



Published in final edited form as:

Proc Int Conf Mach Learn Appl. 2018 December ; 2018: 847–852. doi:10.1109/ICMLA.2018.00136.

Application of a Graphical Model to Investigate the Utility of Cross-channel Information for Mitigating Reverberation in Cochlear Implants

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

Electrical and Computer Engineering, Duke University, Durham, USA

Abstract

Individuals with cochlear implants (CIs) experience more difficulty understanding speech in reverberant environments than normal hearing listeners. As a result, recent research has targeted mitigating the effects of late reverberant signal reflections in CIs by using a machine learning approach to detect and delete affected segments in the CI stimulus pattern. Previous work has trained electrode-specific classification models to mitigate late reverberant signal reflections based on features extracted from only the acoustic activity within the electrode of interest. Since adjacent CI electrodes tend to be activated concurrently during speech, we hypothesized that incorporating additional information from the other electrode channels, termed *cross-channel information*, as features could improve classification performance. Cross-channel information extracted in real-world conditions will likely contain errors that will impact classification performance. To simulate extracting cross-channel information in realistic conditions, we developed a graphical model based on the Ising model to systematically introduce errors to specific types of cross-channel information. The Ising-like model allows us to add errors while maintaining the important geometric information contained in cross-channel information, which is due to the spectro-temporal structure of speech. Results suggest the potential utility of leveraging cross-channel information to improve the performance of the reverberation mitigation algorithm from the baseline channel-based features, even when the cross-channel information contains errors.

Keywords

Reverberation; Cochlear implants; Binary mask; Graphical model; Ising model

I. INTRODUCTION

Cochlear implants (CIs) enable individuals with severe deafness to regain functional levels of sound perception and speech intelligibility. In traditional signal processing algorithms for CIs, an acoustic signal is quantized into frequency bands that are mapped to electrode locations in the implant. The frequency-banded signals are converted to a temporal pattern for stimulating the electrodes implanted in the cochlea [1]. While individuals with CIs can generally understand speech well in quiet, their speech comprehension is degraded in

challenging listening environments due to the limited temporal and spectral information available to the CI user. This limited spectro-temporal resolution is due to the fixed number of electrodes in the implant and the limited temporal resolution achieved with implant stimulation.

Reverberant environments are one type of challenging listening environment in which CI users experience degraded listening performance. Reverberation is the reflection of acoustic waves in an enclosed environment. Reverberant signal reflections degrade the temporal structure of speech by altering the amplitude modulations of the signal, and thus altering the fundamental frequency, an important cue for speech comprehension [2]. For example, the structure of the speech waveform shown in Figure 1(a) is greatly distorted with the addition of reverberation, as shown in Figure 1(b). Reverberation also degrades the spectral structure of speech, mainly by distorting spectral cues used to distinguish speech phonemes and phoneme transitions. For example, in Figure 2, the unvoiced consonant *t* at 730 ms has been obscured by additional reverberant energy from the previous phoneme.

Due to the already limited resolution in time and frequency available to CI users, speech intelligibility is greatly reduced for CI users in reverberant conditions when compared to individuals with normal hearing [3]. To mitigate the effects of reverberation in CI users, the typical approach is to remove portions of the CI stimulus pattern that are estimated to be dominated by reverberation [4], [5]. This type of speech enhancement algorithm is implemented by performing an element-wise multiplication of the CI stimulus pattern with a binary mask. Let $BM(f, t)$ denote a binary mask element for the time-frequency (TF) unit for frequency channel f at time index t . Features are extracted from the acoustic signal at each TF unit and thresholded to retain ($BM(f, t) = 1$) or discard ($BM(f, t) = 0$) TF units based on a criterion that quantifies the amount of reverberation in a given unit.

One promising reverberation mitigation strategy developed by Desmond [6] uses a machine learning approach to generate binary masks. The binary mask distinguishes between two types of active TF units based on the effects of early and late reverberant signal reflections. The first type of TF unit, *self-masking*, is characterized by early reverberant signal reflections that interact with a speech phoneme. The second type, *overlap-masking*, is characterized by late signal reflections that are sustained after termination of a phoneme, which leads to echoes of a preceding phoneme overlapping with that of the subsequent phoneme. Inactive TF units, referred to as *quiet* units, have zero amplitude, and so their mask values are negligible. Figure 2 shows labeled self-masking and overlap-masking TF units of a reverberant signal after CI signal processing. Electrode-specific classifier models are trained on causal features to distinguish between self- and overlap-TF units at each electrode, and units classified as overlap-masking are deleted from the CI stimulus pulse train [7].

Desmond's approach to mitigating reverberation in CIs relies on extracting *channel-based* features for classification using information capturing the activity only within the TF units in the electrode channel of interest. Since the TF units of adjacent electrode channels tend to be activated at the same time, they are more likely to enter the self- or overlap-masking state concurrently. Figure 2 shows the concurrent activation of self- and overlap-masking states

across electrode channels in a reverberant cochlear implant stimulus pattern. Hence, we hypothesize that information about the state of TF units at the other electrode locations at a given time point, termed *cross-channel information*, might be useful for classification when determining the state of a given electrode TF unit. The TF units across all electrodes in the CI stimulus pattern of a reverberant signal can be considered as a *reverberant mask* where each TF unit can exist in one of three states: the two states describing the effect of reverberation, self- and overlap-masking, and one state for quiet. In this work, we explore the utility of cross-channel features extracted from the reverberant mask to improve classification performance in the reverberation mitigation algorithm. Also, since reverberant masks estimated in a real-world scenario will likely contain errors, we analyze the classifier robustness using cross-channel features extracted from reverberant masks with errors. These errors should be clustered in time and frequency, since errors observed when estimating reverberant masks using Desmond's channel-based algorithm typically co-occur in spectro-temporal space.

