS) cues, listeners gained limited benefit from IdBM processing. In their study, the IdBM benefits increased with increasing number of channels.

It has been suggested in the literature that TFS information may be a possible cue for fundamental frequency (F_0) coding which has been shown to play an important role in segregating sound sources into separate auditory streams [1].

http://dx.doi.org/10.21437/Interspeech.2016-1555

In addition, the availability of TFS information is suggested to be important when the listeners "glimpse" the information about the target in the T-F bins with favorable SNRs [14]. In the case of IdBM, all the T-F bins with favorable SNRs (greater than LC in (1)) are available to the listener for glimpsing. As such, consistent with [13], it is reasonable to hypothesize that the masking release resulting from IdBM decreases with the reduction of available TFS information, a deficit observed in hearing-impaired listeners [15].

Another sensory deficit that is suggested to contribute to masking release reduction is impaired frequency selectivity which is a consequence of broadened auditory filters in hearing-impaired listeners [16]. Gnansia et al. in [17] showed that masking release deficits in hearing-impaired listeners were well matched by data obtained from normal hearing listeners presented with speech processed with a $32-$channel tone vocoder. In their study, the bandwidths of the vocoder bandpass filters were increased to account for frequency selectivity reduction and the cutoff frequency of the envelope detection filters was set to 64 Hz to account for TFS degradation.

The purpose of current study is to assess the effects of moderate and severe cochlear hearing loss (i.e. reduced frequency selectivity under conditions where the TFS cues are limited) on masking release (speech intelligibility improvement) obtained from IdBM using similar procedure as in [17].

2. Materials and Methods

Nineteen undergraduate students of the University of Texas at Dallas were recruited to participate in the experiments. Participants received one course credit for their participation. None of the participants reported any hearing loss; this was confirmed by a brief hearing screening conducted prior to the experiment.

The speech material (target) was taken from the IEEE corpus [11]. Each IEEE sentence contains 7 to 12 words, produced by a male speaker in anechoic conditions. The target stimuli were down-sampled from the original sampling rate (25 KHz) to 16 KHz. Speech-shaped noise was added to the target sentences with $\mathrm{SNR} = -5$ dB to generate the degraded (noisy) signals. The spectrum of the speech-shaped noise was determined by averaging the log-magnitude spectra of all the target sentences in the IEEE corpus. This average spectrum was then used to create a $71-$ point, $16-$ kHz finite impulse response filter that was used to shape Gaussian noise to match the average spectrum of the speech signals.

In each trial the target signal was randomly selected (without replacement) from the corpus. The participants were asked to type down the words they heard in a graphical user interface generated in MATLAB. Using an automatic scoring program implemented in MATLAB, the intelligibility scores were calculated as the ratio of the number of correctly identified key words to the total number of presented key words.

The experiment was conducted in a double wall sound-booth located in our laboratory. The experimental design was a 3×3 repeated measure design: three signal processing conditions (quiet, degraded, and IdBM-processed) $\times$ three spectral smearing configurations to account for frequency selectivity in normal hearing listeners, listeners with moderate hearing loss, and listeners with severe hearing loss. In each trial, the participants were presented with stimuli through Tucker-Davis sound system and Sennheiser HD 598 headphones. The stimuli were presented to the listeners at a self-selected comfortable level.

A similar spectral smearing and vocoder processing procedure to that used in [17] was used here to simulate the effects of impaired frequency selectivity and reduced TFS cues. The broad-band signal was passed through a bank of 32 gammatone filters [18], with $1-$, $2-$, or $3-$ ERB, where ERB is the equivalent rectangular bandwidth for the auditory filters [19]. The auditory filters were broadened by a factor of 2 or 3 to account for impaired frequency selectivity caused by moderate and severe hearing loss, respectively [16], [17]. In each frequency band, the envelope was extracted using full-wave rectification and low pass filtering, with a 6^{th} order Butterworth filter whose cutoff was set to 64 Hz to account for reduced TFS information. The IdBM processing mentioned in (1) was then applied to the extracted envelopes in the conditions involving IdBM[1]. The parameter LC in (1) was set to -5 dB. This value is within the optimum range as reported in [20]. A sine-wave carrier with frequency at the characteristic frequency of each of the 32 auditory filters and random starting phase was generated. The sine-waves were then modulated using the extracted (or IdBM-processed) envelopes. The modulated signals were then summed over the 32 frequency bands.

