Parameter tuning of time-frequency masking algorithms for reverberant artifact removal within the cochlear implant stimulus

Lidea K. Shahidi, Leslie M. Collins, Boyla O. Mainsah

Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA

Cochlear implant recipients struggle to understand speech in reverberant environments. To restore speech perception, artifacts due to reverberant reflections can be removed from the cochlear implant stimulus by applying a matrix of gain values, a technique referred to as time-frequency masking. In this study, two common time-frequency masking strategies are implemented within cochlear implant processing, either introducing complete retention or deletion of stimulus components using a binary mask or continuous attenuation of stimulus components using a ratio mask. Parameters of each masking strategy control the level of attenuation imposed by the gain values. In this study, we perceptually tune the parameters of the masking strategy to determine a balance between speech retention and artifact removal. We measure the intelligibility of reverberant signals mitigated by each strategy with speech recognition testing in normal-hearing listeners using vocoding as a simulation of cochlear implant perception. For both masking strategies, we find parameterizations that maximize the intelligibility of the mitigated signals. At the best-performing parameterizations, binary-masked reverberant signals yield larger intelligibility improvements than ratio-masked signals. The results provide a perceptually optimized objective for the removal of reverberant artifacts from cochlear implant stimuli, facilitating improved speech recognition performance for cochlear implant recipients in reverberant environments.

Key Words: Cochlear Implants, Reverberation, Time-frequency masking, ACE processing

Introduction

Reverberant environments can degrade the intelligibility of speech for both normal-hearing (NH) listeners and hearing-impaired (HI) listeners, such as cochlear implant (CI) users (Kokkinakis et al., 2011; Kressner et al., 2018; Neuman et al., 2010). Reverberation occurs when sound reflects off of surfaces in an enclosure. When reverberant reflections arrive at a listener at the same time as the signal’s direct path from the source, the direct path signal becomes distorted. For speech, energy from reverberant reflections obscures gaps between phonemes and degrades active speech segments, blurring the temporal and spectral cues that facilitate speech understanding (Nabelek et al., 1989). CI users experience significant degradations in speech intelligibility in reverberation in part due to the limited temporal and spectral information provided by the CI after speech processing (Loizou, 2006).

To mitigate signal distortions caused by adverse listening environments, gain values can be applied to selectively attenuate components of the distorted signal. When a matrix of gain values is applied to a time-frequency representation of the distorted signal, this approach to speech enhancement is referred to as time-frequency (T-F) masking. Gain values, and thus the amount of attenuation imposed, are determined using a criterion that quantifies the relative amount of the desired and interfering signals; in this case, the direct path and the reverberant reflections, respectively. There are two primary types of ideal T-F masks: the ideal binary mask (IBM) and the ideal ratio mask (IRM) (Loizou, 2007; Wang, 2005). The IBM contains only binary gain values, 0 or 1, leading to deletion or retention of T-F units of the distorted signal, respectively. The IRM contains gain values that assume continuous values between 0 and 1, leading to partial attenuation of highly distorted speech segments.

Typically, T-F masking is used as a pre-processing step in most speech enhancement applications and the time-domain waveform is resynthesized after mask application. T-F masking has demonstrated effective speech enhancement when applied to speech in background noise (Anzalone et al., 2006; Li and Loizou, 2008; Wang, 2008) and recent
efforts have extended T-F masking to improve speech intelligibility in reverberant conditions for NH individuals (Roman and Woodruff, 2011, 2013; Zhao et al., 2016, 2017; Zhao, Xu et al., 2018) and HI listeners (Hazarati, Lee et al., 2013; Hazratı, Omid Sadjadi et al., 2013; Hazratı and Loizou, 2012, 2013; Healy et al., 2019; Kokkinakis et al., 2011; Zhao, Wang et al., 2018). Although some studies have employed binary masking (Hazarati, Lee et al., 2013; Hazratı and Loizou, 2012, 2013; Kokkinakis et al., 2011) or ratio masking (Hazarati, Omid Sadjadi et al., 2013; Kokkinakis and Stohl, 2021) as a pre-processor to remove reverberant distortions prior to CI processing, additional signal distortions can be imposed by the time-domain resynthesis step when large differences between adjacent time or frequency bins are introduced by T-F masking. Implementation of T-F masking to stimulus pulses during CI processing could enable efficient real-time processing and eliminate artifacts imposed by temporal resynthesis. To date, it is unclear whether the binary or ratio masking strategy would result in the most benefit to reverberant speech intelligibility for CI users when masks are applied at the pulse level. While the IBM can eliminate CI stimulus pulses dominated by reverberant distortion, the IRM could enable the restoration of the amplitude modulation of partially distorted pulses. Previous work found that, when mitigating speech in background noise, the IRM resulted in better speech intelligibility outcomes for NH listeners with vocoding but no significant difference in speech intelligibility outcomes for CI users (Koning et al., 2015). Whether similar results are applicable to reverberant speech stimuli remains undetermined.