Due to the spectro-temporal structure of speech, a suitable framework is needed to enforce structured errors in reverberant masks. This framework should model the probabilistic dependence of each TF unit and incorporate the geometric configuration of reverberant masking states that occur in speech. The Ising model is a two-state graphical model that is well suited for modelling pairwise interactions of nearest neighbors between variables that have a specific geometric relationship. Variants of the Ising model have been used in a variety of applications, such as image denoising [8], modeling cancer cell growth [9], and quantifying pairwise interaction strengths in neural populations [10]. Previous work by Kressner *et al.*, 2015 utilized an Ising model to systematically add structured error to binary masks to investigate the robustness of a CI speech enhancement algorithm in a noisy acoustic environment [11]. We extend the Ising model framework to develop a three-state model to systematically add errors to reverberant masks.

This paper is organized as follows. Section II outlines our graphical model for introducing structured errors in CI reverberant masks and describes the feature sets used to analyze the utility of cross-channel information. We present results in Section III that demonstrate the utility of cross-channel features from the reverberant masks to the reverberant speech enhancement algorithm, with reverberant mask errors introduced using our graphical model to assess algorithm robustness. Section IV outlines conclusions and suggestions for future work.

II. METHODS

A. Graphical Model for Adding Structured Errors to Reverberant Masks

Given a map of the true reverberant state labels, referred to as the *ideal reverberant mask*, we would like to generate a *non-ideal reverberant mask* containing a desired amount of self-masking and overlap-masking errors. For our application, we want to add overlap-masking error to self-masking states and self-masking error to overlap-masking states, since the mitigation strategy distinguishes between only these two states. Also, we want to avoid adding quiet errors to reverberant states or errors to quiet states, as it is easy to differentiate between quiet states (inactive TF units) and self- or overlap-masking states (active TF units).

Although we are not adding error to quiet states, quiet states must be included in the masks to maintain the spectro-temporal structure of self-masking and overlap-masking TF units within a speech utterance. To enforce these structured errors in a reverberant mask, we use a variant of the Ising model as it captures information about the relative state of each TF unit and the influence of nearest neighbors.

1) Model Description: The model used in this work is illustrated graphically in Figure 3, which shows the relationship between the ideal reverberant mask and the non-ideal reverberant mask. Each node in the graphical model represents a TF unit in the cochlear implant stimulus pattern of an individual utterance or speech token. Each node can either be in the overlap-masking, self-masking, or quiet state, indicated by the state values 1, 2, or 3, respectively. Nodes in the ideal reverberant mask are represented by $y = \{y_1, \dots, y_N\}$, while nodes in the non-ideal reverberant mask are represented by $x = \{x_1, \dots, x_N\}$, where x_i and y_i are nodes in ideal and non-ideal reverberant mask, respectively, and N is the total number of nodes in a reverberant mask. Edges between nodes illustrate pairwise interactions between nearest neighbors in x (x_i and x_j , where $i \neq j$) and between corresponding nodes in x and y (x_i and y_j).

A set of parameters, $\{h_1, h_2, h_3$ and $\eta\}$, are used to control how the non-ideal reverberant mask is generated from the ideal reverberant mask. State transition probabilities, $p(x_i|y_i)$, are determined by the parameters h_1, h_2 , and h_3 , which control the addition of error to a node in the non-ideal reverberant mask, x_i , when the corresponding node in the ideal reverberant mask, y_i , is in the overlap-masking, self-masking, or quiet state, respectively. The *state transition constants*, h_k , where k denotes one of the three states, can take on any real value. For example, if the $h_1 = h_2 = h_3 = 1$, then all nodes in y will exert equal influence on the nodes in x and thus the non-ideal reverberant mask is likely to be identical to the ideal reverberant mask. Within the non-ideal reverberant mask, the influence of the state of neighboring nodes on the state of the current node is controlled by the *coupling constant*, η , which can take any positive value or zero. For example, if $\eta = 0$, then neighboring nodes have no impact on the state of the current node and uniform errors (in amounts dictated by h_1, h_2 , and h_3) result. If $\eta > 0$, a greater weight is placed on edges between nodes in the same state, encouraging clusters of similar states to appear in the non-ideal reverberant mask.

This model can be described as a Markov Random Field with the following joint distribution over a pair of ideal and non-ideal reverberant masks:

$$p(x, y) = \frac{1}{Z} \prod_{\langle i, j \rangle} \psi_{ij}(x_i, x_j) \prod_{\langle i, i \rangle} \psi_{ii}(x_i, y_j) \prod_i \psi_i(x_i) \quad (1)$$

where $\langle i, j \rangle$ indicates all edges in the graph; ψ_i is a node potential, which dictates the relative probability of a node taking on each state; ψ_{ij} is an edge potential, which dictates the relative probability for each type of state transition; and Z is the partition function which normalizes the distribution. For this application, the node and edge potentials are specified as follows:

$$\begin{aligned}\psi_i(x_i|y_i = 1) &= \begin{bmatrix} e^{h_1} & 1 & 1 \end{bmatrix} \\ \psi_i(x_i|y_i = 2) &= \begin{bmatrix} 1 & e^{h_2} & 1 \end{bmatrix} \\ \psi_i(x_i|y_i = 3) &= \begin{bmatrix} 1 & 1 & e^{h_3} \end{bmatrix}\end{aligned}\quad (2)$$

$$\psi_{ij}(x_i, x_j) = \psi_{ii}(x_i, y_i) = \begin{bmatrix} e^\eta & 1 - e^\eta & 0 \\ 1 - e^\eta & e^\eta & 0 \\ 0 & 0 & 1 \end{bmatrix}\quad (3)$$

Each node potential vector in 2 specifies the potentials for nodes whose ideal counterparts are in the overlap-masking ($y_i = 1$), self-masking ($y_i = 2$), or quiet state ($y_i = 3$). Using three node potential functions allows us to vary the influence of a node in the ideal mask on the corresponding node in the non-ideal mask depending on the state of the node in the ideal mask. Each row in the edge potential matrix in 3 specifies the relative probability assigned to each state (given by the columns) depending on the state of the neighboring node (given by the rows). The form of the edge potential used in this work allows transitions to and from the self- and overlap-masking states while disallowing nodes in the quiet state from transitioning to other states (zero entries in the matrix). This enables the model to only add self- and overlap-masking errors while maintaining the TF structure provided by quiet nodes.