Prior to the actual experiment, participants were asked to listen to 9 practice stimuli (1 stimuli per condition) to become familiar with the vocoded stimuli.

3. Results

Figure 1 shows the speech intelligibility scores in different conditions. A repeated measures analysis of variance indicated a significant main effect of signal processing [$F(2,36) = 1136,\ p < 0.0001$], as well as a significant main effect of spectral smearing [$F(2,36) = 395.3, p < 0.0001$] and a significant signal processing by spectral smearing interaction [$F(4,72) = 126.6, p < 0.0001$] on the speech intelligibility scores.

Post-hoc tests with Bonferroni correction were run on the intelligibility scores in the quiet, degraded (noisy), and IdBM conditions. Interestingly, the differences between the intelligibility scores with the smearing factors of 1 and 2 were *not* significant neither in the quiet nor in the noisy conditions; yet, the differences were significant[2] in the IdBM condition $p < 0.0001$. The differences between the intelligibility scores with the smearing factors of 1 and 3 were significant in the quiet ($p < .0001$), noisy ($p < 0.01$), and IdBM ($p < 0.0001$) conditions. Finally, the differences between the intelligibility scores with the smearing factors of 2 and 3 were significant in

[1] A more plausible design would be performing IdBM before broadening the auditory filters by a factor of 2 or 3 (spectral smearing). We conducted a pilot experiment (5 participants) to see if the results would be different if the IdBM was performed before smearing. Analysis showed that IdBM benefits decrease with impaired frequency selectivity, leading to the same outcomes as presented here.

[2] Throughout the Results section, a significant difference implies a significant decrease of the intelligibility scores in one condition relative to another. Similarly, a non-significant difference implies a non-significant decrease/increase of the intelligibility scores.

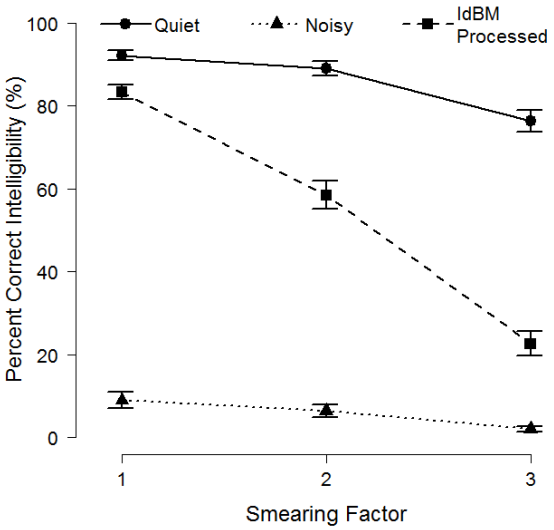

Figure 1: *Speech intelligibility scores versus smearing factor in three different signal processing conditions: Quiet (solid line), Noisy (dotted line), and IdBM (dashed line). The error bars represent the standard error of the means.*

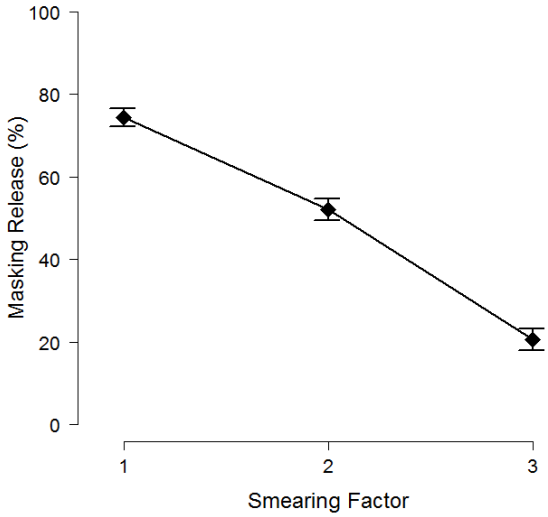

Figure 2: *Masking release, difference between the intelligibility scores in the IdBM and noisy conditions, versus smearing factor. The error bars represent the standard error of the means.*