As a priori knowledge of the direct path signal is required to compute ideal T-F masks, ideal masks are often used as a performance objective for signal processing or machine learning algorithms for reverberant artifact removal (Hazarati, Lee et al., 2013; Hazratı, Omid Sadjadi et al., 2013; Hazratı and Loizou, 2013; Healy et al., 2019; Li et al., 2017; Zhao et al., 2016, 2017; Zhao, Wang et al., 2018). The potential improvement in speech intelligibility due to the T-F masking performance objective is mediated by the parameterization of each masking strategy. The IBM and IRM gain values are determined by parameters that impact the amount of signal retention and therefore the intelligibility of the mitigated speech stimuli. Many studies employing T-F masking strategies for noise mitigation have used default mask parameters that maximized the signal-to-noise ratio (SNR) of the mitigated speech signal (Healy et al., 2019; Zhao et al., 2016, 2017; Zhao, Wang et al., 2018; Zhao, Xu et al., 2018), despite the fact that alternative parameterizations can yield better speech intelligibility outcomes (Brungart et al., 2006; Wang et al., 2009). Other studies employing T-F masking strategies selected parameterizations heuristically without specifying the selection criteria (Hazarati, Lee et al., 2013; Hazratı, Omid Sadjadi et al., 2013; Hazratı and Loizou, 2013). Some studies have explored a range of mask parameterizations of the IBM in reverberation, either with speech intelligibility testing in NH listeners (Roman and Woodruff, 2013) or CI users (Kokkinakis et al., 2011), indicating other parameterizations that led to better speech intelligibility outcomes than the default parameters. More recently, (Kokkinakis and Stohl, 2021) explored parameterizations of the IRM to restore speech intelligibility in reverberation for CI users. Kokkinakis and Stohl, 2021 first performed an offline analysis to select mask parameters that led to the largest improvements in objective intelligibility measure scores and then tested the intelligibility of reverberant speech signals mitigated by the most promising parameterizations in listening studies with CI users. Their analysis indicated that ratio mask parameterizations that differed from the default IRM parameters could significantly improve speech intelligibility outcomes, although they relied on the assumption that objective intelligibility measures reflect subjective intelligibility trends and a thorough search of the ratio mask parameterization space was not conducted using subjective intelligibility testing. These previous studies established a general framework for tuning mask parameters using subjective testing. However, when reverberant speech was considered, the analyses included only one mask type and employed T-F masking as a pre-processing step.

In this study, we investigate the impact of T-F mask parameterizations for both masking strategies on the intelligibility of reverberant speech when masks are used to remove stimulus pulses dominated by reverberant distortions. Masks are applied directly within Advanced Combination Encoder (ACE) processing, a CI processing strategy commonly employed within Nucleus CI devices (manufactured by Cochlear Corporation), avoiding distortions often imposed by time-domain resynthesis and minimizing the delay imposed by the mitigation algorithm. Intelligibility outcomes for the information retained within the mitigated reverberant stimuli are tested with NH listeners using vocoding as a simulation of CI perception.

**Impact of T-F mask parameters on the reverberant CI stimulus pattern**

T-F masking enables the dynamic attenuation of stimulus pulses distorted by reverberation through the application of unique gain values to each stimulus pulse. Gain values are determined by a local measure.
of reverberant distortion in each stimulus pulse. In this study, we use the speech-to-reverberant ratio (SRR) (Naylor and Gaubitch, 2010) to quantify the amount of reverberant distortion in each stimulus pulse. The SRR describes the proportion of energy in a stimulus pulse that is due to the direct path signal, or the speech signal that travels directly from the speaker to the listener, as described in equation (3) of the Supplemental Material (available online). The SRR is analogous to the SNR of a signal in background noise. The SRR will assume negative values when the energy of the reverberant reflections in a stimulus pulse is greater than the energy of the direct path signal and assumes positive values when the contribution of the direct path signal to the stimulus pulse is greater than the contribution of the reverberant reflections.