To generate a non-ideal reverberant mask containing the desired amounts of self- and overlap-masking errors, a sample is drawn based on the joint distribution 1. Due to the intractability of computing the joint probability over a large number of nodes (since the number of possible state configurations described in the partition function grows with 3^N), an approximate sampling algorithm, such as a Markov Chain Monte Carlo (MCMC) sampling technique, is used. We use the toolbox developed by Schmidt [12] to implement and sample from our graphical model. The desired error amounts for each state transition are obtained by altering the value of the respective leading state transition constant of each summation in the energy function (either h_1 , h_2 , or h_3). Given an ideal reverberant mask, the joint distribution is sampled accordingly based on a specific set of model parameters to generate non-ideal reverberant masks. Since the state transition constants, h_k , can take any real value, a look-up table is necessary to map from desired error amounts to the appropriate state transition constants.

2) Model Validation: To determine the desired state transition constants to use when sampling the joint distribution, the constant controlling the coupling of non-ideal mask neighbors (η) is fixed and the state transition constants (h_k) are varied. Preliminary analysis revealed that a coupling constant of 1 gave the smoothest error addition function when state transition constants were varied. As an illustration, Figure 4 shows self- and overlap-masking error rates in a single reverberant mask over a range of h_k values using a coupling constant of -1 , 1 , or 2 . The functional form of the self- and overlap-masking error rates when using an η of 1 suggests a simple sigmoidal functional mapping between desired error rates and h_k values.

To create a look-up table to implement the desired error amounts in self-masking and overlap-masking units, the state transition constants h_1 and h_2 were varied in the simulations, and the amounts of error added were computed by comparing the ideal mask to the non-ideal mask. The value of h_3 had a negligible effect on self- and overlap-masking error amounts due to the form of the edge potential used in this work to avoid the addition of quiet errors to self- and overlap-masking states. For computational reasons, look-up table values were obtained from observing error addition in a 5% subset of all reverberant masks created from individual speech tokens from a given speech database.

To ensure that our model adds the desired error amounts to the ideal reverberant mask given the values determined for the look-up table, we simulated error addition in reverberant masks over a range of state transition constants and determined the actual error added to reverberant masks. Ideal reverberant masks were obtained from all speech tokens in two separate speech corpora, the TIMIT speech corpus [13] and the HINT speech corpus [14]. Each speech corpus uses different speakers and sentence material, and thus will contain different structure of self- and overlap-masking TF units. Since the effect of the state transition constants depends on the structure of self- and overlap-masking units present in each reverberant mask, we analyzed the two speech corpora separately.

Figure 5(a) and (b) shows a comparison of predicted and actual error rates of reverberant masks obtained from both speech databases. Averaged over all masks, the model behaves as desired, with the average actual error similar to the predicted error, although there is some bias and deviations in the amount of error added at higher overlap-masking error rates.

B. Features for Analysis

The approach taken by Desmond [6] for mitigating reverberation in CIs used features that captured trends in signal energy and degradation in a single electrode channel. For each electrode channel, causal pulse-based features are extracted from the CI stimulus pattern. The channel-based feature vector contains features describing pulse amplitude, pulse amplitude difference, window energy, window energy difference, and variance of pulse amplitudes within a window. The windowed features were calculated from 30 ms windows of the CI stimulus pattern.

We extended the work by Desmond [6] for mitigating reverberation in cochlear implants by incorporating information from the reverberant mask into the electrode-specific models used for classification. Let $x_{f,t}$ represent a vector of the channel-based features developed by Desmond for frequency channel f at time t . We augmented the channel-based features with information about the state of other channels at a given time step accordingly:

$$c_{f,t} = [x_{f,t}, l_t \mathcal{V}_{f,t}] \quad (4)$$

where $c_{f,t}$ is the new composite feature vector; and $l_t \mathcal{V}_{f,t}$ represents an ordered list of reverberant state labels from the reverberant mask for all other frequency channels at time t , excluding the state label for channel f . The reverberant state labels are either extracted from ideal reverberant masks, and so reflect the true reverberant state labels of all frequency channels at that time step, or are extracted from non-ideal reverberant masks created using

the model described in the previous section, and so contain errors with spectro-temporal structure.

Either the channel-based feature vectors or the composite feature vectors are scored with an electrode-specific classifier. Relevance vector machine (RVM) classification models with radial basis function kernels are used as our electrode-specific classifiers. A binary mask is implemented by comparing the resulting classifier scores to channel-specific thresholds accordingly:

$$BM(f, t) = \begin{cases} 1, & \text{if } s(f, t) < \tau_f \\ 0, & \text{if } s(f, t) \geq \tau_f \end{cases} \quad (5)$$

where $BM(f, t)$ is the binary mask decision for an electrode channel f at time index t ; $s(f, t)$ is the channel-specific classifier score; and τ_f is the channel-specific classifier threshold value associated with a specified probability of detecting overlap-masking units in a given channel. The classifier threshold value, τ_f , is chosen to maximize the removal of overlap-masking TF units while minimizing deletions of self-masking TF units.