the quiet ($p < .0001$) and IdBM ($p < 0.0001$) conditions; however, the differences were not significant in the noisy condition.

Bonferroni correction for *post-hoc* analysis resulted in a significant difference between the intelligibility scores in the quiet and noisy conditions, quiet and IdBM conditions, and IdBM and noisy conditions in all the smearing factors ($p < 0.0001$, except for the difference between the quiet and IdBM conditions in smearing factor of 1 with $p < 0.01$).

Figure 2 shows the masking release, i.e. improvement of speech intelligibility scores in the IdBM condition relative to the noisy condition, for three spectral smearing configurations. Bonferroni correction for *post-hoc* analysis resulted in a significant difference between all the possible pairs of spectral smearing configurations ($p < 0.0001$). The results in Figure 1 along with Figure 2 support the hypothesis that with limited TFS cues, the benefits of IdBM (masking release) decrease with impaired frequency selectivity.

4. Discussion

The results presented in this paper suggest that the benefits of IdBM (masking release resulted from IdBM) decreased with severity of hearing loss. Consistent with [17], we simulated the effects of moderate and severe cochlear hearing loss with tone-vocoding and broadening of auditory filters. The results are consistent with those reported in [17] which showed that, with degraded TFs cues, masking release decreases with increased spectral smearing. Note that Gnansia et al. in [17] defined masking release as the difference between the speech intelligibility scores for a modulated masker and the scores for a steady-state masker. In our study, masking release is defined in a similar manner: the difference between the intelligibility scores of the IdBM-processed stimuli and the speech degraded with speech-shaped noise (steady-state). IdBM preserves the T-F bins with SNRs higher than the LC and discards those with lower SNRs, i.e. the T-F bins in which the target is absent or substantially lower in power than the masker. This process is equivalent with adding a fluctuating masker (resulted from processing the original steady-state masker with IdBM) to the target signal.

As mentioned in the Results section, the intelligibility scores for the smearing factors of 1 and 2 in the quiet and noisy conditions were not significantly different. For the IdBM condition, however, the intelligibility scores dropped from 83.29% to 58.45% (Figure 1), or equivalently, the IdBM benefit dropped from 74.35% to 52.04% (Figure 2). Similar to glimpsing, these results show that the benefit of IdBM is highly dependent on the availability of TFS cues as well as frequency selectivity. Although the number of channels in the tone-vocoder used in this study was relatively high (32 channels), the limited TFS cues and the spectral smearing caused a decrease in the intelligibility of the IdBM-processed stimuli. The results reported in this study, along with those reported in [13] show that impaired spectral resolution and frequency selectivity may limit the benefits obtained from IdBM.

Wang et al. in [4] showed that, in some noise conditions, the benefit of IdBM for hearing-impaired listeners is even greater than that of normal hearing listeners. There seems to be a discrepancy between the data provided in this study and those reported in [4]. The results in [4] are surprising given the fact that hearing-impaired listeners have reduced frequency selectivity [15], [16] and impaired perception of TFS cues [16]. One possible reason for such a high IdBM benefit for hearing-impaired listeners in [4] may be the use of closed set speech material. In contrast with this study, the testing material in [4] was drawn from a closed set (Dantale II [21]) with predictable temporal and grammatical structure. Thus, lack of available TFS information, as is the case for hearing impaired listeners, may not be crucial to understand the target stimuli used in Wang et al. study [22, pp. 94]. A similar discrepancy has been reported in Lunner et al. [23]. In their study, the available TFS cues in degraded speech signals were

varied. Their results showed that, compared to normal hearing listeners, hearing-impaired listeners obtained less benefit from having access to the original TFS cues. However, when they changed the speech material to Dantale II, the obtained benefit was the same for both normal hearing and hearing-impaired listeners.