For each stimulus pulse, a gain value is calculated from the SRR using a gain function. To retain stimulus pulses predominantly containing the direct path signal, the IBM gain function can be used (Kokkinakis et al., 2011; Roman and Woodruff, 2013). The IBM thresholds the local SRR using the parameter $\tau$, the local threshold in dB, as presented in equation (4) of the Supplemental Material. The local threshold impacts the shift of the IBM gain function, as shown in Fig. 1A. Typically, $\tau$ is set to 0 dB (Li and Wang, 2009), reflecting the assumption that signal components predominantly due to the direct path signal are more intelligible. As an alternative to thresholding the SRR to generate a binary mask, the IRM gain function can be used to map the SRR to continuous gain values between 0 and 1 (Lim and Wang, 2009), reflecting the assumption that signal components predominantly due to the direct path signal are more intelligible. An alternative to thresholding the SRR to generate a binary mask, the IRM gain function can be used to map the SRR to continuous gain values between 0 and 1 (Lim and Oppenheim, 1979; Zhao, Wang et al., 2018), as presented in equation (6) of the Supplemental Material. The IRM gain function is characterized by the parameters $\alpha$ and $\beta$ that control the shift and slope, respectively, of the gain function, as illustrated in Fig. 1B and C. In studies utilizing the IRM for mitigating reverberation, $\alpha$ and $\beta$ are typically set to 1 and 0.5, respectively (Li et al., 2017; Zhao, Wang et al., 2018; Zhao, Xu et al., 2018), as this results in a parameterized IRM that is equivalent to a constrained (Weiner) filter that optimally removes additive background noise (Loizou, 2007).

The choice of mask parameter values impacts the effectiveness of reverberation mitigation as they control the sparsity of the T-F mask, and consequently the amount of retention of the signal being mitigated (Roman and Woodruff, 2013). Figure 2 illustrates the reverberant CI stimuli before and after the application of T-F masks with different parameterizations. Under-attenuation of reverberant signals due to dense masks generated from low parameter values results in mitigated signals that contain large amounts of reverberant artifacts between phonemes (Fig. 2C and D), potentially degrading the spectro-temporal cues that are beneficial for speech intelligibility. Over-attenuation of reverberant signals due to sparse masks generated with high parameter values results in mitigated signals where important speech information may have been removed (Fig. 2G and H), potentially degrading speech intelligibility. At a balance between signal retention and artifact removal, the mitigated stimuli resemble the direct path stimuli (Fig. 2A), as shown in Fig. 2E and F for the binary and ratio mask, respectively.

It remains an open question whether alternative parameterizations of $\tau$ for the IBM and $\alpha$ and $\beta$ for the IRM would benefit speech intelligibility in reverberant conditions when masks are applied within CI processing. In this study, we investigated speech intelligibility outcomes when the CI stimulus is mitigated by the IBM across parameterizations of $\tau$ or by the IRM across parameterizations for $\beta$. We fixed the value of $\alpha$ in the IRM as both $\tau$ and $\alpha$ control the shift of the gain function, as illustrated in Fig. 1. To test the intelligibility of the acoustic information retained in the CI stimulus after mask application, we conducted a subjective intelligibility study in which the mitigated stimulus was vocoded and presented to NH listeners. By testing in NH listeners, we isolated the benefit of the retained information without including confounding individual-specific factors present with CI listeners, such as the duration of deafness or electrode array insertion depth. Results obtained from NH listeners using vocoding often correlate with the performance trends of CI users (Blamey et al., 1984; Dorman et al., 1997; Strydom and Hanekom, 2011). Thus, general trends in mask parameterization strategies are likely to translate to CI users but with a larger variation in the benefits conferred by the mitigation strategy expected across CI users.

**Methods**

This experiment compared the effectiveness of T-F masking strategies applied within ACE CI processing (McDermott et al., 1992) over a range of parameterizations within each mask type. To mitigate reverberation, a binary or ratio mask was generated for each parameter value and directly applied to T-F units of a reverberant CI stimulus pattern. The stimulus pattern was then vocoded and the resulting speech intelligibility of the mitigated reverberant signal was assessed with NH listeners. For comparison, the unmitigated reverberant signal and the direct path reverberant signal were also tested to provide lower and upper bounds, respectively, of speech intelligibility.
Speech stimuli

Speech material consisted of sentences from the Hearing in Noise Test (HINT) speech corpus spoken by a single male speaker (Nilsson et al., 1994). The HINT corpus consists of 25 phonemically balanced lists. Each list contains 10 sentences, with 6–7 syllables in length. Prior to processing, the speech signals were first downsampled from 22160 Hz to 16000 Hz to mirror the sampling rate employed in modern CI devices using the ACE strategy. Reverberant speech material was created by convolving anechoic speech material with a room impulse response (RIR) function that models the effect of a reverberant enclosure on an acoustic signal (Kuttruff, 2009) as in equation (1) of the Supplemental Material. The RIR function was obtained from the Aachen Impulse Response (AIR) database (Jeub et al., 2009). The RIR used in this experiment had a reverberation time (RT60) of 0.8 seconds and was recorded in a lecture hall with a room length of 10.8 m, width of 10.9 m and height of 3.15 m. The source-to-microphone distance was 7.1 m, which is well outside the critical distance of the room, approximated as 1.2 meters (Naylor and Gaubitch, 2010). This reverberant condition was chosen as it provides a realistic scenario in which a CI user would likely have difficulty understanding speech (Kokkinakis et al., 2011; Kressner et al., 2018). To create the direct path signal, the RIR was truncated to remove impulses occurring 5 milliseconds after the initial impulse (Naylor and Gaubitch, 2010) and convolved with the anechoic speech material. The root-mean-square (RMS) level of all reverberant speech material was normalized to the highest RMS value that avoided clipping.

Reverberant signals were passed through a simulation of the ACE processing (implemented using the Nucleus MATLAB Toolbox (Swanson and Mauch, 2006)) resulting in reverberant stimulus patterns. Within the ACE processing simulation, signals are transformed into the frequency domain using a discrete Fourier transform with a window length of

![Figure 1](image1.png)

**Figure 1** Gain function curves mapping speech-to-reverberant (SRR) values to mask values when (A) varying the local threshold \( \tau \) for the ideal binary mask (IBM), (B) varying the shift \( \alpha \) (at slope \( \beta = 1 \)) for the ideal ratio mask (IRM), and (C) varying the slope \( \beta \) (at shift \( \alpha = 1 \)) for the IRM.

![Figure 2](image2.png)

**Figure 2** Electrodograms of the speech token ‘A boy fell from the window,’ showing the CI stimulation pattern of (A) the direct path signal; (B) the reverberant signal; and mitigated reverberant signals obtained by applying (C, E, G) ideal binary masks (IBMs) and (D, F, H) ideal ratio masks (IRMa) with increasing local threshold and slope parameter values, \( \tau \) and \( \beta \), respectively, to the reverberant signal. The mask parameter controls the sparsity of the mask, resulting in under-attenuation (C and D) to over-attenuation (G and H) of the reverberant signal as the parameter value increases.
128 samples and a frame shift of 32 samples. A spectral weighting is used to group the magnitude squared Fourier coefficients into 22 electrode channels. Then, envelopes are extracted by applying a square root transform. In this implementation of ACE processing, 8 of the 22 electrode channels with the largest magnitudes are selected for stimulation in each processing cycle. The stimulus is then scaled to fall between base and saturation levels and logarithmically compressed to match the electrical dynamic range.

T-F masks were computed for reverberant CI stimulus patterns obtained after envelope extraction and applied prior to channel selection. IBMs were created for $\tau$ values in the range -50 to 12 dB while IRMs were created for $\beta$ values in the range 0.01 to 5 with $\alpha$ set to 1. We set $\alpha = 1$ based on the intuition that the magnitudes of the direct path signal and the reverberant reflections should be on the same scale (Naylor and Gaubitch, 2010). The methods used in this study for calculating the SRR, IBM, and IRM, are presented in equations (3), (4), and (6) of the Supplemental Material, respectively.

Vocoded waveform stimuli of the mitigated reverberant stimulus pattern were generated by using the CI pulse amplitudes after logarithmic compression as the amplitudes of sinusoidal carriers, with the frequencies of the sinusoidal carriers corresponding to the frequencies of electrode channels in the default ACE map. All sinusoid vocoded channels were then summed to create the final audio stimulus that was presented acoustically to NH listeners.

Study protocol

Twenty native speakers of American English were recruited to participate in this study, which was approved by the Duke University Institutional Review Board. The subject ages ranged from 19 to 35 years. All subjects had self-reported normal hearing and were paid for their participation. Informed consent was obtained for each subject prior to their participation.

During the test, subjects were seated at a computer in a soundproof booth. Vocoded speech stimuli were presented to the listener diotically through Sony MDR7506 headphones at a sound-pressure level (SPL) of 65 dB. Each sentence was presented once, and subjects were instructed to type the words they were able to hear. A study administrator was seated in the booth with the subject during the study.

Each session began with an adaptive training task to familiarize subjects with listening to vocoded speech. Subjects were presented with feedback and training ended when their performance plateaued over five sentences or when thirty sentences were presented. After training, the direct path condition and the unmitigated condition were presented. Ratio mask conditions and binary mask conditions were then presented in blocks, with $\tau$ and $\beta$ conditions randomized within each block. The presentation order of ratio mask and binary mask blocks was counterbalanced across subjects. Each task used a randomly selected HINT list, sampled without replacement. Each session lasted 45-60 minutes, including breaks.

Results

Figure 3 shows the speech intelligibility results, measured as the percent of correct phonemes, for IBM and IRM conditions across $\tau$ and $\beta$ values, respectively. The direct path condition resulted in speech recognition accuracy of 91.2% ± 6.6%, which is similar to previously reported speech intelligibility performance with vocoded speech in NH users under anechoic conditions (Whitmal et al., 2007). The unmitigated condition resulted in an accuracy of 11.2% ± 9.1%, confirming the difficulty of the reverberant condition. For the mitigated reverberant conditions, there was a concave relationship between speech intelligibility and mask parameter value for both T-F masking strategies, suggesting that speech intelligibility reflects a balance between retention of active speech and deletion of detrimental reverberant signal components.

A repeated measures analysis of variance (ANOVA) test was performed to assess the effectiveness of mask parameterization on speech intelligibility, followed by post hoc Tukey’s HSD (honestly significant difference) test for pairwise comparisons. Subject mean scores were first transformed using the rationalized arcsine transform (Studebaker, 1985) prior to the ANOVA analysis. A significant effect of mask parameter value for both the IRM [F(10, 190) = 50.1, $p < 0.0001$] and the IBM [F(10, 190) = 50.02, $p < 0.0001$] was observed, with large partial $\eta^2$ effect sizes for group mean differences of 0.7250 and 0.7247, respectively. The pairwise comparisons revealed that the unmitigated condition was significantly different ($p < 0.05$) from mitigated reverberation conditions for a subset of parameter values for both masks (see Fig. 3). Averaging across subjects, the maximum performance was achieved at $\tau = -6$ dB for the IBM and $\beta = 0.25$ with $\alpha = 1$ for the IRM, with average speech intelligibility raised by 40.2% and 31.1%, respectively, from the unmitigated condition. These results demonstrate the sensitivity of T-F masking approaches to the choice of parameter settings.

Discussion

This study investigated the potential of applying different T-F masking strategies within CI processing and the impact of mask parameterization on the resulting speech intelligibility in reverberant
The intelligibility of the mitigated CI stimulus pattern was significantly improved after the application of T-F masking, suggesting that T-F masking is a viable option for the mitigation of reverberant artifacts in the CI stimulus and could considerably lower the barriers to speech perception for CI users in reverberant listening environments.

Speech intelligibility was investigated when stimulus patterns were mitigated by either a binary mask, which removed reverberant-dominant stimulus pulses from the stimulus pattern, or a ratio mask, which attenuated stimulus pulse amplitudes depending on the amount of reverberant distortion present. Binary mask-mitigated signals resulted in larger improvements in speech intelligibility than ratio mask-mitigated signals, suggesting that partial retention of distorted speech-dominant stimulus pulses using the ratio mask did not provide useful cues for speech intelligibility. Anecdotal evidence from an informal survey of half of the subjects revealed an overwhelming preference (9 out of 10 subjects) for the binary mask conditions in terms of speech quality. These results suggest that for mitigating reverberant signals with the same time and frequency resolution as the CI stimulus, adjusting the center of the gain function (e.g., by modifying the $\tau$ parameter of the binary mask) may lead to better improvements in speech intelligibility than adjusting the slope of the gain function (e.g., by modifying the $\beta$ parameter of the ratio mask).

By testing over a variety of T-F masking parameter values that resulted in under- to over-attenuation of the reverberant signal, we demonstrated that the parameters that maximized speech intelligibility differed from the default parameters typically employed. In contrast to the default IBM threshold of 0 dB, we found a balance between signal retention and removal using a threshold value of -6 dB. This result is in agreement with studies of binary mask parameterization in reverberation recommending threshold values of -6 dB for NH listeners (Roman and Woodruff, 2013) and -5 dB for CI users (Kokkinakis et al., 2011) when masks were applied to T-F representations with greater spectral resolution than the CI stimulus pattern. Our result reveals that a threshold value of -6 dB yielded improved intelligibility outcomes even when applied directly to the CI stimulus pulses, suggesting that the change in spectral resolution does not impact the tolerance of speech intelligibility for reverberant distortions in the mitigated signal. In ratio masked conditions, subjects benefited from slope parameterizations that resulted in less aggressive attenuations than the default parameterizations, similar to the findings of (Kokkinakis and Stohl, 2021). To our knowledge, the analysis of ratio masking slope parameters presented here is the only such search of speech intelligibility outcomes across gradual changes in slope parameterizations in reverberant conditions. Overall, the perceptually optimized parameterizations indicate that T-F masking has the potential to significantly improve speech intelligibility for CI recipients in reverberant conditions when applied during CI signal processing.

Additional efforts are needed to create a T-F masking mitigation strategy that can benefit speech intelligibility for CI users in real-world conditions. While this study examined speech intelligibility outcomes of the mitigated CI stimulus using NH listeners, intelligibility outcomes in CI users must still be examined. Further, the masks examined in this study resulted from ideal knowledge of the anechoic speech signal and the impulse response of the room.
A real-world T-F masking strategy would require blind estimation of the SRR from the reverberant signal while this study assumed knowledge of the RIR and the anechoic speech signal to calculate the SRR.

This work supplies an informed parameterization objective for a potential machine learning paradigm that aims to estimate T-F masks in real-world scenarios. For example, a machine learning algorithm could be trained to estimate IBMs using only knowledge of the reverberant signal, with ground truth binary masks based on IBMs generated at a threshold of -6 dB. Towards the goal of creating a T-F masking strategy implementable in CI devices, our preliminary work indicates that machine learning models have the potential to estimate T-F masks for application within CI processing, resulting in improved intelligibility of the mitigated CI stimulus pattern (Chu et al., 2021; Shahidi et al., 2021).

Conclusions
In this study, we investigated the intelligibility outcomes of T-F masking strategies when used to identify and remove reverberant artifacts within the CI stimulus. T-F masks were formulated across a range of mask parameterizations for either the binary or ratio mask and applied within the ACE processing chain prior to channel selection. The intelligibility of the speech information retained after the application of T-F masks was tested with NH listeners using vocoding as a simulation of CI perception. Both binary and ratio masking strategies improved reverberant speech intelligibility when applied within ACE processing. Parameterizations of binary and ratio masks were identified that lead to the largest improvements in speech intelligibility at a balance between signal retention and artifact removal. In general, binary masked conditions led to greater improvements in intelligibility than ratio masked conditions, suggesting that partial retention of distorted speech-dominant stimulus segments did not provide useful cues for speech intelligibility. When developing algorithms to detect and mitigate reverberant artifacts within the CI stimulus, the mask type and mask parameterization should be selected to maximize the potential improvements in intelligibility conferred by the algorithm.

Acknowledgements
The authors would like to thank the subjects who participated in this experiment.

Notes on contributors
Lidea K. Shahidi received her PhD in Electrical and Computer Engineering from Duke University in May 2022. She is currently a Research Associate at the University of Cambridge working in the Cambridge Hearing Group. Her research interests include speech enhancement and speech coding strategies for cochlear implants.

Leslie M. Collins (PhD) is a Professor of Electrical and Computer Engineering at Duke University.

Boyla O. Mainsah (PhD) is an Assistant Research Professor of Electrical and Computer Engineering at Duke University.

Disclaimer statements
Contributors None.

Funding This work was funded by the National Institutes of Health via a grant administered by the National Institute on Deafness and Other Communication Disorders (R01-DC014290).

Conflicts of interest None.

Ethics approval This study was approved by the Duke University Institutional Review Board.

Supplemental data
Supplemental data for this article can be accessed at https://doi.org/10.1080/14670100.2022.2096182.

ORCID
Lidea K. Shahidi http://orcid.org/0000-0003-1369-6011
Leslie M. Collins http://orcid.org/0000-0002-9043-6531
Boyla O. Mainsah http://orcid.org/0000-0003-2612-0679

References
Anzalone, M.C., Calandrucio, L., Doherty, K.A., Carney, L.H. 2006. Determination of the potential benefit of time-frequency gain manipulation. Ear and Hearing, 27(5): 480–492. doi:10.1097/01.aud.0000233918.86909.df.

Blamey, P.J., Dowell, R.C., Tong, Y.C., Clark, G.M. 1984. An acoustic model of a multiple-channel cochlear implant. The Journal of the Acoustical Society of America, 76(1): 97–103. doi:10.1121/1.391012.

Brungart, D.S., Chang, P.S., Simpson, B.D., Wang, D. 2006. Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. Journal of the Acoustical Society of America, 119(6): 4007–4018. doi:10.1121/1.2363929.

Chu, K. M., Collins, L. M., & Mainsah, B. O. (2021). Phoneme-based time-frequency mask estimation for reverberant speech enhancement for cochlear implant users. Conference on Implantable Auditory Prostheses, Online/virtual conference.

Dorman, M.F., Loizou, P.C., Rainey, D. 1997. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. The Journal of the Acoustical Society of America, 102: 2403–2411. doi:10.1121/1.419603.

Hazzati, O., Lee, J., Loizou, P.C. 2013. Blind binary masking for reverberation suppression in cochlear implants. The Journal of the Acoustical Society of America, 133(3): 1607–1614. doi:10.1121/1.4789891.

Hazzati, O., Loizou, P.C. 2012. Tackling the combined effects of reverberation and masking noise using ideal channel selection. The Journal of Speech, Language, and Hearing Research, 55(2): 500–510. doi:10.1044/1092-4388(2011/11-0073).

Hazzati, O., Loizou, P.C. 2013. Reverberation suppression in cochlear implants using a blind channel-selection strategy.
The Journal of the Acoustical Society of America, 133(6): 4188–4196. doi:10.1121/1.4804313.

Hazarati, O., Omid Sadjadi, S., Loizou, P.C., Hansen, J.H.L. 2013. Simultaneous suppression of noise and reverberation in cochlear implants using a ratio masking strategy. The Journal of the Acoustical Society of America, 134(5): 3759–3765. doi: 10.1121/1.4822389.

Healy, E., Delfarad, M., Johnson, E., Wang, D. 2019. A deep learning algorithm to increase intelligibility for hearing-impaired listeners in the presence of a competing talker and reverberation. The Journal of the Acoustical Society of America, 145 (3): 1874–1874. doi: 10.1121/1.5101778.

Jeub, M., Schafer, M., & Vary, P. (2009). A binaural room impulse response database for the evaluation of time-frequency masks. 16th International Conference on Digital Signal Processing (Vols 1 and 2, pp. 550–554), Santorini, Greece.

Kokkinakis, K., Hazrati, O., Loizou, P.C. 2011. A channel-selection criterion for suppressing reverberation in cochlear implants. The Journal of the Acoustical Society of America, 129(5): 3221–3232. doi: 10.1121/1.359683.

Kokkinakis, K., Stohl, J.S. 2021. Optimized gain functions in ideal time-frequency masks and their application to dereverberation for cochlear implants. Journal of the Acoustical Society of America, 144(2): 1113–1122. doi: 10.1121/1.5105640.

Kuttruff, H. 2009. Room acoustics. London: CRC Press.

Li, X., Li, J., & Yan, Y. (2017). Ideal ratio mask estimation using deep neural networks for monoaural speech segregation in noisy environments. Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH) (pp. 1203–1207), London. doi: 10.21437/Interspeech.2017-549.

Li, N., Loizou, P.C. 2008. Factors influencing intelligibility of ideal binary time-frequency masks for speech degradation and noise reduction. The Journal of the Acoustical Society of America, 123(3): 1673–1682. doi: 10.1121/1.2832617.

Li, Y., Wang, D.L. 2009. On the optimality of ideal binary time-frequency masks. Speech Communication, 51(3): 230–239. doi: 10.1016/j.specom.2008.09.001.

Lin, J.S., Oppenheim, A.V. 1979. Enhancement and bandwidth compression of noisy speech. Proceedings of the IEEE, 67 (12): 1586–1604. doi: 10.1109/PROC.1979.11540.

Loizou, P.C. 2006. Speech processing in vocoder-centric cochlear implants. Advances in Oto-Rhino-Laryngology, 69: 104–143. doi: 10.1159/000094648.

Loizou, P.C. 2007. Speech enhancement: theory and practice. Boca Raton, FL: CRC Press. doi: 10.1201/9781420015836.

McDermott, H.J., McKay, C.M., Vandali, A.E. 1992. A new portable sound processor for the University of Melbourne/nucleus limited multielectrode cochlear implant. The Journal of the Acoustical Society of America, 91(6): 3367–3371. doi: 10.1121/1.402826.

Nabelek, A.K., Letowski, T.R., Tucker, F.M. 1989. Reverberant overlap-masking and self-masking in consonant identification. The Journal of the Acoustical Society of America, 86(4): 1259–1265.

Naylor, P.A., Gaubitch, N.D. 2010. Speech dereverberation. London: Springer-Verlag. doi: 10.1007/978-1-84996-056-4_1.

Neuman, A.C., Wroblewski, M., Hajicek, J., Rubinstein, A. 2010. Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. Ear and Hearing, 31(3): 336–344. doi: 10.1097/AUD.0b013e3181d3f514.

Nilsson, M., Soli, S.D., Sullivan, J.A. 1994. Development of the hearing in noise test for the measurement of speech recognition thresholds in quiet and in noise. The Journal of the Acoustical Society of America, 95(2): 1085–1099.

Roman, N., Woodruff, J. 2011. Intelligibility of reverberant noisy speech with ideal binary masking. The Journal of the Acoustical Society of America, 130(4): 2153–2161. doi: 10.1121/1.3631668.

Roman, N., Woodruff, J. 2013. Speech intelligibility in reverberation with ideal binary masking: effects of early reflections and signal-to-noise ratio threshold. The Journal of the Acoustical Society of America, 133(3): 1707–1717. doi: 10.1121/1.4759883.

Shahidi, L.K., Collins, L.M., & Mainsah, B.O. (2021). A comparison of real-time feasible mask estimation models for reverberant artifact removal in cochlear implant pulse trains. Conference on Implantable Auditory Prostheses, Online/virtual conference.

Strydom, T., Hanekom, J.J. 2011. The performance of different synthesis signals in acoustic models of cochlear implants. The Journal of the Acoustical Society of America, 129(2): 920–933. doi: 10.1121/1.3518760.

Studebaker, G.A. 1985. A “rationalized” arcane transform. Journal of Speech and Hearing Research, 28: 455–462.

Swanson, B., Mauch, H. 2006. EI0511DD: Nucleus MATLAB Toolbox 4.20 Software User Manual (4.20). Article 4.20.

Wang, D. 2005. On ideal binary mask as the computational goal of auditory scene analysis. In: P. Delorme, (ed.) Speech perception by humans and machines, p. 181–197. New York, NY: Springer US.

Wang, D. 2008. Time-frequency masking for speech separation and its potential for hearing aid design. Trends in Amplification, 12(4): 332–353. doi: 10.1177/1087138808326455.

Wang, D., Kjems, U., Pedersen, M.S., Boldt, J.B., Lunner, T. 2009. Speech intelligibility in background noise with ideal binary time-frequency masking. Journal of the Acoustical Society of America, 125(4): 2346–2347. doi: 10.1121/1.3083233.

Whitmal, N.A., Poissant, S.F., Freyman, R.L., Helfer, K.S. 2007. Speech intelligibility in cochlear implant simulations: effects of carrier type, interfering noise, and subject experience. The Journal of the Acoustical Society of America, 122(4): 2376–2388. doi: 10.1121/1.2773993.

Zhao, Y., Wang, D., Johnson, E.M., Healy, E.W. 2018. A deep learning based segregation algorithm to increase speech intelligibility in hearing-impaired listeners in reverberant conditions. The Journal of the Acoustical Society of America, 143(3): 1717–1727. doi: 10.1121/1.4974011.

Zhao, Y., Wang, D., Merks, I., & Zhang, T. (2016). DNN-based enhancement of noisy and reverberant speech. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6525–6529), Shanghai, China. doi: 10.1109/ICASSP.2016.7472934.

Zhao, Y., Wang, Z-Q., & Wang, D. (2017). A two-stage algorithm for noisy and reverberant speech enhancement. International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (pp. 5580–5584), New Orleans, LA.

Zhao, Y., Xu, B., Giri, R., & Zhang, T. (2018). Perceptually guided speech enhancement using deep neural networks. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5074–5078), Calgary, AB.