III. RESULTS

We tested the performance of each electrode-specific classification model when given channel-based features only and composite features with information obtained either from an ideal reverberant mask or a non-ideal reverberant mask created using the model described in the previous section. Simulated reverberation was added to speech tokens [15] and channel-based features and ideal reverberant masks were extracted from the reverberant speech tokens. Then, errors were added to the ideal reverberant masks using the graphical model described in Section II-A, to generate non-ideal reverberant masks. Classification models were trained on 2000 feature vectors obtained from a random subset of the TIMIT speech corpus [13] and tested on 1000 feature vectors obtained from a random subset of the HINT speech corpus [14]. To evaluate the robustness of classifier performance to errors in cross-channel information, errors in increasing amounts were added to self- and overlap-masking units in each reverberant mask. Classifier performance was evaluated using the area under the receiver operating characteristic curve, AUC. The AUC metric allows us to compare classifier performance regardless of the classifier operating point (τ_f) chosen to create the binary mask.

Figure 6 shows classifier performance with channel-based features only, composite features with ideal reverberant state labels, and composite features with non-ideal reverberant state labels with varying combinations of self- and overlap-masking errors added to reverberant masks. Features with ideal reverberant state labels significantly improved classifier performance across all electrodes compared to channel-based features. There was a significant effect of feature set employed in each classification algorithm, as determined by a one-way repeated measures analysis of variance ($p < 0.005$). Performance improvements were also observed with features with non-ideal state labels across most of the electrode locations. Based on similarity in performance trends, the electrodes can be grouped by

frequency ranges: electrodes 1–7, 8–16, and 17–22, for high-, mid-, and low-frequency channels, respectively. When non-ideal reverberant state labels were used, classifier performance improvements were observed in the high- and low-frequency channels, while minimal changes in performance were observed in the mid-frequency channels.

Figure 7 shows classifier performance for different combinations of self- and overlap-masking error rates at representative electrodes locations that summarize performance trends in the each of three frequency ranges. Classifier performance is given as the difference between the AUC resulting from only channel-based features and the AUC resulting from additional reverberant mask features containing errors. Figure 7(a) shows results in the mid-frequency channels, with minimal changes in classifier performance observed as the state error combination rates varied. These results further support the previous finding that information of the state of other channels provides minimal benefit to overlap-masking classification in the mid-frequency channels. Figure 7(b) and 7(c) show results in the low- and high-frequency channels, respectively. Initially, with increasing amounts of self- and overlap-masking errors, a decrease in classifier performance can be observed. As the amount of added error further increased, an improvement in classifier performance is observed. While it might be counter-intuitive that higher amounts of errors in the reverberant masks confers information beneficial for overlap-masking classification, this reversal in performance trends is due to the spectro-temporal structure of reverberant speech. When high amounts of error are added to the reverberant mask, errors in the overlap-masking units become saturated, creating dense regions of self-masking units resembling the spectro-temporal structure normally observed in reverberant speech. The errors in the self-masking units do not attain this same level of saturation, due to the model's limited range of error addition demonstrated in Figure 4. The sparsely distributed self-masking errors and the dense overlap-masking errors create an effective error rate much lower than the implemented error rate, yielding improved classifier performance at high error rates. Overall, the low-frequency channels are more robust to errors than the high-frequency channels, evident by the slower change in gradient in the AUC performance map in Figure 7(b) when compared to Figure 7(c), respectively.

IV. CONCLUSIONS AND FUTURE WORK

We investigated the utility of cross-channel state information extracted from reverberant masks to improve the performance of a CI reverberant mitigation algorithm. To generate realistic mask estimation errors based on speech structure, we adapted the Ising model to systematically add structured errors to reverberant masks and validated our new graphical model framework using data for commonly used speech databases in the literature. Results show that features with additional cross-channel state information, even when containing errors, have the potential to improve the performance of the reverberation mitigation algorithm over the baseline of channel-based features. A sensitivity analysis of the robustness of the algorithm to reverberant mask errors revealed that the benefits conferred with additional cross-channel state information are frequency- and state-dependent. Future work includes conducting speech recognition studies to investigate the most effective strategy for incorporating cross-channel information to the CI reverberation mitigation algorithm to yield improvements in speech intelligibility.

Acknowledgments

This work was sponsored by the National Institute of Health, via a grant administered by the National Institute on Deafness and Other Communication Disorders (R01-DC014290-03).