It is important to note that broadening the auditory filters (spectral smearing) without degrading the TFS cues cannot account for the masking release deficits observed in individuals with moderate and severe hearing loss. This is shown in [16]. In contrast, Gnansia et al. in [17] showed that tone-vocoding along with spectral smearing was able to account for deficits caused by cochlear hearing loss. Their results were in good correspondence with those obtained from actual hearing-impaired listeners. This provides us a future direction to evaluate the IdBM benefits in actual hearing impaired-listeners and compare the results with those obtained in this study.

To perform the IdBM in study, we assumed *a priori* knowledge of the ratio of the target and masker energy values in each T-F bin. In practice, however, this ratio needs to be estimated using a signal processing algorithm, such as in [24]. Inevitably, such estimation results in false alarms (retaining a noise-dominated T-F bin) as well as misses (discarding a target-dominated T-F bin). As shown in [20], the intelligibility of IdBM-processed speech decreases dramatically with the increase of false alarms. As such, further investigation is needed to extend the current study to the non-ideal binary masking (nIdBM) conditions, in which false alarms and misses are artificially introduced to the IdBM-processed speech, i.e. to investigate the effects of hearing loss on the benefits of nIdBM.

By comparing the intelligibility of broadband IdBM-processed speech in [20] and [25], it seems that using Fast Fourier Transform (FFT) for IdBM results in higher intelligibility scores, compared to the case in which speech signals are passed through band-pass filters. The present study has used a bank of 32 band-pass filters and applied the IdBM processing to the envelopes extracted from the output of each filter. Thus, it is likely that the intelligibility scores would have been higher if FFT had been used instead of band-pass filters. Regardless of the levels of intelligibility, the key point is that IdBM benefits decrease with impaired frequency selectivity.

5. Conclusions

In this study, we evaluated the effects of reduced frequency selectivity and degraded TFS information on the speech intelligibility benefits obtained from IdBM. Three smearing factors were considered here to simulate the effects of reduced frequency selectivity in normal hearing and hearing impaired with moderate and severe hearing loss, respectively. The data suggests that the intelligibility of the IdBM-processed stimuli decreased with increasing the spectral smearing. Future directions may entail the evaluation of IdBM benefits for actual hearing impaired individuals and comparison of their intelligibility scores with those obtained in this study. Future studies will also evaluate the effects of cochlear hearing loss on the benefits obtained from nIdBM.

6. Acknowledgements

The authors would like to thank all the individuals who participated in this study.

7. References

[1] A. Bregman, *Auditory scene analysis*. Cambridge, Mass.: MIT Press, 1990.

[2] D. Wang and G. Brown, *Computational auditory scene analysis*. Hoboken, N.J.: Wiley Interscience, 2006.

[3] D. Wang, "On ideal binary mask as the computational goal of auditory scene analysis", in *Speech Separation by Humans and Machines*, 1st ed., P. Divenyi, Ed. Dordrecht: Kluwer Academic, 2005, pp. 181-187.

[4] D. Wang, U. Kjems, M. Pedersen, J. Boldt and T. Lunner, "Speech intelligibility in background noise with ideal binary time-frequency masking", *The Journal of the Acoustical Society of America*, vol. 125, no. 4, pp. 2336-2347, 2009.

[5] O. Hazrati and P. Loizou, "Comparison of two channel selection criteria for noise suppression in cochlear implants", *The Journal of the Acoustical Society of America*, vol. 133, no. 3, p. 1615, 2013.

[6] O. Hazrati, J. Lee and P. Loizou, "Blind binary masking for reverberation suppression in cochlear implants", The Journal of the Acoustical Society of America, vol. 133, no. 3, p. 1607, 2013.