REFERENCES

- [1]. Loizou PC, "Signal-processing techniques for cochlear implants," IEEE Eng Med Biol Mag, vol. 18(3), pp. 34–46, 1999. [PubMed: 10337562]
- [2]. Nabelek AK, Letowski TR, and Tucker FM, "Reverberant overlap-masking and self-masking in consonant identification," Journal of the Acoustical Society of America, vol. 86, no. 4, pp. 1259–1265, 1989. [PubMed: 2808901]
- [3]. Kokkinakis K and Loizou PC, "The impact of reverberant self-masking and overlap-masking effects on speech intelligibility by cochlear implant listeners," J Acoust Soc Am, vol. 130(3), pp. 1099–1102, 2011. [PubMed: 21895052]
- [4]. Hazrati O, Lee J, and Loizou PC, "Blind binary masking for reverberation suppression in cochlear implants," J Acoust Soc Am, vol. 133(3), pp. 1607–1614, 2013. [PubMed: 23464030]
- [5]. Hazrati O and Loizou PC, "Reverberation suppression in cochlear implants using a blind channel-selection strategy," J Acoust Soc Am, vol. 133(6), pp. 4188–4196, 2013. [PubMed: 23742370]
- [6]. Desmond JM, "Using Channel-Specific Models to Detect and Mitigate Reverberation in Cochlear Implants," Ph.D. dissertation, Duke University, 2014.
- [7]. Desmond JM, Collins LM, and Throckmorton CS, "The effects of reverberant self- and overlap-masking on speech recognition in cochlear implant listeners," J Acoust Soc Am, vol. 135, no. 6, pp. EL304–10, 2014. [PubMed: 24907838]
- [8]. Besag J, "On the statistical analysis of dirty pictures," J R Stat Soc, vol. 48(3), pp. 259–302, 1986.
- [9]. Sun L, Chang YF, and Cai X, "A discrete simulation of tumor growth concerning nutrient influence," Int J Mod Phys B, vol. 18(17–19), pp. 2651–2657, 2004.
- [10]. Schneidman E, Berry MJ II, Segev R, and Bialek W, "Weak pairwise correlations imply strongly correlated network states in a neural population," Nature, vol. 440(20), pp. 1007–1012, 2006. [PubMed: 16625187]
- [11]. Kressner AA and Rozell CJ, "Structure in time-frequency binary masking errors and its impact on speech intelligibility," J Acoust Soc Am, vol. 137(4), pp. 2025–2035, 2015. [PubMed: 25920853]
- [12]. Schmidt M. (2007) UGM: A Matlab toolbox for probabilistic undirected graphical models. [Online]. Available: <https://www.cs.ubc.ca/schmidtm/Software/UGM.html>
- [13]. Victor Z, Seneff S, and Glass J, "TIMIT acoustic-phonetic continuous speech corpus," Speech Commun, vol. 9(4), pp. 351–356, 1990.
- [14]. Nilsson M, Soli SD, and Sullivan JA, "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J Acoust Soc Am, vol. 95(2), pp. 1085–1099, 1994. [PubMed: 8132902]
- [15]. Lehman EA and Johansson AM, "Prediction of energy decay in room impulse responses simulated with an image-source model," J Acoust Soc Am, vol. 124(1), pp. 269–277, 2008. [PubMed: 18646975]

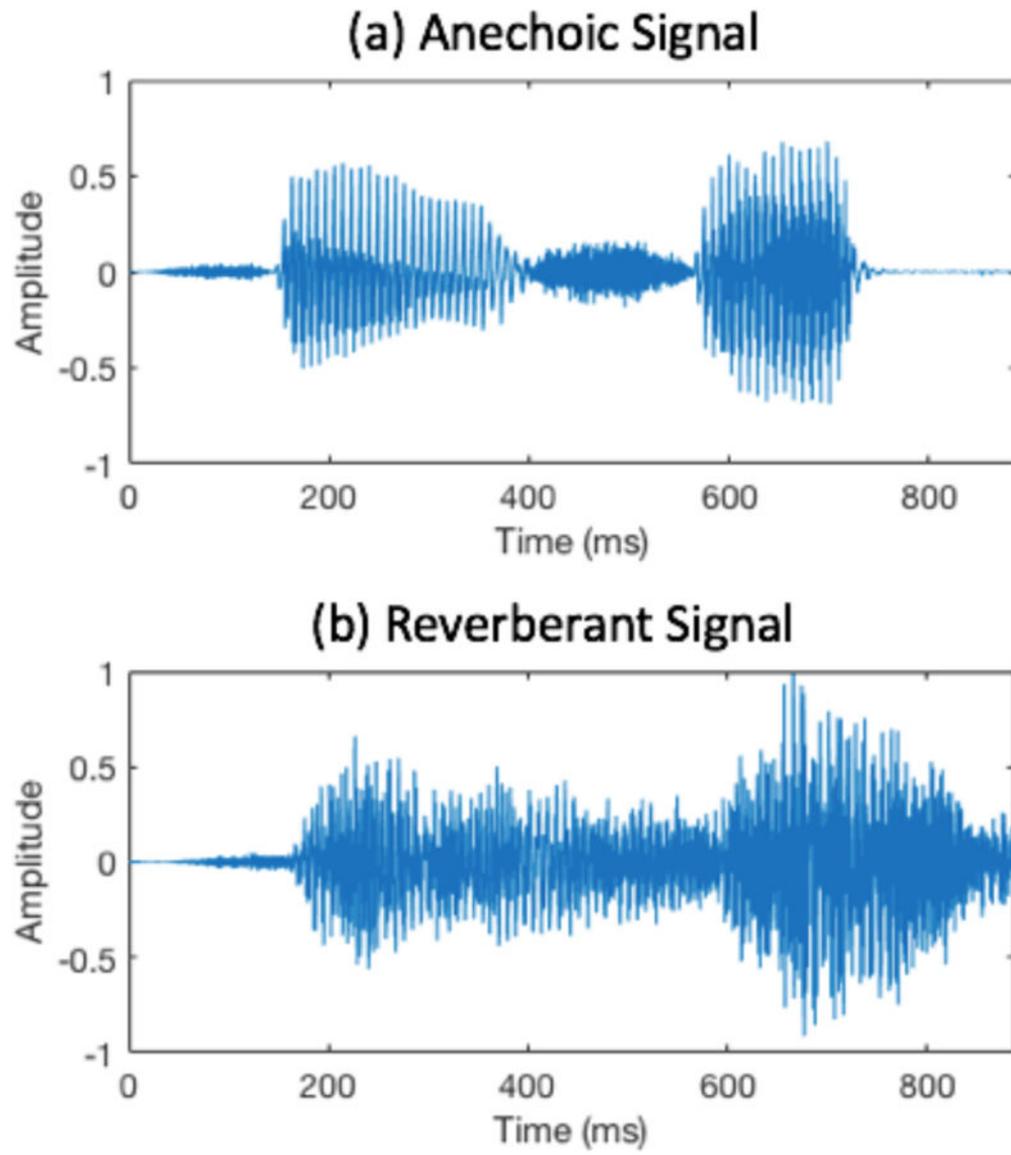


Fig. 1: The acoustic waveform for the utterance *sunset* in (a) anechoic and (b) reverberant conditions (with a reverberation time of 1.5 seconds).

Cochlear Implant Stimulus Pattern of a Reverberant Signal

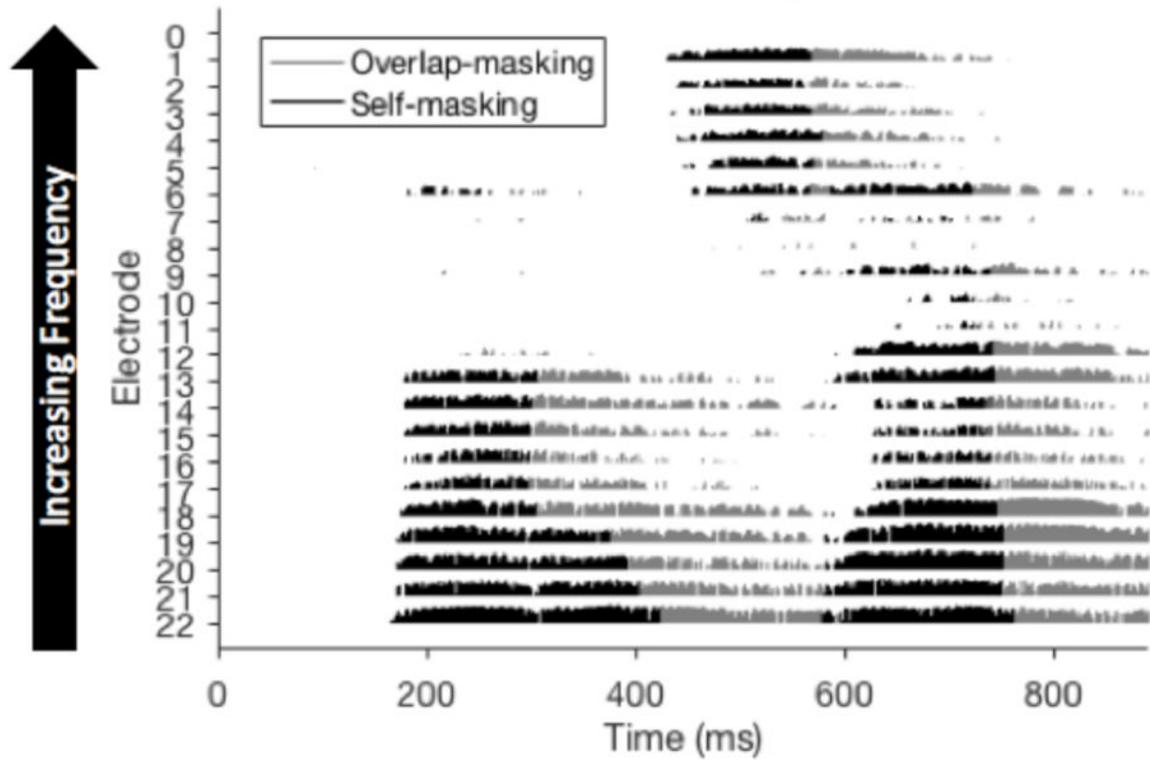


Fig. 2: The cochlear implant stimulus pattern for the utterance *sunset* in reverberant conditions.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

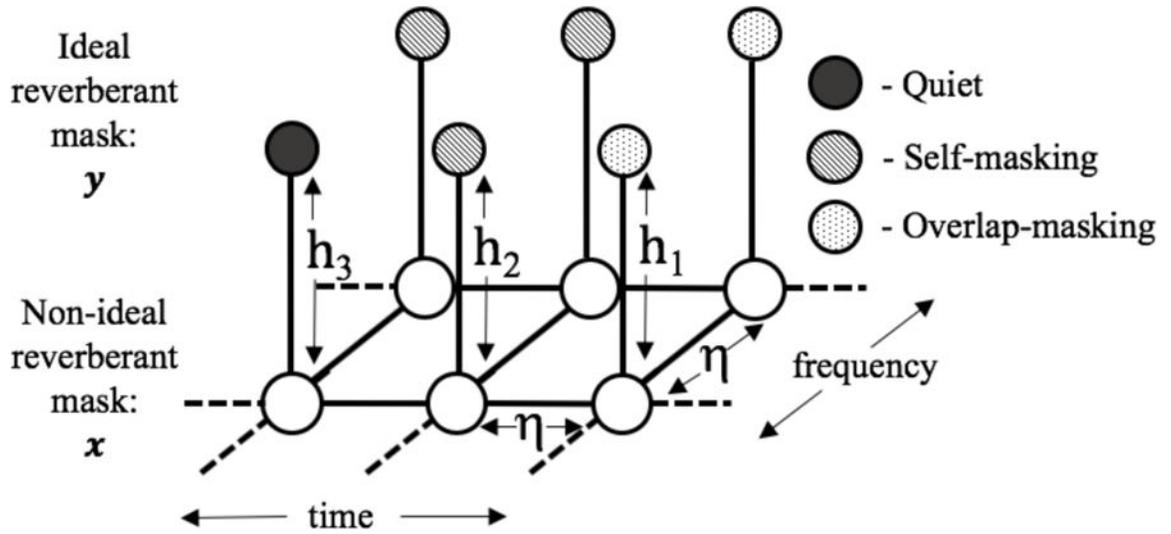


Fig. 3: Graphical model representing the relationship between an ideal reverberant mask (top unconnected layer) and the corresponding non-ideal reverberant mask (bottom connected layer). A node represents the state of a TF unit in a cochlear implant stimulus pattern, where a unit can exist in a quiet state, or one of two reverberant states, self- or overlap-masking. State-specific error probabilities are controlled by state transition parameter constants, h_1 , h_2 , and h_3 , for overlap-masking, self-masking, and quiet states, respectively. The clustering of labels across neighboring states is controlled by the coupling constant η .