[7] Y. Hu and P. Loizou, "A new sound coding strategy for suppressing noise in cochlear implants", *The Journal of the Acoustical Society of America*, vol. 124, no. 1, p. 498, 2008.

[8] G. Kim, Y. Lu, Y. Hu and P. Loizou, "An algorithm that improves speech intelligibility in noise for normal-hearing listeners", *The Journal of the Acoustical Society of America*, vol. 126, no. 3, p. 1486, 2009.

[9] E. Healy, S. Yoho, Y. Wang and D. Wang, "An algorithm to improve speech recognition in noise for hearing-impaired listeners", *The Journal of the Acoustical Society of America*, vol. 134, no. 4, p. 3029, 2013.

[10] Y. Hu and P. Loizou, "Environment-specific noise suppression for improved speech intelligibility by cochlear implant users", *The Journal of the Acoustical Society of America*, vol. 127, no. 6, p. 3689, 2010.

[11] "IEEE Recommended Practice for Speech Quality Measurements", *IEEE Transactions on Audio and Electroacoustics*, vol. 17, no. 3, pp. 225-246, 1969.

[12] T. May and T. Dau, "Requirements for the evaluation of computational speech segregation systems", The Journal of the Acoustical Society of America, vol. 136, no. 6, pp. EL398-EL404, 2014.

[13] N. Li and P. Loizou, "Effect of spectral resolution on the intelligibility of ideal binary masked speech", *The Journal of the Acoustical Society of America*, vol. 123, no. 4, pp. EL59-EL64, 2008.

[14] K. Hopkins and B. Moore, "The importance of temporal fine structure information in speech at different spectral regions for normal-hearing and hearing-impaired subjects", *The Journal of the Acoustical Society of America*, vol. 127, no. 3, p. 1595, 2010.

[15] K. Hopkins and B. Moore, "The contribution of temporal fine structure information to the intelligibility of speech in noise",

The Journal of the Acoustical Society of America, vol. 123, no. 5, p. 3710, 2008.

[16] A. Léger, B. Moore, D. Gnansia and C. Lorenzi, "Effects of spectral smearing on the identification of speech in noise filtered into low- and mid-frequency regions", *The Journal of the Acoustical Society of America*, vol. 131, no. 5, p. 4114, 2012.

[17] D. Gnansia, V. Péan, B. Meyer and C. Lorenzi, "Effects of spectral smearing and temporal fine structure degradation on speech masking release", *The Journal of the Acoustical Society of America*, vol. 125, no. 6, p. 4023, 2009.

[18] M. Slaney, "Auditory Toolbox", Interval Research Corporation, 1998.

[19] 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.

[20] N. Li and P. Loizou, "Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction", *The Journal of the Acoustical Society of America*, vol. 123, no. 3, p. 1673, 2008.

[21] K. Wagener, J. Josvassen and R. Ardenkjær, "Design, optimization and evaluation of a Danish sentence test in noise", *International Journal of Audiology*, vol. 42, no. 1, pp. 10-17, 2003.

[22] B. Moore, *Auditory Processing of Temporal Fine Structure*. World Scientific Publishing Company, 2014.

[23] T. Lunner, R. Hietkamp, M. Andersen, K. Hopkins and B. Moore, "Effect of Speech Material on the Benefit of Temporal Fine Structure Information in Speech for Young Normal-Hearing and Older Hearing-Impaired Participants", Ear and Hearing, vol. 33, no. 3, pp. 377-388, 2012.

[24] G. Hu and D. Wang, "Monaural Speech Segregation Based on Pitch Tracking and Amplitude Modulation", *IEEE Trans. Neural Netw.*, vol. 15, no. 5, pp. 1135-1150, 2004.

[25] D. Brungart, P. Chang, B. Simpson and D. Wang, "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation", *The Journal of the Acoustical Society of America*, vol. 120, no. 6, p. 4007, 2006.