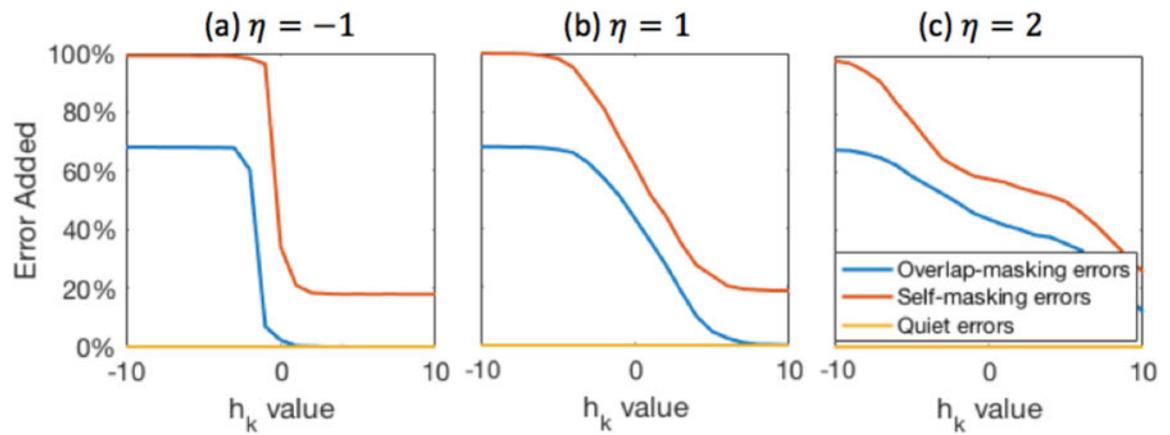
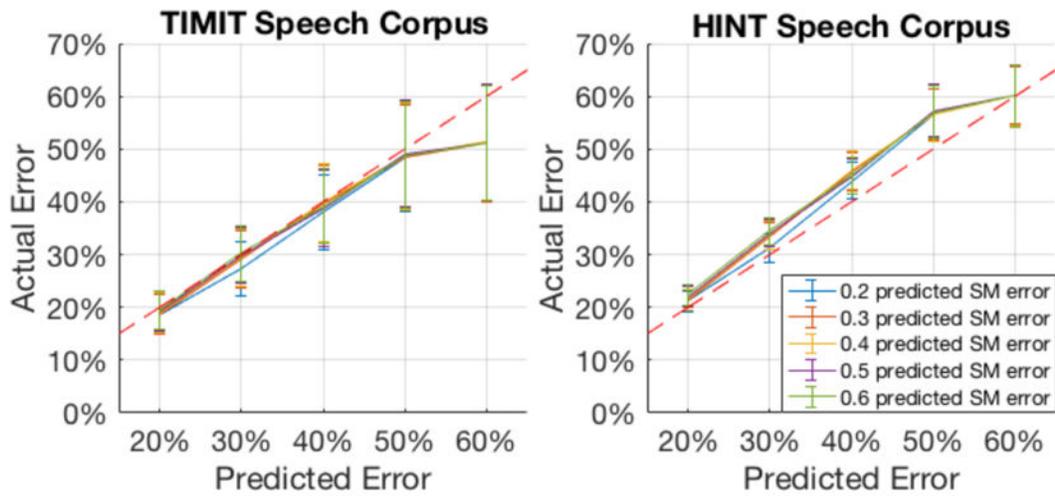
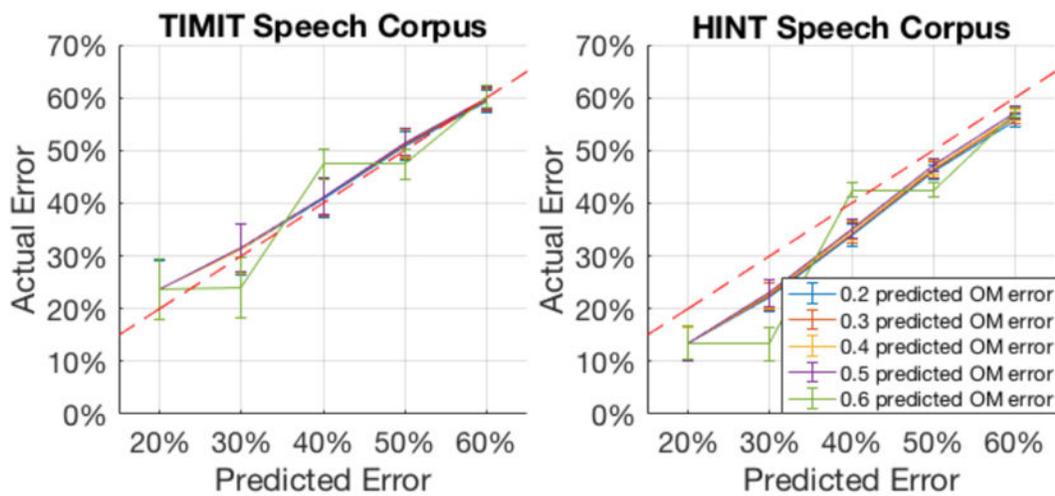


Fig. 4:

The amount of state-specific errors added to the non-ideal reverberant mask of a single speech token over a range of state transition constants, h_k , using (a) $\eta = -1$, (b) $\eta = 1$, and (c) $\eta = 2$. Setting η to a single value facilitates a direct mapping from desired error amount to h_k .



(a) Error added to overlap-masking units.



(b) Error added to self-masking units.

Fig. 5:

Plots of actual added versus (a) overlap- and (b) self-masking errors in non-ideal reverberant masks created from all speech tokens in two speech databases (the TIMIT speech corpus [13] and HINT speech corpus [14]). Each line represents a single state transition constant, h_1 or h_2 , used to determine self-masking (SM) of overlap-masking (OM) error added to reverberant masks, respectively. The ideal mapping of predicted to actual error is given by the dashed red line.

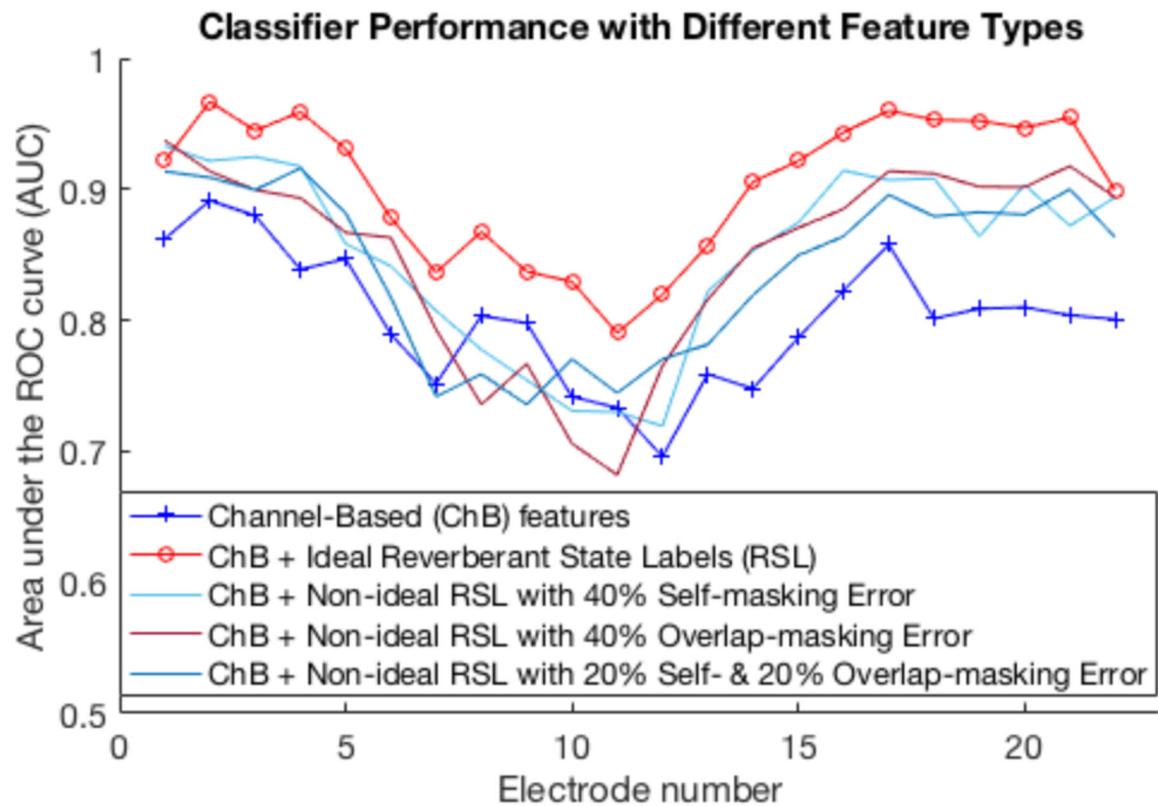


Fig. 6:

Performance of channel-specific classification models when given ideal or non-ideal reverberant state labels as additional features. Classification performance given only channel-based (ChB) features is provided as a baseline.

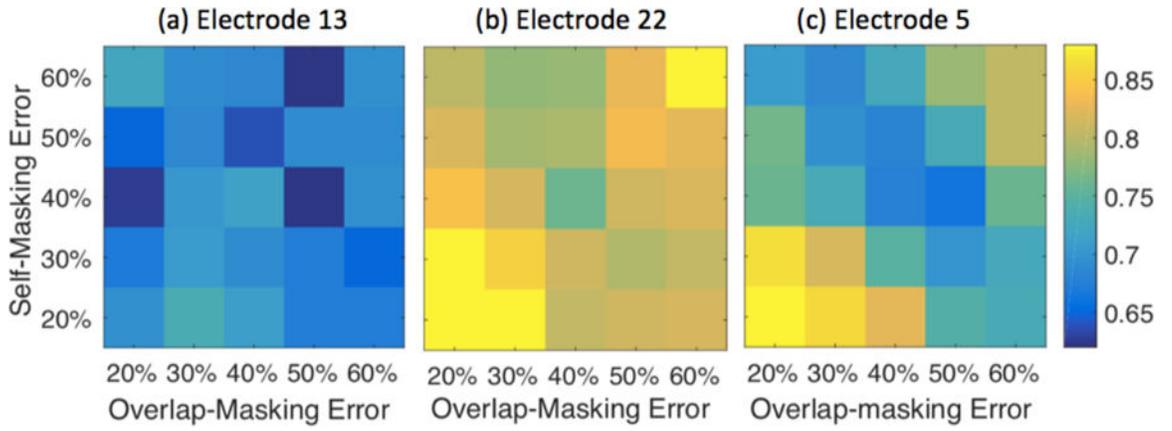


Fig. 7: Classifier performance measured as the difference between the area under the ROC curve (AUC) when classifiers were given channel-based features only and when given a range of self-masking and overlap-masking errors added to reverberant masks in addition to channel-based features. Classifier performance for frequency-time units in frequency channel 13, 22, and 5, are given in the left, middle, and right plots respectively, as a representation of results from mid-, low- and high-frequency channels.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript