Modeling the Medial Olivocochlear Efferent in the Descending Auditory Pathway with a Dynamic Gain Control Feedback System

by

Afagh Farhadi

Submitted in Partial Fulfillment of the

Requirements of the Degree

Doctor of Philosophy

Supervised by

Professor Laurel H. Carney

Department of Electrical and Computer Engineering
Arts, Sciences and Engineering
Edmund A. Hajim School of Engineering and Applied Sciences

University of Rochester
Rochester, New York

2023

Table of Contents

List of Tables

List of Figures

Biographical Sketch

Afagh Farhadi was born in Iran. She attended Iran University of Science and Technology, and graduated with the Bachelor of Science degree in Electrical Engineering in 2015. She began her doctoral studies in Electrical and Computer Engineering at University of Rochester in 2016. She received the Master of Science degree in Electrical and Computer Engineering from University of Rochester in 2018. She pursued her research in computational modeling of the efferent system in the descending auditory pathway under the direction of Dr. Laurel H. Carney.

The following publications were a result of work conducted during doctoral study:

Journal Publications:

Farhadi, Afagh, Jennings, S. G., Strickland, E. A., & Carney, L. H. "Subcortical Auditory Model including Efferent Dynamic Gain Control with Inputs from Cochlear Nucleus and Inferior Colliculus." Submitted to Journal of the Acoustical Society of America, Under Revision.

Brennan, M. A., Svec, A., Farhadi, A., Maxwell, B. N., & Carney, L. H. (2023). Inherent envelope fluctuations in forward masking: Effects of age and hearing loss. The Journal of the Acoustical Society of America, 153(4), 1994-1994.

Hamza, Y., Farhadi, A., Schwarz, D. M., McDonough, J. M., & Carney, L. H. (2023). Representations of fricatives in sub-cortical model responses: comparisons with human consonant perception. (Accepted to Journal of the Acoustical Society of America).

Conference Publications:

Farhadi, A., Jennings, S. G., Strickland, E. A., & Carney, L. H. (2021, June). A Closed-Loop Gain-Control feedback model for the medial efferent system of the descending auditory pathway. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 291-295). IEEE.

Farhadi, A., & Carney, L. H. Predicting thresholds in an auditory overshoot paradigm using a computational subcortical model with efferent feedback. accepted in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 2023 (WASPAA2023).

Farhadi, A., Agarwalla, S., & Carney, L. H. (2023). A Subcortical Auditory Model With Efferent Gain Control Explains Perceptual Enhancement. Abstract accepted for presentation at the 46th Midwinter Meeting of the Association for Research in Otolaryngology, Orlando, FL, February 2023. Presented as a poster presentation.

Maxwell, B. N., Farhadi, A., Brennan, M. A., Svec, A., & Carney, L. H. (2023). Simulating Physiological and Psychoacoustic Forward Masking in a Subcortical Model with Efferent Gain Control. Abstract accepted for presentation at the 46th Midwinter Meeting of the Association for Research in Otolaryngology, Orlando, FL, February 2023. Presented as a poster presentation.

Carney, L. H., Farhadi, A., & Maxwell, B. N. (2022). Pitch Perception of Complex Tones: Predictions based on the Neural Fluctuation-Place Model. Abstract accepted for presentation at the 45th Midwinter Meeting of the Association for Research in Otolaryngology, held virtually, February 2022.

Acknowledgments

I would like to express my sincere gratitude to my advisor, Prof. Laurel H. Carney, for her patience and guidance throughout my PhD journey. She not only instructed me with invaluable knowledge about the fascinating world of auditory system, but also became a role model for me in both academic and personal life. Her success as a scientist and her remarkable qualities as a person have been a constant source of inspiration for me. I am truly grateful to her for making my PhD experience enjoyable and memorable. I would also like to extend my heartfelt thanks to Prof. Marc Bocko, who provided me with great support and also introduced me to my advisor. Additionally, I am grateful to Prof. Zhiyao Duan for his invaluable assistance and guidance, as what I have learned from him in his classes and during my committee sessions significantly influenced my research. Prof. Ross K. Maddox has constantly provided insightful feedback on my work in different scientific events and helped me to grow as a researcher. I am deeply grateful to my collaborators, Prof. Elizabeth A. Strickland, Prof. Skyler G. Jennings, Prof. Marc Brennan, and Prof. Adam Svec for their patience and unwavering support. They have been actively helping me overcome obstacles by providing exceptional guidance and innovative ideas. I am also grateful to the current and previous Carney lab members, Dr. Daniel Guest, Dr. Swapna Agarwalla, Dr. Braden Maxwell, Paul Mitchell, Johanna Fritzinger, Dr. Langchen Fan, Dr. Yasmeen Hamza,

Doug Schwarz, Kris Abrams whose constant support, insightful perspectives, and constructive comments contributed significantly to the development of my work throughout my research journey. Prof. Kenneth S. Henry and his graduate students, Yingxuan Wang and Leslie Gonzales have also provided me with support and valuable feedback during my PhD studies. It is impossible to mention everyone who has helped and inspired me in my research, I would like to express my gratitude to the auditory neuroscience and hearing community as a whole. Attending conferences and being a part of this wonderful scientific family has been a constant source of motivation and inspiration for me. Finally, I would like to acknowledge my family in Iran—my parents Maryam Javadi and Houshang Farhadi and my brother Ashkan Farhadi—for their patience and non-stop support despite the distance that separates us. Their support and their belief in me have been invaluable throughout these years. Lastly, I want to thank to my supportive and loving husband, Ahmad Azizimanesh, whose patience, understanding, and sacrifices helped me to overcome the challenges I faced during this journey. Thank you all for being a part of my thesis journey and for contributing to my growth as a researcher and as an individual.

Abstract

Computational modeling is a powerful tool in hearing research, as it helps to test hypotheses regarding the underlying mechanisms involved in different auditory scenes, including speech in noise, which is one of the most challenging problems in hearing studies. The medial olivocochlear (MOC) efferent system is suggested to play a crucial role in enhancing auditory processing in noisy backgrounds. The MOC system is a part of the descending auditory system that includes pathways that ultimately project to the outer hair cells (OHCs) in the cochlea, which are responsible for cochlear amplification. The MOC system adjusts the gain of the OHCs dynamically based on the characteristics of the incoming sound signal. Auditory models have mostly been focused on the ascending pathway of the auditory system, which is responsible for transmitting sensory information from the cochlea to higher areas. However, the descending pathway or efferent system, which originates from higher auditory centers and projects back to lower levels of the auditory system, has received relatively little attention. My dissertation aimed to provide a more accurate and comprehensive model of the auditory system which included the efferent system in addition to ascending pathways to better understand the effect of the MOC efferent system in auditory processing. The MOC model that was proposed in this study incorporated a feedback projection from MOC neurons which dynamically adjusted cochlear gain based on inputs received by the MOC. The two primary inputs to the MOC that were examined in the model were the excitatory projections from wide-dynamic-range cells located in the cochlear nucleus, and the fluctuation-driven information from IC cells situated in the midbrain. The model parameters were optimized using previous actual

neural recordings from IC cells in awake rabbits responding to AM noise. The model with efferents and the optimized set of parameters was successful in simulating the trend that was observed in recorded neural responses, whereas the model without efferent responses did not simulate the same trend. The optimized parameters also matched the physiological evidence for the dynamics of the MOC efferent system. The proposed model with efferents was tested using several psychoacoustical detection experiments and was found to predict human listener thresholds better than the model without efferents. For example, the model was used to predict thresholds for simultaneous masking focused on the overshoot effect, in which detecting a short tone in a longer noise masker is easier when there is a delay between the onset of the noise and the added tone. The results showed that the model with efferents was better at estimating human thresholds than the model without efferents. Specifically, the proposed model performed better with longer delay conditions, consistent with human performance, whereas the model without efferents showed the opposite trend, with worse performance for longer delays. These results demonstrate the significance of the efferent system in analyzing the mechanisms underlying various psychoacoustic phenomena, including not only simultaneous masking but also forward masking and auditory enhancement. Investigating the efferent system's influence on neural fluctuations will help reveal its potential role in speech perception. I will focus on neural fluctuations coding of speech-liked sound. The neural fluctuation coding benefits from the physiological nonlinearities in the normal auditory system at conversational levels. In contrast, level-dependent cues are not very informative at conversational levels, as the average rate of the majority of auditory-nerve responses saturate. Simulation results will

test the hypothesis that the efferent system can improve contrast in neural fluctuation profiles in response to vowel-like sounds. This result would have implications for the development of new strategies for improving speech intelligibility in noisy environments.

Contributors and Funding Sources

This work was supervised by a dissertation committee consisting of Professor Laurel H. Carney (advisor) of the Department of Biomedical Engineering and Neuroscience, Professors Mark Bocko and Zhiyao Duan of the Department of Electrical and Computer Engineering, and Professor Ross K. Maddox of the Departments of Biomedical Engineering and Neuroscience at the University of Rochester. The work described in Chapter 2 was done in collaboration with Professor Elizabeth A. Strickland from Purdue University and Professor Skyler G. Jennings from the University of Utah. Dr. Daniel R Guest, a postdoctoral fellow from University of Rochester, collaborated on the cross-frequency implementation code developed for Chapter 4. All other work conducted for the dissertation was completed by the student independently, under the supervision of Prof. Laurel H. Carney. Graduate study was supported by the National Institute of Health and National Institute on Deafness and Other Communication Disorders under grant number NIH-R01-DC010813.

Chapter 1. Introduction

Hearing speech in background noise is a very challenging task for listeners with hearing loss even with the assistance of state-of-the-art hearing aids. Listeners with healthy hearing not only outperform listeners with hearing loss, even with assistance of hearing aids, but also have better speech intelligibility in background noise than most advanced automatic-speech-recognition devices (Spille et al., 2018; Humes et al., 2013, Chou et al., 2019). These findings highlight the importance of understanding the underlying mechanism in healthy periphery auditory system for speech in background noise. The auditory processing in the healthy peripheral auditory system is doing a great job to prepare the neural response to speech in noise sound, for speech perception at the level of auditory cortex. However, studying the auditory mechanism in human listeners is challenging with noninvasive methods. We could also benefit from the results of more invasive methods to study the hearing mechanism in animals. Computational modeling can help to make a bridge between human and animal models. An accurate comprehensive computational model that can simulate the neural response to any arbitrary desired sound is a very powerful tool which can provide valuable insight into the human auditory system. Using the computational model we can test different hypotheses in hearing research and learn more about the auditory system.

In addition, hearing devices such as hearing aids, cochlear implants, automatic speech recognition (ASR) devices, and sound-localization units in AR/VR devices, can make progress benefiting from more accurate physiologically plausible auditory models.

Different studies show that the audio devices' performance improves specifically in more challenging auditory scenes such as in background noise or with reverberation when these devices use physiological inspired models (Stern et al., 2012; Stern & Morgan, 2012, Schädler et al., 2016, Meyer and Kollmeier, 2011).

The auditory system is a complex network of neural pathways that is responsible for processing sound information. However, most computational models of the auditory system only focus on the ascending pathway from the cochlea to the auditory cortex, ignoring the significant contribution of the descending pathway (Verhulst et al., 2018; Zilany et al., 2014; Zweig, 2015; Meddis et al., 2013, Meddis et al., 2001). The descending pathway is at least as large as the ascending pathway in terms of the number of projections (Schofield, 2011). The medial olivocochlear (MOC) efferent system is a subset of the descending pathway or the efferent system, which includes the projection from MOC neurons to outer hair cells adjusting the cochlear gain (Guinan, 2018). MOC system has been proposed to improve speech perception in noise (Brown et al., 2010; Yasin et al., 2020; Martes et al., 2018). The aim of this project is to develop a new computational model that is more comprehensive and includes both the ascending and descending auditory pathways by modeling the MOC efferent system to investigate the role of MOC efferent system in different auditory tasks and speech perception in background noise.

The auditory model proposed in this study includes input projections to MOC neurons from two sources: wide dynamic range (WDR) cells in the brainstem and inferior colliculus (IC) cells in the midbrain. These two feedback systems provide important and distinct information about the stimulus that MOC system uses for adjusting the cochlear gain at

different frequency channels. The WDR projections convey information about the stimulus level and function as a negative feedback gain control system. This pathway adjusts cochlear gain based on the overall level of the stimulus. If the sound is too loud or too soft, this feedback system ensures that there are variations in the neural responses across different frequency channels in response to a sound with a non-flat spectrum. To achieve this, the WDR negative feedback increases the gain for lower sound levels and decreases the gain for higher sound levels, maintaining an overall input level to the AN nerve at a moderate average rate. On the other hand, the IC midbrain cells carry more detailed information related to stimulus fluctuations. The IC projection dynamically adjusts the gain to enhance the response to stimulus peaks and reduce the gain for valleys in the frequency spectrum. This adjustment helps sharpen the response profile across the auditory-nerve population to the stimuli. The IC cells act as a positive feedback mechanism, contributing to an increased contrast in the rate profile of the auditory nerve. In summary, the model incorporates both WDR and IC feedback systems, each providing unique and essential information about the stimulus. The WDR feedback controls the overall gain based on the stimulus level, while the IC feedback enhances responses to peaks and reduces responses to valleys, increasing the contrast in the auditory nerve's rate profile.

The proposed model will provide more physiological details and specific properties of the human auditory system with healthy hearing. The long-term benefits of this model include gaining a deeper understanding of the MOC efferent system and the human auditory system in general, as well as inspiring the development of hearing aids, ASR devices, and cochlear implants that are more effective in noisy background.

1.1 Background

The role of the auditory periphery is to process the sound waveform and convert the sound waveform to neural coding suitable for perception of the sound in the brain. One interesting aspect of the auditory system is the presence of nonlinearities, which are essential for various aspects of auditory processing. In the auditory system, nonlinearities can arise at different stages of sound processing, including within the cochlea, auditory nerve, and central auditory pathways. Nonlinearities play a crucial role in enhancing auditory perception and improving the encoding of complex acoustic signals. Here we introduce a few examples of how nonlinearities in the auditory system that will shape the neural response to input sound.

The first nonlinearity discussed here is saturation of average discharge rate of auditory-nerve (AN) fibers (Liberman, 1978; Costalupes et al., 1984). Even in the absence of external stimuli, AN fibers exhibit randomly timed spikes known as spontaneous rate. As the stimulus level increases, the spiking rate of the neurons typically rises until it reaches a saturation point at the synapse between inner hair cells and the AN. AN fibers can be classified into three groups based on their spontaneous rates: high-spontaneous-rate (HSR) fibers, medium-spontaneous-rate (MSR) fibers, and low-spontaneous-rate (LSR) fibers. HSR fibers, which are the majority, project to brainstem cells and initiate the ascending pathway (Liberman, 1978, Ye et al., 2000). These different groups of AN fibers have different nonlinear response properties, with HSR fibers saturating at relatively low sound

pressure levels (SPLs). Consequently, the majority of AN fibers have saturated average rates at conversational sound levels (55-66 dB) (Olsen, 1998). This saturation leads to reduced robustness of perception cues that rely on level-dependent information at these medium sound levels.

Another nonlinearity in the auditory periphery, which is an independent mechanism from AN saturation, is inner-hair-cell (IHC) saturation. IHC saturation in the normal ear occurs gradually at 50-70 dB SPL (Russell and Sellick, 1983, Russell et al., 1986), which is around the range of sound levels at which we normally speak (Olsen, 1998). Due to IHC saturation, slow temporal fluctuations of AN fibers are reduced in frequency channels near spectral maxima; these channels are dominated by a single harmonic or spectral peak, called capture (Deng and Geisler, 1987; Deng et al., 1987; Miller et al., 1997; Sachs et al., 2002; Zilany and Bruce, 2007). In response to a tone in a noise masker, the AN responses with CFs near the tone frequency are dominated by the added target tone, which pushes the IHC transduction nonlinearity further into saturation. However, for frequency channels far from spectral maxima, the AN responses to noise have large amplitude, low-frequency fluctuations due to cochlear filtering (Delgutte and Kiang, 1984a).

The slow temporal fluctuations in AN fibers, known as neural fluctuations (Carney, 2018), have a robust profile across various frequency channels. This profile preserves information about the stimulus even when the average firing rate of the AN becomes saturated. This cue effectively utilizes the saturation in IHC within specific frequency channels, along with low-frequency temporal fluctuations in other channels, to facilitate stimulus perception. Therefore, the limited dynamic range of IHC, which is nonlinear in

nature, not only affects this system but also serves as the underlying reason for the existence of the fluctuation profile cue.

The occurrence of IHC saturation at conversational sound levels raises exciting questions about its potential role in shaping our vocal communication. It is possible that we naturally speak at specific sound levels to create a contrast in the depth of neural fluctuation across different frequency channels. This contrast serves as a cue for perceiving various stimuli and learning about the peaks and valleys within the stimulus spectrum.

Another interesting physiological aspect of the auditory system that emphasizes the significance of neural fluctuation profiles is the modulation transfer function (MTF) of the IC cells in the auditory midbrain. The MTF represents the average neural spiking rate in response to amplitude-modulated stimuli as a function of modulation frequency. The auditory brainstem projects to these IC cells to transmit information to higher auditory processing stages in the brain. The MTF characteristics of these cells reveal that they are sensitive to the same frequency range as the neural-fluctuation frequencies observed in AN fibers (Kim et al., 2020). Consequently, the information encoded in AN fluctuations is conveyed by the IC cells through their firing rates.

There are different types of IC cells based on their MTFs (Kim et al., 2020). Band-enhanced (BE) cells are excited by low-frequency modulation in their input, while band-suppressed (BS) cells are inhibited by low-frequency modulation. Hybrid IC cells have MTFs that with a combination of behaviors seen in both BS and BE cells. For instance, some hybrid cells have BS-like behavior at relatively higher modulation frequencies and

BE-like behavior at relatively lower modulation frequencies. There are other types of IC cells, such as non-responding and flat cells, which are not the focus of this study.

In summary, the profile of neural-fluctuation depths across frequency channels serves as a robust cue in conversational sound levels, carrying essential information about the stimulus to higher stages within the auditory system. This cue arises due to the nonlinearity present in the transduction function of the IHC. Unlike level-dependent perception cues, neural-fluctuation profiles remain unaffected by the saturation of AN fibers at low to medium sound levels. Neural fluctuations are a robust coding for spectral information in the stimuli and maintaining this coding requires the control of the cochlear amplification and the efferent system, explained next, is responsible for this process.

1.2 Modeling the MOC efferent system

In hearing research, there is more focus on the ascending auditory pathway in the auditory system, which includes the process from the inner ear to the auditory cortex. This pathway involves the conversion of sound signals into neural responses, which are then analyzed and perceived as sound in the brain. However, it is fascinating to note that there exists a significant feedback in descending pathway that extends all the way from the auditory cortex, with projections at various stages in the auditory system, back to the cochlea. It should be noted that the efferent system includes an equally significant number of projections as the ascending pathway. This fact highlights that the efferent system is essentially half of the story in the auditory system, but it is still missing from most of the models for the auditory system (Guinan, 2018, Schofield, 2011).

The efferent system consists of both the lateral and medial efferent systems (Guinan, 2018). The lateral olivocochlear (LOC) efferent system directly projects to inner hair cells, although its precise functioning is still not fully understood at this stage of research. On the other hand, the MOC efferent system projects to outer hair cells, which serve as amplifiers in the ear. The projections from MOC to OHCs adjusts the cochlear gain across different frequency channels. By adjusting the cochlear gain, the MOC system may help enhance auditory processing in challenging listening environments.

The lack of information about the efferent system can be linked to challenges associated with studying this system. First, the projection in the descending pathway is highly complex, involving both contralateral and ipsilateral connections (Guinan, 2018, Schofield, 2011). This complex circuitry adds difficulty to understanding functioning of the efferent system. Additionally, the location of MOC neurons deep within the brainstem presents challenges in accessing and studying these cells. There is limited available physiology data on MOC neurons, further limiting our understanding of their properties and functions. Furthermore, many recordings of AN fibers and other tissues are often conducted using anesthetized animal models. However, anesthesia significantly affects efferent activity, leading to potential distortions in the recorded data (Aedo et al., 2015; A. R. Chambers et al., 2012; Guitton et al., 2004). This limitation highlights the need for caution when interpreting findings from such studies. Noninvasive methods, such as otoacoustic emissions (OAEs) and the cochlear microphonic, have been used to study the MOC system (Jennings & Aviles, 2023; Jamos et al., 2020, 2021; Lichtenhan et al., 2016; Smith et al., 2017). However, these methods require careful consideration of the potential

effects of the middle ear reflex (Sun, 2008, Guinan, 2018). Activation methods for studying the MOC system should be chosen with care to avoid unintended activation of the middle-ear reflex, which can affect the gain of the system.

In light of these challenges, we propose a model of the MOC efferent system. This model can serve as a valuable tool for designing future physiological and psychophysical experiments, facilitating the study of efferent function.

Implementation of the proposed model with MOC efferent feedback system:

There exists a robust and physiologically detailed model from previous study (Zilany et al., 2014) that can simulate the AN fiber response to any desire stimulus. Additionally, separate models for the midbrain auditory cells known as the IC cells were developed in the previous studies in the lab (Mao et al., 2013, Nelson and Carney, 2004), The output of the AN model serves as the input for the CN/IC model. The integration of these two models has proven to be highly successful in accurately simulating certain properties of the auditory periphery. However, as the efferent pathway is not included in these previous models, there were limitations in simulating some of the aspects of hearing that are related to MOC efferent system activity. We used this combination of AN and IC models to develop a subcortical auditory model with inclusion of the MOC efferent system.

Incorporating the efferent system into the computational AN model was a challenging task due to the need for computing the cochlear gain for each stimulus sample in the time domain based on the AN and IC model responses to previous samples of the stimulus in the time domain. To address this challenge, the AN model underwent reconfiguration to

enable the computation of responses for the entire model, one sample at a time. The states of each filter had to be saved from each sample to the next for accurate processing. In this new configuration, the feedback signals, including the neural responses of the model to the current sample, were utilized to calculate the gain required for processing the next sample. Careful testing was conducted to ensure that the newly configured model, with the modeled efferent system disabled, produced responses that matched the responses of the previous version of the model.

The proposed model for simulating the MOC efferent system incorporates two feedback systems. The first feedback system originates from the output of the AN LSR model, representing the projections from different types of WDR cells in the cochlear nucleus that project to MOC neurons. The second feedback system stems from the output of the IC model, which receives its input from the output of the HSR AN model. These two inputs are combined using a scaler that accounts for potential differences in their rates and the strength of their projections. Further investigation is required to fully understand these differences, and conducting physiological and anatomical studies would greatly benefit this aspect. The combined input signals are fed into the MOC block, processing them one time sample at a time. The MOC block includes a low-pass filter that follows the time constant of the efferent system. The output of the low-pass filtered input signal then passes through a physiologically plausible input-output function, which maps the input to MOC neurons to a scaler for adjusting the cochlear gain. The inputs to the MOC block carry information about the level and also spectral shape of the stimulus and MOC modify the cochlear gain for each frequency channel based on this information.

To incorporate the MOC model, three additional parameters were introduced to the existing list of model parameters. These parameters play essential roles in shaping the behavior of the MOC system and include:

1. Time Constant or Cutoff Frequency of the Low-Pass Filter: This parameter determines the rate at which the MOC system responds to changes in the input signal. It controls the filtering effect of the low-pass filter, mimicking the time constant of the efferent system.

2. Scaler for the IC Signal: This parameter accounts for potential differences in the rates and strengths of the input projections from the IC model. It enables the scaling of the IC signal to match the input from the AN LSR model.

3. Nonlinear Input-Output Function Parameter of the MOC: This parameter characterizes the nonlinear relationship between the input to the MOC system and the resulting adjustment of cochlear gain. It defines the shape and magnitude of the input-output function, influencing how the MOC responds to different levels of input and regulates cochlear gain accordingly.

The next crucial step in accurately modeling the MOC efferent system involved optimizing the model parameters using a previously recorded physiological dataset. Specifically, I utilized neural response recordings from IC cells in the midbrain of awake rabbits, caused by AM noise, to guide the parameter optimization process (Carney et al., 2014). The goal was to align the model's neural responses with the actual neural responses observed in this dataset.

An intriguing finding from this analysis was that approximately 80% of the BE cells in the dataset exhibited an increase in rate over time. This observation deviated from the common trend of neural adaptation, but it aligned with our hypothesis regarding the influence of the efferent system over time on these response rates. By taking a different approach to studying the dataset, focusing on the change in neural rate over time, we gained valuable insights into the impact of the efferent system. This novel perspective opens up possibilities for revisiting other physiological datasets and exploring the efferent effects by examining changes in response rates over time.

The model without the efferent system failed to simulate the observed trend of gradually increasing IC rates over time in response to AM noise, as seen in the physiological dataset. However, with the incorporation of the proposed model including the efferent system and subsequent adjustment of the parameters, we achieved successful simulation of the same increasing rate trend observed in the dataset. This outcome provides compelling support for the importance of implementing the efferent system and contributes to auditory models that more accurately simulate recorded responses in physiological datasets, particularly those obtained from un-anesthetized recordings.

1.3. Test and refine the model using psychoacoustic data

The efferent system has been proposed to influence the performance of listeners in psychoacoustic tasks. Psychoacoustic datasets are a valuable source to use for examine the efferent model's ability to simulate human auditory behavior. In this project the thresholds of listeners were compared with the model threshold in various psychoacoustic tasks, both with and without efferent activation in the model.

One potential approach to evaluate the impact of the efferent system is by exploring masking effects. Auditory masking occurs when the presence of a competing sound, such as noise, impairs the ability of human listeners to detect or perceive a target sound, such as a tone (Harris and Dallos, 1979; Moore and Oxenham, 1998). This phenomenon can be established in two forms: simultaneous masking and forward masking. Simultaneous masking occurs when the masker sound and the target sound are presented concurrently in the stimulus. The efferent system plays a role in this type of masking by modulating the cochlear gain based on the characteristics of the masker, thereby influencing the threshold for detecting the target sound. Forward masking, on the other hand, involves the perception of the target sound being affected by a preceding masker sound presented within a specific time interval. Here again, the efferent system could be involved in adjusting the cochlear gain in response to the masker and, based on the delay between the masker and target, it can have different levels of effectiveness on the perceptual threshold for detecting the target sound.

Studying both forward and simultaneous masking effects can provide us with a better understanding of how the efferent system influences speech intelligibility in noisy environments. For example, in forward masking, observing the masker's impact on perceiving sounds that follow it can resemble the effects of co-articulation, where the preceding sound (for example a vowel) influences the articulation of the subsequent sound (for example a consonant). Similarly, simultaneous masking experiments can shed light on how the efferent system affects the neural coding of speech in the presence of background noise. These experiments offered a valuable means to explore the role of the efferent

system in shaping auditory perception and further emphasized the importance of incorporating this system in computational models of auditory processing. Moreover, incorporating the findings from these experiments into the model implementation led to improvements in the model's ability to simulate and replicate the observed behaviors in real-world listening tasks. By considering the simulation results from these masking tasks, we were able to refine and validate the model's implementation of the efferent system.

For the purpose of testing the proposed model with psychoacoustic data, I will explain three collaborative projects briefly here.The fifth project, which is an independent study, will be discussed in more detail in Chapter 3 of the thesis.

1.3.1. Forward Masking: studying the effect of temporal fluctuation in masker and hearing loss on tone detection thresholds

This project was a collaboration with Marc Brennan, Adam Svec, Braden Maxwell, and Laurel Carney and was published in the Journal of the Acoustic Society of America (Brennan et al., 2023).

In this experiment, both the model with efferent and the model without efferent were used to estimate the threshold of human listeners in detecting a tone after the presence of a masker. To compare the model's threshold with that of human listeners, we utilized an existing psychoacoustic dataset collected by our co-authors, Brennan and Svec.

The dataset consisted of human listeners' thresholds for detecting a tone at 4000 Hz with duration of 10 ms after different time intervals (25, 75, or 150 ms) following a noise masker with a duration of 400 ms. Two types of maskers were examined: low fluctuation

noise (LFN) and Gaussian noise (GN). The Gaussian noise had a bandwidth of one third of the equivalent rectangular bandwidth (1/3 ERBN) centered at 4000 Hz. The Gaussian masker has a more fluctuating temporal envelope. The other masker used in this study was a low fluctuation noise, achieved by dividing the Gaussian noise by its envelope, using a Hilbert function, and repeating the procedure 10 times to obtain a noise with a flat envelope. For both maskers, the sound level was 80 dB SPL. In this dataset, the listeners were categorized into three groups based on their audiograms: normal hearing, mild hearing loss, and severe hearing loss. The psychoacoustic results revealed that, in general, the listeners exhibited higher thresholds or lower performance when detecting the tone after the Gaussian fluctuating masker. The difference between the threshold of tone detection after LFN and GN is referred to as Gaussian Noise Disruption (GND). For all groups of human listeners, GND is positive for the 25 ms interval, indicating that the threshold for detecting the tone is higher with the GN masker compared to the LFN masker. As the interval delay between the masker and target tone increases, GND and overall thresholds for both masker conditions decreases, indicating a reduction in the masking effect. Notably, the GND is significantly higher for listeners with mild hearing loss compared to the other two groups in the 25-ms condition.

Efferent activity can provide an explanation for the observed trends in this dataset. Based on the MOC efferent system dynamics, our hypothesis suggests that the GN masker elevates the threshold more compared to the LFN masker due to a greater decrease in cochlear gain in response to the fluctuating GN masker. This effect is because of the increased rate of the BE IC cells in response to fluctuating masker. This higher IC rate are

part of the projection to the MOC, as a result, the efferent activity increases, leading to further suppression of cochlear gain. Consequently, the lower cochlear gain after the GN masker makes the detection of the tone more challenging.

However, as the delay between the masker and target tone increases, the cochlear gain begins to recover due to the minimal response from the AN LSR and IC during the silent interval. The recovery in cochlear gain during the silent interval reduces the differences between the effect of two maskers on detection of target tone. Additionally, with a longer silent delay, the gain at the onset of the target tone is higher, leading to an overall decrease in the threshold for both masker conditions. It is worth noting that the timing of the delay interval for masking release aligns with the time constant of the efferent system.

To validate our hypotheses and gain a deeper understanding of the neural mechanisms underlying forward masking, we conducted experiments using both the model with efferent and the model without efferent. We replicated the same experiment that was conducted with human listeners and estimate the detection thresholds for different listener groups for both types of maskers and various interval durations. We simulate the hearing loss by changing the related parameters of the model based on the average audiogram of subjects in the psychoacoustic dataset. In our simulations, we utilized the output of the IC model within a specific time window, starting from 375 ms after the onset of the masker and extending until at least 50 ms after the end of the probe tone. We compared the maximum simulated IC rate between two intervals: one with the target tone and the other without the target tone. Based on the maximum IC rate, the model predicted the target interval, and we measured the accuracy of the model by conducting 50 repetitions at different tone levels.

This allowed us to estimate the detection thresholds for both the model with efferent and the model without efferent in this task.

The simulation results strongly supported the role of MOC activity in forward masking, as the model with efferent successfully replicated the observed masking effects. In contrast, the model without efferent failed to simulate the GND effect for the 25 ms delay condition, which was significant for human listeners. However, it is important to note that the adjustment of cochlear gain by the MOC system alone cannot fully explain the GND effect. This is evident from the observation of a GND of approximately 10 dB in the model without efferent for the severe hearing loss simulation. We hypothesize that listeners may have confused the GN masker more than the LFN masker with the target. In the peripheral auditory system world, this can be explained: the increased IC rate in response to the fluctuating GN masker compared to the flat-envelope LFN masker can lead to a higher level of confusion, as these responses can be mistakenly interpreted as responses to the target tone by the IC cells. While this factor may contribute to the GND observed in the 25-ms condition, it cannot fully account for the differences observed among different groups of listeners based on their hearing status.

The model incorporating the MOC efferent system not only successfully simulated the observed GND in different interval durations but also captured the trend observed in the psychoacoustic dataset, specifically the greater influence of temporal fluctuations in the GN masker on listeners with mild hearing loss compared to those with normal and severe hearing loss. This trend can be justified by the activity of the MOC efferent system. In the case of normal hearing, the high sound level and narrow bandwidth of the masker result in

strong input to the IHCs, pushing them into saturation. As a result, the responses of AN fibers become flat for both GN masker and LFN masker, leading to less observed GND in this group of listeners. In contrast, for mild hearing loss, the impaired cochlear gain causes the IHC response to the GN masker to be less saturated, resulting in more fluctuating AN fiber responses and higher rates of activity in the IC. The increased IC rate not only reduces the cochlear gain due to the projections of these cells to the MOC system and subsequent activation of MOC activity, but it also increases the likelihood of confusion between the response to the masker and the response to the target tone. Both factors contribute to making the detection of the tone more challenging after the GN masker compared to the LFN masker in listeners with mild hearing loss. In the case of severe hearing loss, the overall cochlear gain is already very low, leaving minimal room for gain reduction and limited increase in IC rate. Consequently, the differences in detection thresholds between the LFN and GN masker conditions are less obvious in this group.

1.3.2. Forward Masking: studying the effect of tonal masker level and tone detection thresholds

This project was a collaboration with Braden Maxwell, Marc Brennan, Adam Svec, and Laurel Carney and was presented at The Association for Research in Otolaryngology midwinter meeting (Maxwell et al., 2023).

In this project, the focus was on investigating the relationship between masker level and the threshold for detecting a tone in the presence of a tone masker. Both physiological studies, which involved recording neural responses from IC cells, and psychoacoustic

studies with human listeners have demonstrated that as the tone masker level increases, the threshold for tone detection also increases. This phenomenon is known as growth of masking. Previous studies have suggested that the emergence of growth of masking occurs at the IC level, as it was not observed in AN fiber recordings. However, it is not clear how growth of masking emerges in the IC, and existing physiological auditory models for IC cells have not been able to fully explain this effect. An alternative approach to understanding the neural mechanisms underlying growth of masking is the involvement of MOC efferent activity. The MOC system includes the WDR projection, which carries information about the masker level. As the masker level increases, the WDR firing rates also increase, leading to an increase in MOC activity. This increased MOC activity results in a reduction in cochlear gain at the target frequency, making the detection of the target tone more challenging. In other words, it raises the detection threshold. In the context of this task, the WDR projection plays a crucial role in mediating the effect of growth of masking, particularly when the IC responses are low and not changing with masker level a lot due to the flat envelope of a tone masker. The contribution of MOC activity as a potential neural mechanism for growth of masking can provide an explanation for the lack of observation of this phenomenon in AN recordings. During experiments where animals are anesthetized, the efferent activity is significantly affected. Anesthesia suppresses the MOC efferent system, leading to reduced modulation of cochlear gain. As a result, the effects of growth of masking may not be observed in AN recordings under anesthesia.

We performed simulations using both the model with efferent feedback and the model without efferent feedback at the AN and IC levels to estimate tone-detection thresholds.

The stimuli used in our study were matched to those used in the original studies. To estimate thresholds, we used a similar method to our previous work, comparing the maximum response of the model for two internal tasks: one with the target tone present and the other without. By adjusting the tone level, we determined the threshold corresponding to an accuracy of 70.7%. Our findings demonstrated that both the AN model and the IC model with efferent feedback exhibited similar thresholds and displayed the growth of masking effect, consistent with physiological and psychophysical evidence. However, the model without efferent feedback failed to simulate this effect, highlighting the crucial role of the efferent system in the growth of masking phenomenon. Furthermore, the model with efferent feedback successfully replicated the release from masking by increasing the delay between the masker and target tone. This release from masking can be linked to the recovery of cochlear gain during the extended silent interval, enabling improved discrimination of the target tone. Overall, our results emphasize the significance of the efferent feedback mechanisms in auditory processing and provide valuable insights into the neural mechanisms underlying the perception of masked tones.

1.3.3. Forward Masking: Studying the effects of spectral properties of masker on detection of the tone

This project was a collaboration with Swapna Agarwalla and Laurel Carney and was presented at The Association for Research in Otolaryngology midwinter meeting (Farhadi et al., 2023).

We also investigated the auditory enhancement effect as another psychoacoustic task to test and refine the model with MOC efferent feedback and explore the influence of this feedback system on the underlying neural mechanisms for auditory enhancement. Auditory enhancement involves the detection of a tone embedded in a background sound, which is enhanced when preceded by a precursor stimulus with same background sound but without containing the frequency of the target tone. This phenomenon is referred to as the "target sound popping out". In auditory enhancement, not only the detection of the target sound embedded in a background masker improves when proceed by the precursor described above (signal enhancement) but the presence of the precursor also enhances the effectiveness of the combination of the target and masker as a forward masker for the subsequent detection of a tone presented alone (masker enhancement). This effect was observed and well-studied in the human listener's psychoacoustic experiment.

In our investigation of the auditory enhancement effect, we specifically focused on studying masker enhancement, which offers a more direct and accessible approach to studying auditory enhancement phenomenon. The main objective of this study was to predict masker enhancement effect in individuals with normal (NH) and impaired (HI) hearing, as previously explored by Kreft and Oxenham (2019). Surprisingly, their findings revealed unexpected effects of sound pressure level (SPL) and hearing impairment as the masker enhancement effect significantly decreases for hearing impaired listeners and for lower sound levels. These observations could not be easily explained by the conventional explanation of adaptation of inhibition as the underlying neural mechanisms for auditory enhancement. In this study, we explored the possible role of the MOC efferent system in

explaining the masker enhancement effect in different hearing and sound level conditions. To examine the underlying mechanisms of masker enhancement, we hypothesized that a subcortical model incorporating the dynamics of the efferent system could account for the observed effects.

We simulate the same experiment as in the original study with five different stimulus conditions

- MSK (Baseline): Masker that includes the frequency of the target signal.
- ENH (Enhanced condition): Precursor added before the masker, but without the frequency of the target signal.
- CON (No Enhancement): Control condition where the precursor contains the frequency of the target signal.
- MSK0: Masker without the frequency of the target signal.
- ENH0: Both the precursor and masker lack the frequency of the target signal. This condition serves as a control to assess the impact of masker and precursor duration.

The precursor duration was 500 ms, followed by a masker and a tone probe with duration of 100 ms and 20 ms respectively, each separated by 20-ms gaps. The masker was four equal-amplitude, logarithmically spaced sinusoids geometrically centered around the target frequency. In a study by Kreft and Oxenham (2019), probe level was manipulated to determine the detection threshold for two models: normal hearing (NH) and hearing-

impaired (HI). The researchers investigated three different level configurations for the NH thresholds.

1- SPL matched to level that was used for HI (85 dB SPL/comp): In this configuration, the probe level for the NH model was set to match the level used for the HI model, which was 85 dB sound pressure level (SPL) per component.
2- Sensation level (SL) matched to average HI SL: In this configuration, the probe level for the NH model was reduced to match the average sensation level (SL) of the HI listeners. Sensation level refers to the level of sound above an individual's hearing threshold.
3- SPL and SL matched at same time. In this condition, the SPL is set 85 dB and SL matched to average HI listener using threshold equalizing noise (TEN). In this configuration, the probe level for the NH model was set to 85 dB SPL, similar to the first configuration. Additionally, the SL was adjusted to match the average SL of the HI listeners using TEN. TEN is a type of noise used to equalize hearing thresholds across different frequencies.

In each two-interval trial, the decision variable was used to determine which interval contained the target tone and which interval did not. The decision variable involved comparing the maximum average rate of IC rate between the two intervals. The IC rate used in the decision variable was obtained from summing the responses of nine IC BE cells. These cells had CFs that covered the frequency range of the masker components. The decision variable time window 105-ms, which included the final 25 ms of the masker duration and the entire duration of the probe tone. Model thresholds for probe detection

were estimated by varying the probe level until a 70.7% accuracy in tone detection was achieved.

The masker enhancement effect was measured by comparing the tone-detection thresholds with and without the precursor stimuli. In human listeners' dataset, it was found that listeners with normal-hearing (NH) showed a significant masker-enhancement effect when the sound pressure level (SPL) was matched to hearing loss. However, when the sensation level (SL) or both SPL and SL were matched to those for hearing-impaired listeners, the enhancement effect was reduced. Additionally, listeners with hearing loss did not show a significant masker enhancement effect in the original psychoacoustic study.

The simulation results of the model that included efferent gain control closely matched the thresholds and trends of masker enhancement observed in human listeners. However, the model without efferent gain control failed to capture these effects. The key finding is that the model with efferent gain control successfully predicted the masker enhancement observed in the NH SPL-matched condition, while showing little enhancement in all other conditions, which aligns with the psychophysical data obtained from human listeners.

To gain a better understanding of how a computational model with a MOC efferent system contributes to more accurately estimating listeners' thresholds in an auditory enhancement task, we examined the simulated neural responses at different stages of the auditory pathway using the model. The responses of model stages were analyzed for a CF of 4 kHz. The stimuli were presented to both normal hearing (NH) and hearing-impaired (HI) models to investigate the mechanism behind auditory enhancement. The probe level

was set above the individual threshold for both the NH and HI model in these figures. In the NH model, the cochlear gain is reduced during the precursor. This reduction in cochlear gain resulted in a more effective masking of the target sound in the ENH condition, as indicated by a lower response to target tone in IC response to the stimuli with the precursor. The enhanced masking effect leading to increased threshold to detect target sound. On the other hand, in the HI model, the cochlear gain was already reduced due to the hearing impairment. The addition of the precursor had little effect on further reducing the already reduced cochlear gain. As a result, the enhancement observed in the HI model was less than NH condition leading to reduced auditory enhancement.

We also investigated the neural responses that were used in the decision variable. The decision variable was calculated using the sum of the BE IC responses over different frequencies. Upon analyzing these neural responses, we observed that in the NH SPL-matched condition, there was a reduction in the response to the probe. This reduction is consistent with the phenomenon of masker enhancement, where the presence of the precursor leads to a more effective masking of the probe tone. The smaller differences in the neural responses observed in other conditions, as well as in the HI condition, are in line with the reduced enhancement observed experimentally (Kreft & Oxenham, 2019).

Interestingly, these simulated neural responses also had a stronger summed IC BE response during the masker phase for the ENH condition. This finding aligns with the concept of signal enhancement, indicating that the precursor may enhance the representation of the masker in the neural responses. The reason for this simulation result is that the efferent activity driven by IC BE neurons, which are excited by fluctuations,

reduces cochlear gain, further increasing low-frequency fluctuations in AN responses, and thus further increasing IC BE responses. As a result of reduction in the cochlear gain, AN HSR responses to the masker have strong fluctuations for low-CFs, where multiple component frequencies fall within AN tuning curves. These AN HSR fluctuations result in stronger IC BE masker response rates at low CFs, which would improve detection of the target (Auditory enhancement).

In summary, a subcortical model with MOC efferents simulated masker enhancement and also signal enhancement effects observed in NH and HI listeners in three sound-level conditions, whereas a model without efferents failed to do so. These findings support the hypothesis that efferent activity could explain auditory enhancement.

1.3.4. Simultaneous Masking: Studying the Overshoot effect

This work is accepted as a proceeding paper to 2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (Farhadi and Carney, 2023) and was presented at The Association for Research in Otolaryngology midwinter meeting (Farhadi and Carney, 2023).

For the investigation of a simultaneous-masking temporal effect known as overshoot, psychoacoustic studies have demonstrated that the detection of a short tone added to a longer masker noise can be improved when there is a sufficiently long delay between the onset of the masker and the addition of the target tone. In contrast, when the delay between the onset of the tone and the onset of the masker is very short, the detection of the tone is more challenging.

To gain a deeper understanding of this phenomenon, we utilized both the model with the MOC efferent system and the model without the efferent system. The intermediate responses for both models at different levels of auditory pathway in response to overshoot stimulus is shown in Chapter 3, Fig. 3.1 We examined the computational model with and without efferent to estimate human listeners' thresholds for the long and short delay conditions in the auditory overshoot experiment. Similar to the original study, the tone level was constant, while the masker level was adjusted to reach the threshold level. This threshold level represents the maximum noise level that the model can tolerate while still detecting the tone. In the short delay condition of 2 ms, the thresholds were found to be similar for both the model with efferent thresholds and the model without efferent thresholds. This similarity exists because 2 ms is too short of a delay for the efferent system to influence the cochlear gain considering the time constant of around 200 ms for the MOC efferent system. Therefore, the presence or absence of efferent system does not affect the thresholds in the short-delay condition.

However, as the delay increased from 2 ms to 200 ms, we observed that the model with efferent thresholds exhibited improved thresholds, which is consistent with the findings observed in human listeners. The reason for this finding could be that due to the reduction in cochlear gain, as a result of efferent activity in response to masker noise, the IHC response is pulled away from saturation. As a result, the response to the tone would have a strong peak in the IC BE response, while for the model without efferent the IHC is saturated and the AN response is flat, and as a result the IC BE response is much weaker. The efferent system, given sufficient time, can effectively modulate the cochlear gain and

increase the auditory-overshoot effect. This improvement aligns with the behavior observed in human listeners. In contrast, the model without efferent thresholds showed worse thresholds as the delay increased, contrary to the behavior observed in human listeners. More detailed information about this experiment and its findings are included in Chapter 3 of this thesis.

1.4. Study of the effect of efferent feedback on neural coding of speech sound

Improving speech intelligibility in background noise is an important long-term goal of this line of research. Many studies that work on speech enhancement focus on improving the acoustic waveform of the speech sound and test the improvement using neural-network based models for the auditory cortex (Fu et al., 2017; Tan and Wang, 2018). However, in this method, one important step is missing, the auditory peripheral system. The transformations that the auditory periphery applies to the audio waveform play an important role in translating the information in the acoustic sound to the neural response of AN fibers. Malfunctioning of the auditory peripheral system is in most cases the reason behind lower speech perception in listeners with hearing loss (Rance, 2005). Therefore, understanding the auditory periphery and focusing on improvement of the neural responses of the auditory system is potentially more effective for speech intelligibility purposes.

Neural coding of vowels

Speech is a complex phenomenon that involves the production of various sounds, including both voiced and unvoiced segments. Voiced segments are produced when the vocal cords vibrate, generating a periodic waveform. These segments include vowel sounds and certain consonants. For this project, we mostly focus on the vowel, an important sound

in any spoken language. Vowels have higher energy and stronger periodicity compared to the consonants, and in the existence of background noise, they carry more information than consonants (Parikh and Loizou, 2005; Rao and Carney, 2014), possibly due to the structure of vowel waveforms. The spectrum of a vowel waveform has several strong peaks that are called vocal formants. These strong peaks could preserve their structure even in the existence of the background noise. The formants in vowel sounds are the result of vocal-tract resonances, and the frequency of the first two formants and their relationship are what researchers in the acoustic and linguistic area used for classification of vowels (Fant, 1960). Based on the frequencies of first two formants, vowels are distributed in a vowel space. This acoustical vowel space has the same distribution for most languages, regardless of the number of vowels in that language (Lindblom, 1986; Stevens, 1989; Engstrand and Krull 1991, Livijn, 2000; Al-Tamimi and Ferragne , 2005).

The spectral peaks or vocal formants of the vowel waveform can be coded in the neural fluctuation profile of AN fibers. Because of capture effect and IHC saturation, the AN fibers with CF close to the formant frequency will have flat responses; AN fibers with CF away from formant frequencies will have fluctuating responses (Carney and McDonough, 2019; Carney et al., 2015; Young and Sachs, 1979; Deng and Geisler, 1987; Miller et al., 1997). In BE IC cell responses, the flatness in the AN fiber response will be coded as a dip in average-rate responses of neurons tuned near the formant frequency in the IC response profile and a peak in responses of neurons tuned to frequencies further from formants.

Effect of MOC efferent on neural coding of vowels

As we discussed, the neural fluctuation profile is a reliable cue at conversational level for perceiving different sound including speech. Here we focus on improvement in the contrast across frequency channels in the neural fluctuation profile as a measure for improvement in speech intelligibility. We hypothesized that the MOC efferent system can improve this contrast and as a result can enhance speech intelligibility. In this work, the IC feedback in the MOC efferent model is hypothesized to improve this contrast in the IC pattern, as it will increase the gain in channel near formant frequencies and decrease the gain at other frequencies.

For the stimuli tested with the model incorporating the MOC efferent system so far, we used simple and narrowband spectra, while speech signals are more complex and broadband. In response to a peak in the spectrum corresponds to the narrowband stimulus, the inner hair cell (IHC) becomes saturated, resulting in a flat envelope in the AN response. This leads to a decrease in the IC BE response and an increase in the IC BS response. The efferent system, through the IC BE projection pathway, enhances this profile by increasing the cochlear gain, further saturating the IHC and flattening the AN response. Simultaneously, for other frequencies, the non-saturating IHC and higher neural fluctuation in the AN result in an increase in IC rate and a decrease in cochlear gain, pulling the IHC away from saturation and increasing fluctuation in the AN response.

For wideband stimuli, the fine-tuning implementation of WDR projection can flatten the profile of the AN response. If there is a peak in the stimulus spectrum, the CF-by-CF, WDR projection will have a high response to the peak, reducing the gain for that frequency

and pushing the IHC away from saturation. If the IHC is not saturated in that frequency, there will not be enough contrast in that neighborhood for the IC pathway to enhance, and it will not enter the positive feedback loop that continuously improves the contrast.

Updated structure for the MOC efferent model with wide WDR projection

Anatomical and physiological evidence suggests that the MOC projection to the cochlea has a broader tuning compared to AN fibers (Guinan, 2018). Some studies indicate the presence of two distinct types of MOC projections (Liberman & Brown,1986; Brown, 2014) with different bandwidth characteristics. While the specific differences between these two types and their input sources are not yet fully understood, our hypothesis that there are cross-frequency WDR-driven and within-channel IC-driven projections from MOC to OHCs . In our model, we implemented this effect by separating the MOC neurons' projections based on whether they receive input from the WDR or IC pathway. The WDR-driven projections are broader, allowing for a wider adjustment of the cochlear gain, while the IC-driven projections are sharper, enabling more localized and detailed adjustments. Essentially, the WDR-driven projections modulate the overall range of cochlear gain that the IC can work with across a wide frequency neighborhood, while the IC projections fine-tune the gain for each specific cochlear frequency within that range, based on local peaks and valleys in that neighborhood.

The observations in our model are consistent with findings from hearing aids that employ a higher number of channels (Holube et al., 2016; Bor et al., 2008; Plomp, 1988; Stone and Moore, 2008). Because of necessary compression feature in hearing aid implementation, when there is a peak in the stimulus spectrum, a fine tined hearing aid

amplifies the sound minimally. Conversely, for frequencies in the spectral valleys, the sound is amplified maximally. This results in a flatter spectral response or similar response across all frequencies, which can lead to a reduction in the perception of spectral information in the sound.

For evaluating speech signals, we employed the wider WDR projection implementation described earlier and adjusted the bandwidth of the WDR projection to optimize the sharpness of the responses to a single formant frequency in a silent condition.

Single-Formant Simulations

We investigated the neural fluctuation profile using the adjusted parameters in response to a single-formant stimulus. The single-formant stimulus is harmonic complex with fundamental frequency of 200 Hz, duration of 300 ms, peak frequency at 2 kHz with a spectral slope of 50 dB/octave. This stimulus was chosen as a starting point as its spectrum is similar to a vowel spectrum but only has one formant frequency. Consistent with our hypothesis the IC BE response to single formant stimulus has a sharper response as it has low rates in peak frequency and higher rates in other frequency when compared to the model without efferent.

We further explored the effects of adding background speech-shaped noise (Byrne et al., 1994) to the stimulus. The model with efferent MOC exhibited more robust responses with the addition of noise at different signal-to-noise ratio (SNR) levels, compared to the model without efferent. Additionally, we examined the impact of stimulus level variations

in a silent condition, and the results supported the hypothesis that the efferent system preserves the sharp responses in wider range sound levels, compared to the model without efferent.

Synthesized Vowel (double-formant) Simulations

The next step we explore the neural response to double-formant synthesized vowels made using Klatt software (Klatt, 1980). This stimulus is very similar to the single-formant stimulus but since it has two formant peaks it better resembles a vowel signal. The properties of the synthesized-vowel stimuli used here were similar to those in Carney et al., (2023). The peak frequencies were 500 and 2100 Hz with a bandwidth of 100 Hz used for both formants. The duration of the stimuli was 300 ms and the stimulus level was 75 dB. The comparison between the models with efferent and without efferent shows a sharper response to peak frequencies in IC BE response affected by efferent activation.

1.3. Overview of the thesis

The first chapter of this thesis provides essential background information that is crucial for understanding the subsequent sections of this document. It begins with an explanation of the neural-fluctuation profile as a robust and significant cue for auditory perception, along with the underlying neural mechanisms. The importance of considering the descending efferent system in the auditory pathway is also highlighted. The architecture of the proposed model with MOC efferent is briefly reviewed, along with the process of parameter optimization. Furthermore, an overview of the psychophysical tests used to evaluate the model's performance in estimating human listeners' thresholds is presented. These results not only validate the model's accuracy in predicting human auditory behavior

but also emphasize the significant role of the MOC efferent system in various auditory tasks. Additionally, a brief exploration of the neural coding of vowels as components in speech signals and the effects of MOC efferent on the neural response profile is discussed. Finally, this chapter briefly study the impact of MOC efferent on improving speech intelligibility in noisy environments.

Chapter 2 has been submitted for publication in Journal of Acoustic Society of America (JASA) and explains the proposed MOC model structure and implementation in details. This chapter also discussed the process of optimization of the parameter proposed in this thesis using the physiological dataset including the neural responses of IC cells in awake rabbit in response to AM noise. Finally, this chapter includes the validation of the model for simulating the physiological dataset that were previously successfully simulated with model without efferent to make sure that MOC addition to the code is not disturbing those neural responses.

Chapter 3 has been accepted in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. This chapter explains the independent study on testing the efferent model with psychoacoustic data from an overshoot experiment. In this chapter an explanation of the stimuli and the procedure is explained and the estimated threshold using the model with efferent and the model without efferent were compared to human behavioral data.

Chapter 4 proposes a new implementation of the MOC efferent model with cross-frequency projections and explores the effect of this model on simulation the neural

response to speech-like stimuli in different sound and noise levels. This chapter investigates the role of the MOC efferent system on enhancing of the neural-fluctuation profile coding of speech-like sound over time.

Finally, in Chapter 5, a summary and discussion of all three projects is presented. Ideas for future work are discussed at the end of the thesis.

Bibliography

Spille, C., Kollmeier, B., & Meyer, B. T. (2018). Comparing human and automatic speech recognition in simple and complex acoustic scenes. Computer Speech & Language, 52, 123-140.

Humes, L. E., Kidd, G. R., & Lentz, J. J. (2013). Auditory and cognitive factors underlying individual differences in aided speech-understanding among older adults. Frontiers in systems neuroscience, 7, 55.

Chou, K. F., Dong, J., Colburn, H. S., & Sen, K. (2019). A physiologically inspired model for solving the cocktail party problem. Journal of the Association for Research in Otolaryngology, 20, 579-593.

Stern, R. M., & Morgan, N. (2012). Hearing is believing: Biologically inspired methods for robust automatic speech recognition. IEEE Signal Processing Magazine, 29(6), 34–43.

Stern, R. M., Morgan, N., Virtanen, T., Raj, B., & Singh, R. (2012). Features Based on Auditory Physiology and Perception. Techniques for Noise Robustness in Automatic Speech Recognition, 193227.

Schädler, M. R., Warzybok, A., Ewert, S. D., & Kollmeier, B. (2016). A simulation framework for auditory discrimination experiments: Revealing the importance of across-frequency processing in speech perception. The journal of the acoustical society of America, 139(5), 2708-2722.

Meyer, B. T., & Kollmeier, B. (2011). Robustness of spectro-temporal features against intrinsic and extrinsic variations in automatic speech recognition. Speech Communication, 53(5), 753-767.

Verhulst, S., Altoè, A., & Vasilkov, V. (2018). Computational modeling of the human auditory periphery.

Zilany, M. S., Bruce, I. C., & Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. The Journal of the Acoustical Society of America, 135(1), 283-286.

Zweig, G. (2015). Linear cochlear mechanics. The Journal of the Acoustical Society of America, 138(2), 1102-1121.

Meddis, R., Lecluyse, W., Clark, N. R., Jürgens, T., Tan, C. M., Panda, M. R., & Brown, G. J. (2013). A computer model of the auditory periphery and its application to the study of hearing. Basic aspects of hearing: physiology and perception, 11-20.

Meddis, R., O'Mard, L. P., & Lopez-Poveda, E. A. (2001). A computational algorithm for computing nonlinear auditory frequency selectivity. The Journal of the Acoustical Society of America, 109(6), 2852-2861.

Schofield, B. R. (2011). Central descending auditory pathways. In Auditory and vestibular efferents (pp. 261–290). Springer.

Guinan Jr, J. J. (2018). Olivocochlear efferents: Their action, effects, measurement and uses, and the impact of the new conception of cochlear mechanical responses. Hearing Research, 362, 38–47.

Brown, G. J., Ferry, R. T., & Meddis, R. (2010). A computer model of auditory efferent suppression: implications for the recognition of speech in noise. The Journal of the Acoustical Society of America, 127(2), 943-954.

Yasin, I., Drga, V., Liu, F., Demosthenous, A., & Meddis, R. (2020). Optimizing speech recognition using a computational model of human hearing: effect of noise type and efferent time constants. IEEE Access, 8, 56711-56719.

Mertes, I. B., Wilbanks, E. C., & Leek, M. R. (2018). Olivocochlear efferent activity is associated with the slope of the psychometric function of speech recognition in noise. Ear and hearing, 39(3), 583.

Liberman, M. C. (1978). Auditory-nerve response from cats raised in a low-noise chamber. The Journal of the Acoustical Society of America, 63(2), 442-455.

Costalupes, J. A., Young, E. D., & Gibson, D. J. (1984). Effects of continuous noise backgrounds on rate response of auditory nerve fibers in cat. Journal of neurophysiology, 51(6), 1326-1344.

Ye, Y., Machado, D. G., & Kim, D. O. (2000). Projection of the marginal shell of the anteroventral cochlear nucleus to olivocochlear neurons in the cat. Journal of Comparative Neurology, 420(1), 127–138.

Olsen, W. O. (1998). Average speech levels and spectra in various speaking/listening conditions.

Russell, I. J., & Sellick, P. M. (1983). Low-frequency characteristics of intracellularly recorded receptor potentials in guinea-pig cochlear hair cells. The Journal of physiology, 338(1), 179-206.

Russell, I. J., Richardson, G. P., & Cody, A. R. (1986). Mechanosensitivity of mammalian auditory hair cells in vitro. Nature, 321(6069), 517-519.

Deng, L., & Geisler, C. D. (1987). Responses of auditory-nerve fibers to nasal consonant–vowel syllables. The Journal of the Acoustical Society of America, 82(6), 1977-1988.

Deng, L., Geisler, C. D., & Greenberg, S. (1987). Responses of auditory-nerve fibers to multiple-tone complexes. The Journal of the Acoustical Society of America, 82(6), 1989-2000.

Miller, R. L., Schilling, J. R., Franck, K. R., & Young, E. D. (1997). Effects of acoustic trauma on the representation of the vowel/ɛ/in cat auditory nerve fibers. The Journal of the Acoustical Society of America, 101(6), 3602-3616.

Sachs, M. B., Bruce, I. C., Miller, R. L., & Young, E. D. (2002). Biological basis of hearing-aid design. Annals of Biomedical Engineering, 30, 157-168.

Zilany, M. S., & Bruce, I. C. (2007). Representation of the vowel/ɛ/in normal and impaired auditory nerve fibers: model predictions of responses in cats. The Journal of the Acoustical Society of America, 122(1), 402-417.

Delgutte, B., & Kiang, N. Y. (1984). Speech coding in the auditory nerve: I. Vowel-like sounds. The Journal of the Acoustical Society of America, 75(3), 866-878.

Carney, L. H. (2018). Supra-threshold hearing and fluctuation profiles: implications for sensorineural and hidden hearing loss. Journal of the Association for Research in Otolaryngology, 19(4), 331-352.

Kim, D. O., Carney, L., & Kuwada, S. (2020). Sensory Processing: Amplitude modulation transfer functions reveal opposing populations within both the inferior colliculus and medial geniculate body. Journal of Neurophysiology, 124(4), 1198.

Aedo, C., Tapia, E., Pavez, E., Elgueda, D., Delano, P. H., & Robles, L. (2015). Stronger efferent suppression of cochlear neural potentials by contralateral acoustic stimulation in awake than in anesthetized chinchilla. Frontiers in Systems Neuroscience, 9, 21.

Chambers, A. R., Hancock, K. E., Maison, S. F., Liberman, M. C., & Polley, D. B. (2012). Sound-evoked olivocochlear activation in unanesthetized mice. Journal of the Association for Research in Otolaryngology, 13, 209-217.

Guitton, M. J., Avan, P., Puel, J.-L., & Bonfils, P. (2004). Medial olivocochlear efferent activity in awake guinea pigs. Neuroreport, 15(9), 1379–1382.

Jennings, S. G., & Aviles, E. S. (2023). Middle ear muscle and medial olivocochlear activity inferred from individual human ears via cochlear potentials. The Journal of the Acoustical Society of America, 153(3), 1723-1732.

Jamos, A. M., Kaf, W. A., Chertoff, M. E., & Ferraro, J. A. (2020). Human medial olivocochlear reflex: contralateral activation effect on low and high frequency cochlear response. Hearing Research, 389, 107925.

Jamos, A. M., Hosier, B., Davis, S., & Franklin, T. C. (2021). The role of the medial olivocochlear reflex in acceptable noise level in adults. Journal of the American Academy of Audiology, 32(03), 137-143.

Lichtenhan, J. T., Wilson, U. S., Hancock, K. E., & Guinan Jr, J. J. (2016). Medial olivocochlear efferent reflex inhibition of human cochlear nerve responses. Hearing research, 333, 216-224.

Smith, S. B., Lichtenhan, J. T., & Cone, B. K. (2017). Contralateral inhibition of click-and chirp-evoked human compound action potentials. Frontiers in neuroscience, 11, 189.

Sun, X.-M. (2008). Contralateral suppression of distortion product otoacoustic emissions and the middle-ear muscle reflex in human ears. Hearing Research, 237(1–2), 66–75.

Mao, J., Vosoughi, A., & Carney, L. H. (2013). Predictions of diotic tone-in-noise detection based on a nonlinear optimal combination of energy, envelope, and fine-structure cues. The Journal of the Acoustical Society of America, 134(1), 396–406.

Nelson, P. C., & Carney, L. H. (2004). A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. The Journal of the Acoustical Society of America, 116(4), 2173-2186.

Carney, L. H., Zilany, M. S. A., Huang, N. J., Abrams, K. S., & Idrobo, F. (2014). Suboptimal use of neural information in a mammalian auditory system. Journal of Neuroscience, 34(4), 1306–1313.

Harris, D. M., & Dallos, P. E. T. E. R. (1979). Forward masking of auditory nerve fiber responses. Journal of neurophysiology, 42(4), 1083-1107.

Moore, B. C., & Oxenham, A. J. (1998). Psychoacoustic consequences of compression in the peripheral auditory system. Psychological review, 105(1), 108.

Brennan, M. A., Svec, A., Farhadi, A., Maxwell, B. N., & Carney, L. H. (2023). Inherent envelope fluctuations in forward masking: Effects of age and hearing loss. The Journal of the Acoustical Society of America, 153(4), 1994-1994.

Maxwell, B. N., Farhadi, Brennan, M. A., Svec, & Carney, L. H. (2023). Simulating Physiological and Psychoacoustic Forward Masking in a Subcortical Model with Efferent Gain Control. Abstract, Association for Research in 553 Otolaryngology, SA200, pg.301.

Farhadi, A., Agarwalla, S. & Carney, L. H. (2023). A Subcortical Auditory Model With Efferent 552 Gain Control Explains Perceptual Enhancement. Abstract, Association for Research in 553 Otolaryngology, SU198, pg.449.

Kreft, H. A., & Oxenham, A. J. (2019). Auditory enhancement under forward masking in normal-hearing and hearing-impaired listeners. The Journal of the Acoustical Society of America, 146(5), 3448-3456.

Farhadi, A., & Carney, L. H. (2023). Predicting thresholds in an auditory overshoot paradigm using a computational subcortical model with efferent feedback. In WASPAA 2023. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY.

Fu, S. W., Hu, T. Y., Tsao, Y., & Lu, X. (2017, September). Complex spectrogram enhancement by convolutional neural network with multi-metrics learning. In *2017 IEEE 27th international workshop on machine learning for signal processing (MLSP)* (pp. 1-6). IEEE.

Tan, K., & Wang, D. (2018, September). A convolutional recurrent neural network for real-time speech enhancement. In *Interspeech* (Vol. 2018, pp. 3229-3233).

Rance, G. (2005). Auditory neuropathy/dys-synchrony and its perceptual consequences. *Trends in amplification*, *9*(1), 1-43.

Parikh, G., & Loizou, P. C. (2005). The influence of noise on vowel and consonant cues. *The Journal of the Acoustical Society of America*, *118*(6), 3874-3888.

Rao, A., & Carney, L. H. (2014). Speech enhancement for listeners with hearing loss based on a model for vowel coding in the auditory midbrain. *IEEE Transactions on Biomedical Engineering*, *61*(7), 2081-2091.

Fant, G. (1960). Acoustic theory of speech production, s'-Gravenhage. *Mouton and Co*.

Lindblom, B. (1986). Phonetic universals in vowel systems. *Experimental phonology*, 13-44.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of phonetics*, *17*(1), 3-45.

Engstrand, O., & Krull, D. (1991). Effect of inventory size on the distribution of vowels in the formant space: preliminary data for seven languages. *PERILUS*, *13*, 15-18.

Livijn, P. (2000). Acoustic distribution of vowels in differently sized inventories–hot spots or adaptive dispersion. *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS)*, *11*.

Al-Tamimi, J. E., & Ferragne, E. (2005). Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. In *9th European Conference on Speech, Communication and Technology (Interspeech 2005)*. Newcastle University.

Carney, L. H., & McDonough, J. M. (2019). Nonlinear auditory models yield new insights into representations of vowels. *Attention, Perception, & Psychophysics*, *81*, 1034-1046.

Carney, L. H., Li, T., & McDonough, J. M. (2015). Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations. *Eneuro*, *2*(4).

Young, E. D., & Sachs, M. B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. The Journal of the Acoustical Society of America, 66(5), 1381-1403.

Miller, R. L., Schilling, J. R., Franck, K. R., & Young, E. D. (1997). Effects of acoustic trauma on the representation of the vowel/ɛ/in cat auditory nerve fibers. *The Journal of the Acoustical Society of America*, *101*(6), 3602-3616.

Liberman, M. C., & Brown, M. C. (1986). Physiology and anatomy of single olivocochlear neurons in the cat. *Hearing research*, *24*(1), 17-36.

Brown, M. C. (2014). Single-unit labeling of medial olivocochlear neurons: the cochlear frequency map for efferent axons. *Journal of neurophysiology*, *111*(11), 2177-2186.

Holube, Inga, Volkmar Hamacher, and MEAD C. Killion. "Multi-channel compression: Concepts and (early but timeless) results." *Hearing Review* 23.2 (2016): 20-26.

Bor, S., Souza, P., & Wright, R. (2008). Multichannel compression: Effects of reduced spectral contrast on vowel identification.

Plomp, R. (1988). The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *The Journal of the Acoustical Society of America*, *83*(6), 2322-2327.

Stone, Michael A., and Brian CJ Moore. "Effects of spectro-temporal modulation changes produced by multi-channel compression on intelligibility in a competing-speech task." *The Journal of the Acoustical Society of America* 123.2 (2008): 1063-1076.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., ... & Ludvigsen, C. (1994). An international comparison of long-term average speech spectra. *The journal of the acoustical society of America*, *96*(4), 2108-2120.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *the Journal of the Acoustical Society of America*, *67*(3), 971-995.

Carney, L. H., Cameron, D. A., Kinast, K. B., Feld, C. E., Schwarz, D. M., Leong, U. C., & McDonough, J. M. (2023). Effects of sensorineural hearing loss on formant-frequency discrimination: Measurements and models. *Hearing Research*, *435*, 108788.

Chapter 2. Subcortical Auditory Model including Efferent Dynamic Gain Control with Inputs from Cochlear Nucleus and Inferior Colliculus

This chapter has been submitted for publication in Journal of Acoustic Society of America (JASA).

2.1 Abstract

We developed an auditory model with a time-varying, gain-control signal based on the physiology of the efferent system and the sub-cortical neural pathways. The medial olivocochlear (MOC) efferent stage of the model receives excitatory projections from both fluctuation-sensitive model neurons of the inferior colliculus (IC) and wide-dynamic-range model neurons of the cochlear nucleus. The response of the model MOC stage dynamically controls the gain of simulated cochlear outer hair cells. In response to amplitude-modulated (AM) noise, firing rates of most IC neurons with band-enhanced modulation transfer functions in awake rabbits increase over a time course consistent with the dynamics of the MOC efferent feedback. These changes in the rates of IC neurons in awake rabbits were employed to adjust the parameters of the efferent stage of the proposed model. Responses of the proposed model to AM noise were able to simulate the increasing IC rate over time, while the model without the efferent system did not show this trend. The proposed model with efferent gain control provides a powerful tool for testing hypotheses, shedding insight on mechanisms in hearing, specifically those involving the efferent system.

2.2 Introduction

The descending auditory pathway, or efferent system, is as large as the ascending pathway in terms of number of projections, but this part of the system not been as extensively studied as the ascending pathway. The efferent system consists of several feedback projections (Guinan, 2006; Warr, 1992) spanning all levels of the auditory pathway, from the auditory cortex to the cochlea. The focus of this study is the medial olivocochlear (MOC) system, which includes projections from MOC neurons to the outer hair cells, which directly control cochlear gain. Although several studies suggest the critical role of the MOC efferent system for interpreting data from psychoacoustic measures (Lopez-Poveda, 2018; Jennings, 2021; Lauer et al., 2022), it is still unclear the extent to which the gain control by the MOC efferent system is involved in hearing. Understanding the role of the efferent system in auditory processing is essential to gain a comprehensive understanding of auditory function. In this study, we develop a computational model for the sub-cortical auditory system, providing a tool for exploring efferent mechanisms of cochlear gain control.

Despite several studies of the physiology of the efferent system (Guinan, 2006; Gummer et al., 1988; Liberman & Guinan, 1998), many aspects of this system are not well understood, such as the effects of anesthesia, the role of higher-level projections, and the independent effects of the MOC and middle-ear muscle reflexes. This gap in understanding is potentially explained by its complexity and the limited experimental methods for understanding the efferent system (Schofield, 2011). Most of the information available is based on a relatively small number of recordings from single neurons located deep in the

brainstem in anesthetized animals (Guinan, 2018); however, anesthesia suppresses MOC responses (Aedo et al., 2015; Chambers et al., 2012; Guitton et al., 2004). Noninvasive methods used to study MOC effects on cochlear gain include measurements of either otoacoustic emissions (OAEs) (Guinan, 2006) or the cochlear microphonic (Jennings & Aviles, 2023; Jamos et al., 2020, 2021; Lichtenhan et al., 2016; Smith et al., 2017), with and without MOC activation by electrical stimulation or contralateral broadband noise. These methods provide useful insights about the efferent system; however, they do not provide information about higher-level projections in the pathway, such as those from the IC or auditory cortex to the MOC (Delano et al., 2007). Furthermore, care must be taken to separate effects of the middle-ear muscle reflex on OAEs (Sun, 2008, Guinan, 2018). Here we propose a model of the MOC efferent system that can help in designing future physiological and psychophysical experiments for studying efferent function.

A new framework for auditory coding proposes an explicit role for the efferent system. Carney (2018) suggests that low-frequency temporal fluctuations in auditory-nerve (AN) responses can be a reliable cue for auditory perception. It is often assumed that AN average-rate profiles encode the spectrum of complex sounds; however, the majority of AN fibers, the high-spontaneous-rate (HSR) fibers, have average rates that are saturated at conversational speech levels (55–65 dB SPL, Olsen, 1998). Therefore, in many conditions, average discharge rates of HSR AN fibers do not contain enough information to explain perception of complex sounds (Bharadwaj et al., 2014; Carney, 2018; Liberman, 1978). On the other hand, temporal fluctuations of HSR AN responses, referred to as neural fluctuations, have robust information at conversational speech levels, potentially encoding

peaks and valleys in the stimulus spectrum because the fluctuation depths are smaller for AN fibers tuned near peak frequencies and larger for other frequency channels (Carney et al., 2015; Carney, 2018). If neural fluctuations play a key role in encoding complex sounds, then these fluctuations should be involved in the efferent gain-control system to maintain contrasts in the neural-fluctuation profile. Interestingly, neurons in the auditory midbrain (inferior colliculus, IC) are sensitive to low-frequency (i.e., <100Hz) fluctuations of their inputs, and the IC projects to MOC neurons (Joris et al., 2004; Huffman & Henson, 1990; Krishna & Semple, 2000).

The MOC efferent system receives several convergent inputs; the model presented here focuses on subcortical pathways, specifically inputs from the IC and cochlear nucleus

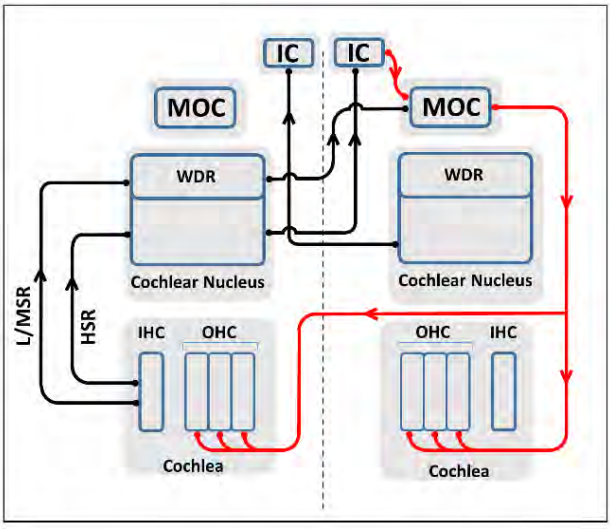

FIG. 2.1. (color online) Schematic illustration of the peripheral afferent (black arrows) and MOC efferent (red arrows) pathways. Here for a simple presentation we have shown the afferent pathways on the left side and the efferent pathways on the right side, but anatomically both pathways exist on both sides of the brainstem.

(CN). MOC neurons receive descending projections from IC cells in the midbrain and exhibit band-pass modulation transfer functions (MTFs) that are possibly inherited from these inputs (Gummer et al., 1988a; Huffman & Henson, 1990; Schofield, 2011). MOC neurons also receive ascending projections from cells types in the CN that have wide dynamic ranges (WDR), including T-stellate cells (Blackburn & Sachs, 1990; Guinan, 2018; Romero & Trussell, 2021) and neurons in the small cell cap, which receive low- and medium-spontaneous-rate AN inputs (Ye et al., 2000; Fekete et al., 1984) (Fig. 2.1). For the proposed model, IC projections to the MOC provide neural-fluctuation information to the MOC gain-control system, as follows: 1) IC neurons that are excited by fluctuations (band-enhanced (BE) IC cells) reduce cochlear gain in response to fluctuating inputs, 2) this reduction in cochlear gain increases AN fluctuation depth by reducing inner-hair-cell saturation, 3) this increase in AN fluctuation depth results in a further increase in IC activity until the gain stops decreasing, either because a minimum cochlear gain is reached or when fluctuation-driven feedback is balanced by WDR driven feedback (described below). This fluctuation-driven feedback loop, in which IC activity ultimately reinforces IC activity, is effectively a positive feedback loop. Note that, a decrease in AN fluctuation depth results in positive feedback in the opposite direction, that is a reduction in IC responses ultimately leads to a further reduction in IC responses.

The positive-feedback mechanism described above could contribute to spectral coding of complex sounds, as follows (see Carney, 2018): 1) AN frequency channels tuned near local peaks in the stimulus spectrum have responses that are “captured” by the spectral peak, and thus have relatively small fluctuations (Deng et al., 1987; Delgutte and Kiang,

1984; Young and Sachs, 1979). Consequently, for these frequency channels, BE IC cells would have lower response rates. Activity of MOC neurons reduces cochlear gain in a frequency/place-specific manner (Guinan, 2018); therefore, input to the MOC from BE IC channels near a spectral peak would be relatively small, as compared to that from channels driven by large fluctuations. The relatively higher cochlear gain for channels near spectral peaks would drive IHCs closer to saturation, resulting in further reduction in the neural fluctuations in AN HSR responses over time (i.e., positive feedback). 2) For frequency channels tuned away from spectral peaks, AN fluctuations are larger and BE IC cells therefore have higher response rates. The higher rates in the BE IC input to the MOC would result in a reduction in cochlear gain, reducing IHC saturation and further increasing fluctuation amplitudes in the AN response. Overall, the IC input to MOC neurons provides a positive feedback that would enhance the fluctuation contrast across AN channels tuned near spectral peaks, and thus the rate contrast across IC channels. This feedback potentially enhances the neural representation of spectral peaks and valleys at the level of the IC (see Figs. 2 and 3 in Carney, 2018).

In the proposed model, the other input to the MOC stage is from the WDR neurons in the CN, which carry the stimulus sound-level information. This level-dependent input to the MOC acts as a negative feedback system, i.e., an increase in stimulus level would increase activity of the WDR neurons which excite MOC neurons, ultimately reducing cochlear gain. This negative feedback is hypothesized to keep the operating point of the saturating IHC transducer function near its knee point, to ensure sufficient baseline fluctuation contrast across frequency channels. The IC feedback signal, described above,

would then enhance this baseline contrast. If there were no contribution to gain control from the WDR cells, the AN channels could all have fluctuating or flat responses when exposed to stimuli with low or high levels, respectively, regardless of the peaks and valleys in the stimulus spectrum. Both extremes would reduce the contrast in fluctuation depth across IC frequency channels.

According to the neural-fluctuation profile framework and the feedback mechanisms of the MOC efferent system described above, sustained fluctuations in a stimulus could cause a progressive decrease in cochlear gain over time. As a result, in response to amplitude-modulated (AM) noise, the average rate of BE IC neurons is expected to increase over a time course of 100-200 ms due to the MOC-related reduction in cochlear gain. The rationale for this prediction is that the low-frequency fluctuations of AN fiber responses to modulated noise stimuli excite IC neurons with BE MTFs. BE IC neurons are assumed to excite MOC cells, and MOC activity decreases cochlear gain. Reduction in cochlear gain effectively decreases the slope of the IHC transduction nonlinearity and therefore increases the effective modulation depth at the output of the IHC (Fig. 2.2), ultimately increasing the BE IC response. Thus, IC rates would increase over time until the efferent control system is stabilized. Interestingly, re-examination of IC responses to AM noise in awake rabbits (Carney et al., 2014) supported this hypothesis and revealed rates that increased over a time course that is consistent with the dynamics of

MOC efferent gain control (Farhadi et al., 2021). This increase in IC rate is an unexpected effect in neural responses, which typically adapt and therefore exhibit a decrease, rather than increase, in the neural rate over time. We quantified the increase in IC average rate over the time course of responses to AM noise, and used that information to tune the parameters of the MOC efferent model proposed in this work.

Here, we refine and test an AN and subcortical model with both fluctuation- and wide-dynamic-range-driven gain control (Farhadi et al., 2021). We first introduce the novel control system used to implement the MOC efferent pathway. Next, we describe how the parameters of the MOC efferent model were adjusted based on a physiological dataset from a previous study of IC responses to AM noise from awake rabbits (Carney et al., 2014). We then show that the model with efferents outperforms a model without efferents in simulating the dynamics of physiological responses in the IC response to AM noise stimuli.

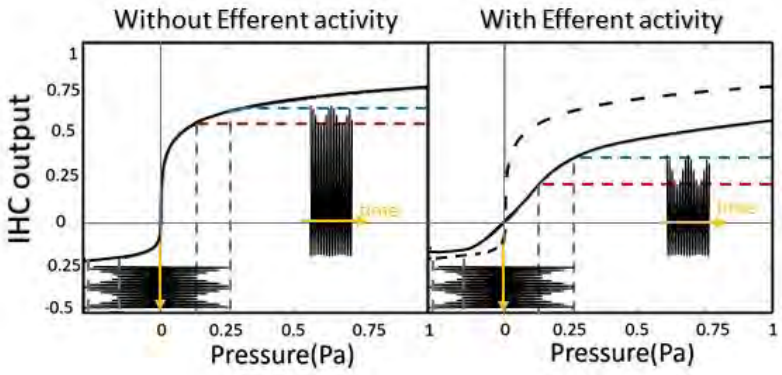

FIG.2. 2. (color online) Effect of MOC efferent activity on IHC input-output function. Left panel: The effective modulation depth in IHC output voltage is limited due to the saturation. Right panel: with the same input pressure, effective modulation depth is increased due to the MOC efferent activity and decrease in gain.

This finding supports the hypothesis that the efferent system is potentially the underlying mechanism for the increase in rate of IC cells over the time course of AM noise. Implications of the proposed model for responses to other sounds, and interactions with other neural mechanisms involved in long-term changes in response rates, will be discussed.

2.3 Methods

We adapted existing models that were designed to simulate responses of AN fibers (Zilany et al., 2009, 2014) and IC cells (Mao et al., 2013) to implement the proposed model with MOC efferent gain control. First, we briefly introduce the original AN and IC models. Then, we describe the modifications that were applied, including changes to the structure of the existing models and new extensions added to simulate the MOC efferent system. Finally, we introduce the methods used to estimate the MOC efferent-system model parameters.

2.3.1 Existing models

For the original Zilany et al. (2009, 2014) AN model, the stimulus first passes through a filter that simulates the middle-ear transfer function. The output of the middle-ear filter is the input to the cochlear-model stage (Fig. 2.3). Cochlear tuning is modeled using two filters, a relatively sharp chirping filter in the signal path tuned at the characteristic frequency (CF) and a wider bandpass filter in the feedforward control path tuned slightly higher than CF (Zhang et al., 2001; Tan & Carney, 2003; Zilany & Bruce, 2006). The output of the control-path filter and OHC model stage adjusts the time constant (and thus the gain and bandwidth) of the chirping filter. The cochlear-gain

parameter (C_{OHC}) in this model can be varied from 1 (normal OHC function) to 0 (completely impaired OHC function) to simulate hearing loss related to OHC impairment. The IHC is modeled with a nonlinear transduction function followed by a low-pass filter. Hearing loss from decreased IHC sensitivity is implemented by varying the C_{IHC} parameter between 1 (normal IHC function) and 0 (completely impaired IHC function).

We used the BE IC model from Mao et al. (2013), which consisted of a band-pass filter centered at the best modulation frequency (BMF) of the IC cell. The response of a simulated HSR AN model was the input to the IC model; this structure is a simplified

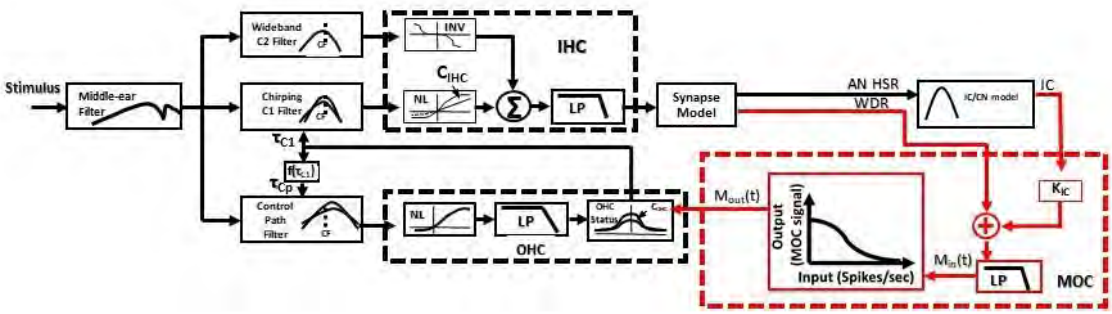

FIG. 2.3. (color online) Schematic illustrating overall structure of the new model components (red) with respect to the Zilany et al. (2014) AN model and Mao et al. (2013) IC/CN model (black). After (Zilany et al., 2009) and (Farhadi et al., 2021). The new model components include the two feedback signals from IC and WDR cells to the MOC unit (dashed red box), where the two input signals are scaled and summed before going through a first-order low-pass filter. The low-pass filtered signal goes to the rational input-output MOC function, which maps the input signal to a cochlear-gain scaler between 0 and 1, for that time sample.

version of an IC model for AM tuning that also included a CN stage (Nelson and Carney, 2004).

2.3.2 Proposed model with MOC efferents

One of the challenges of including the efferent system in the computational AN model was to reconfigure the implementation of the model to work sample by sample. In previous IC model implementations (e.g., Mao et al., 2013), the AN model output for the entire stimulus waveform served as the input to the IC stage. In the sample-by-sample configuration of the proposed efferent model (Fig. 2.3), a new value of COHC(t) was applied to each sample to adjust cochlear gain based on the output of a new stage: the model MOC (Fig. 2.3, dashed red box). The dynamics of the MOC system were modeled with a first-order low-pass filter, thus the output of the MOC model was a function of the current and previous time samples of the MOC input signal, which was a combination of the IC-model rate function (time-varying rate, in spikes/sec) and the CN WDR response provided by a LSR AN-model rate-function. For simplicity, we used the LSR AN model response to represent all WDR inputs to the MOC, such as the T-stellate cells in the anteroventral cochlear nucleus (Oertel et al., 2011; Romero & Trussell, 2021) and cells in the small cell cap of the CN that receive exclusive inputs from LSR and MSR AN fibers (Fekete et al., 1984; Ryugo, 2008). In the sample-by-sample implementation, the previous states of all the filters of the AN model were preserved in order to process each time sample of the simulated model response.

In the proposed model, the output of the MOC efferent stage controlled the gain parameter based on the current and previous samples of the neural signals at the input to

this stage. Specifically, the output of the MOC stage, a scaler between 0 to 1, was multiplied by the OHC gain parameter (COHC) to modulate the cochlear gain. This multiplication occurred for every time sample. The model with this new sample-by-sample configuration was tested to ensure that it matched the response of the original model when the efferent stage was deactivated.

The two inputs to the proposed MOC model (Fig. 2.3, red) were the descending IC feedback signal and the ascending WDR signal, which were combined to determine the degree of cochlear gain reduction. Specifically, the excitatory input to the MOC stage was the sum of the IC and the WDR rate functions. A constant scalar, KIC, was applied to the IC rate signal to approximately balance the amplitudes of the two inputs to the MOC (Table 4.1). The MOC system was modeled with a low-pass filter representing dynamics of the MOC efferent system, followed by a decaying rational function (Eqn. 1) that mapped the MOC inputs to an output signal that modulated the cochlear-gain parameter, C_{OHC}, on a sample-by-sample basis. The rational function, chosen to be second-order to ensure asymptotic behavior for both small and large input signals, was described by

$$M_{out}(t) = \frac{1}{1 + [B \times \{M_{in}(t)\}^2]} \quad , \quad (Eqn. 1)$$

where $M\_in\ (t)$ is the sum of the responses from the WDR and IC paths to the MOC unit after low-pass filtering, and B is a constant that adjusts the shape of MOC input-output

function (Fig. 2.3). The optimization of the value of parameter B is discussed later (Section C of the Methods). Figure 2.4 shows the intermediate responses at each level of the

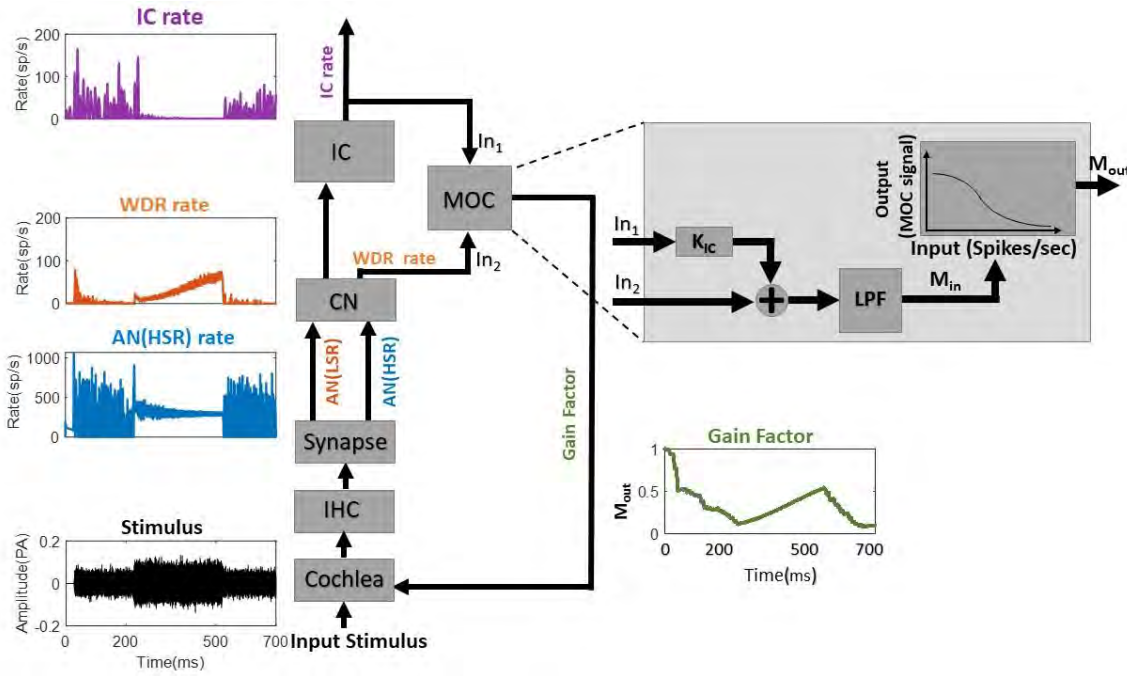

FIG. 2.4. (color online) The MOC efferent model diagram with intermediate responses at different levels of auditory periphery in response to tone-in-noise stimulus. The stimulus (black) is initially a wide band noise; after 200 ms a tone is added to the noise for a duration of 300 ms. The response of the model AN (HSR) stage (blue) shows that during the noise-only stimulus (first 200 ms) the fluctuation depth increases over time as the gain decreases (green), and as a result the IC (purple) rate increases. After the addition of the tone (200 -500 ms) the fluctuation depths in the AN HSR response is reduced, resulting in a drop in IC response rate. As the In1 input to the MOC block (on the right side of this figure) decreases, the gain starts to recover (green). After the offset of the tone (500-700 ms) the gain decreases again, and the AN fluctuations and IC rate increase.

composite model to a tone-in-noise stimulus. The first 200 ms of the stimulus is noise only (Fig. 2.4, black). The BE IC response to this initial noise is robust (Fig. 2.4, purple); because of the fluctuation in AN HSR rates (Fig. 2.4, blue), thus driving the MOC model and decreasing the cochlear gain (Fig. 2,4, green). Between 200 and 500 ms, a tone at the model CF (5 kHz) was added to the noise, resulting in saturation of the IHC (not shown) and flattening the neural fluctuations in the AN HSR response. As a result of the reduced AN HSR neural fluctuations, the BE IC rate response decreased. The reduced IC rate resulted in a weaker MOC input and thus a relative increase in cochlear gain (Fig. 2.4, green). Note that the mean rate of the AN HSR fiber was not changed appreciably by the addition of the tone, due to rate saturation; however, the model WDR rate function increased during presentation of the tone. After the tone ended, the IHC was no longer saturated, AN fluctuations reappeared, and the IC response increased. The increase in the IC response excited the MOC stage, which resulted in a reduction in cochlear gain. The changes in the WDR response rate (Fig.2.4, red) also influenced cochlear gain; as described above, the WDR signal maintains the IHC operating point such that the IC response has an effect on cochlear gain. In contrast to the CF channel tuned to the tone frequency (Fig. 2.4), AN fibers tuned away from the tone would have responses that were dominated by the noise masker; at moderate sound levels, these responses would have strong fluctuations. In summary, a flat AN response flanked by fluctuating responses provides a cue for detecting a tone in noise. Model code will be made freely available as downloadable files.

2.3.3 Adjusting the parameters of the MOC efferent model

Parameters of the MOC efferent model include those of the MOC input/output function, *B* (Eqn.1, Fig.2.4), the scaler constant for IC rate (K_{IC}), and the low-pass filter (LPF) time constant (cut-off frequency). We adjusted these parameters based on an existing physiological dataset that contained IC responses in awake rabbits to AM wideband noise (Carney et al., 2014).

Stimuli

The stimulus for both the physiological dataset and the model simulation during the parameter fits was a sinusoidally amplitude-modulated (SAM) wideband noise (100 Hz to 20 kHz) with modulation depths that varied from -30 to 0 dB (fully modulated) in 5 dB steps, and an unmodulated wideband noise with the same bandwidth. The modulation frequency was matched to the best modulation frequency (BMF) of the recorded cells, and was in a ranging from 50 to 100 Hz. The stimulus spectrum level was 36 dB re 20μPa (RMS: 76 dB SPL overall level). The duration of the stimuli was 500 ms, including 50-ms raised-cosine ramps.

Physiological data and analysis

The physiological dataset included previously published extracellular, tetrode recordings of responses in the central nucleus of the IC in three awake rabbits. Stimuli were presented to the contralateral ear. Details of the physiological methods can be found in Carney et al. (2014). Briefly, IC cells were classified based on their MTFs into BE, band-suppressed (BS), hybrid, flat and unusual types (Kim et al., 2015, 2020) (Fig. 2.5).

BE IC cells are excited over a band of modulation frequencies, whereas IC BS cells are suppressed over a band of modulation frequencies (Kim et al., 2020). Hybrid cells have rates that are enhanced over one band of modulation frequencies, and suppressed over another range (Kim et al., 2020). The focus of this study was on BE cells because this IC MTF type was hypothesized by Gummer et al., (1988b) to project to MOC neurons, based on the BE MTFs reported for MOC neurons. MTF shapes were classified similar to Kim et al. (2020), using a criterion ratio of 1.2 peak MTF rate to the reference rate response to an unmodulated wideband noise which yielded approximately equal numbers

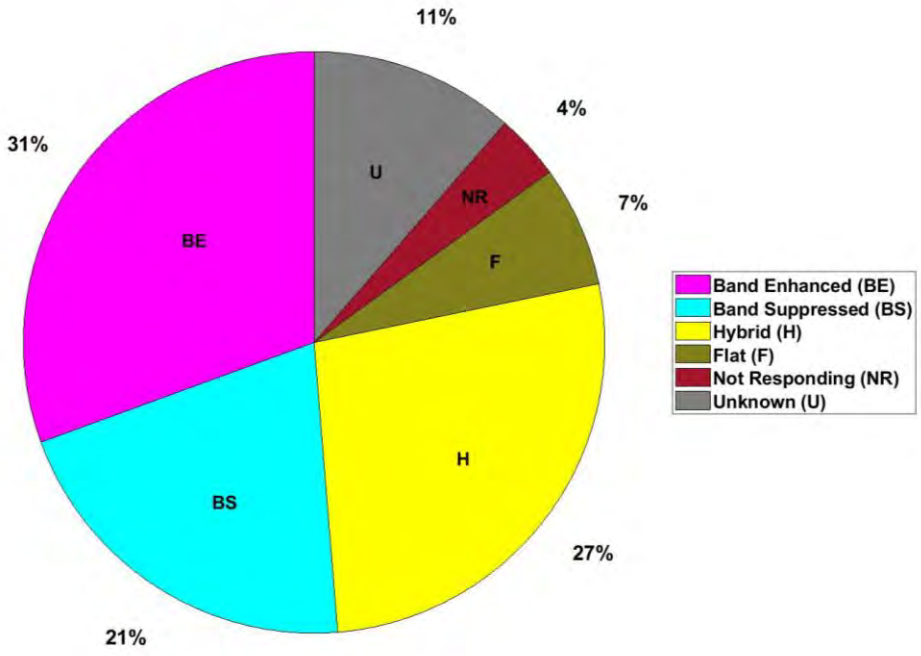

FIG. 2.5. (color online) Distribution of MTF classes of the IC neurons in the physiological dataset (Carney et al., 2014) using the MTF-categorization method of Kim et al. (2020) (D. O. Kim et al., 2020)

of BE and BS MTFs (Fig. 2.5). The frequency of the peak rate in the MTF identifies the cell's BMF. The physiological responses used in this study had stimulus modulation frequency that was within an octave of the recorded cell's BMF.

To analyze the physiological responses to AM noise, a Gaussian filter was used to smooth the spike trains to avoid aliasing introduced by PSTH bin-widths (Lehky, 2010). The standard deviation of the Gaussian filter was inversely proportional to each cell's BMF; the number of points in the filter was six times the standard deviation of the Gaussian window. Responses were averaged over 50 repetitions of the stimulus.

Trends in firing rate over the course of the response, after an initial 200 ms which included the onset response, were characterized using a simple linear regression. Based on the efferent model framework, the BE IC rates were hypothesized to increase over time as a result of reduction in cochlear gain and increase in the temporal fluctuation depths in AN responses. The sign of the linear model's slope indicated whether the response increased or decreased over time. BE IC responses with rates that increased over time were used to find the parameters in the MOC efferent model by maximizing the correlation between the BE IC spiking rate and the model responses to a stimulus with a modulation depth of -10 dB. A modulation depth of -10 dB was chosen here because the majority of the IC neurons had strong increases in rate over time at this modulation depth as compared to higher modulation depths, for which IC rates sometimes saturated. A brute-force method was used to search a reasonable range of parameters and find a set of parameters values that maximized the correlation between the model response and the physiological data over the analysis window. Parameters were estimated for each neuron in the dataset. We

simultaneously fit the time constant, τ_{MOC}, the rational function parameter, B (Eqn. 1), and the IC constant scaler, K_{IC}. A 300-ms -duration analysis window was used for both physiological and model responses. The analysis window started 200 ms after stimulus onset and continued until the end of stimulus. Properties of the model such as CF, BMF, and modulation frequency were matched to each cell in the physiological dataset. Based on the parameter value fits for each of the 33 neurons in the dataset, a distribution of the parameters was established, and the means of the distributions were chosen for the final model parameter values. The sensitivity of the model to the parameter values was studied by comparing the average correlation between the physiological and the model responses when each model parameter was either set to the value fit to each neuron or to the mean parameter value for the population.

2.4 Results

Section A below reviews our hypothesis regarding the MOC efferent model response to AM noise stimuli, analysis of the physiological dataset in response to AM noise, and implementation of the MOC efferent model with its parameters (KIC, B, and τMOC). Section B describes the distribution of the optimum parameters in the dataset and compares the responses of models with and without the MOC efferent control system to physiological data. In section C we use the new model with MOC efferent control to simulate responses to a few different stimuli and compare these responses with physiological data from the literature.

2.4.1. The physiological dataset

Consistent with our hypothesis, the IC rates increased over time in most IC neurons with BE MTFs. Figure 2.6 presents six examples of BE IC neuron responses that illustrate the trend of increasing rate over time; these examples were chosen as they have different BMFs and CFs. The left panel in each example in Fig. 2.6 is the rate MTF, and the right panel is the IC response to AM noise modulated at a frequency near the neuron's BMF, with a range of modulation depths (different colored lines). The gray time range is the time

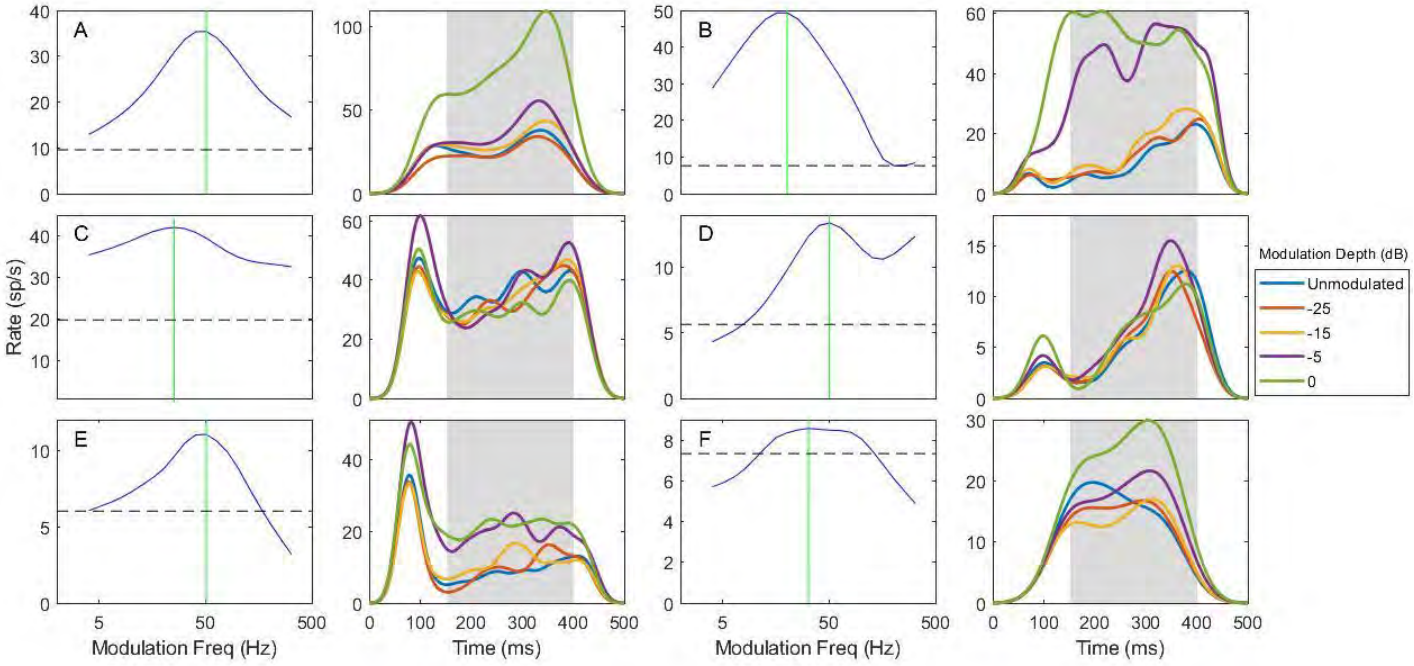

FIG. 2.6. (color online) A-F) Six examples of BE IC neuron smoothed responses to AM noise stimulus. The left-hand panel in each example is the MTF for AM noise stimuli; the vertical green lines indicate BMF. The right-hand panels show smoothed PSTHs that illustrate the time course of the rate during responses to AM noise stimuli with a modulation frequency near each neuron's BMF and a range of modulation depths (see legend).

window after the onset response over which we analyzed the increase in rate. All of the examples in Fig. 2.6 had increasing rate over time during the analysis window, especially in response to lower modulation depths. In some neurons (e.g., Fig. 2.6B) the rate was nearly constant over time in response to fully modulated noise, but the IC rate was high for this modulation depth from the beginning of the analysis window. In the framework of our hypothesis, this observation suggests that because of the strong modulation in the stimulus, the fluctuation in AN PSTHs was large enough to maximize the IC rate, such that a change in cochlear gain as a result of MOC efferent activity could not further increase the IC rate. In these examples, the PSTHs in response to the unmodulated noise stimulus also showed an increasing trend in rate. This trend in the noise responses could also be explained by efferent-system activity, because unmodulated wideband noise includes inherent temporal fluctuations after peripheral filtering, which would increase IC rate and lower the cochlear gain.

The trend of increasing rate over time (See Methods Section C) occurred in approximately 80% of neurons with BE MTFs in response to modulation depths less than -15 dB (Fig. 2.7). This percentage decreased at high modulation depths (see Fig. 2.6). As described above, the decrease at high modulation depths can be explained by a ceiling effect on the IC rate. In general, neural response rates are typically expected to decrease over time during sustained stimuli, as a result of rate adaptation, making this increase in rate particularly striking. The finding that 80% of the BE IC cells had rates that increased over the time course of the AM noise stimuli is consistent with the hypothesis that MOC efferent activity results in an increase in BE IC rates over time.

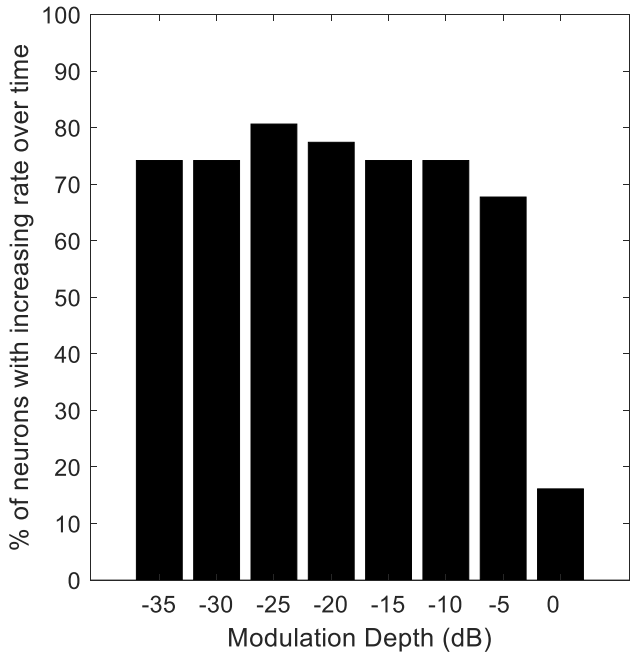

FIG. 2.7. Percentage BE IC neurons with increasing rate over the time course of wideband AM noise stimuli.

2.4.2. Adjusting model parameters

The distribution of the optimized parameters for neurons in this study was very broad (Fig.2.8). The median of the distribution for each parameter was used as the final value for that parameter (K_{IC}=30, τ_{MOC}=260 ms, B=0.01) (Eqn.1). The median of the estimated time constant distribution was 260 ms, which was used to determine the cutoff frequency of the LPF in the MOC's model, τ_{MOC}. This value agrees with values of the time constant that have been suggested for the MOC efferent system based on other types of physiological and psychophysical measurements (Backus & Guinan, 2006; Roverud & Strickland, 2010; Warren & Liberman, 1989).

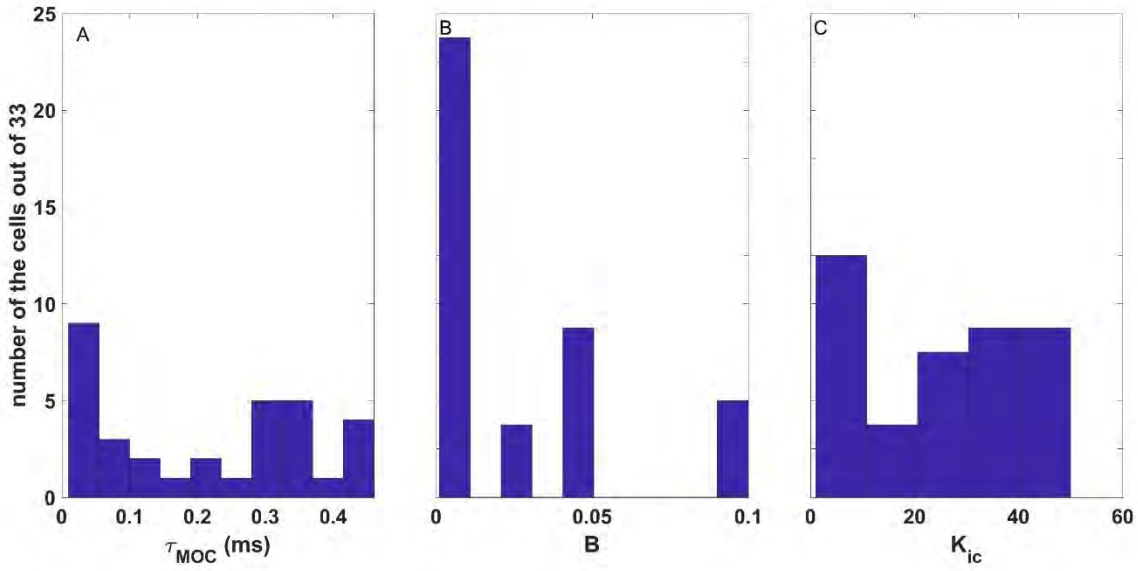

FIG. 2.8. (color online) Distribution of the fit values of, A) time constant (τ_{MOC}), B) the rational function parameter (B), and C) the IC scaler (K_{IC}) for the model for 29 neurons in this study for which the rates increased during the SAM stimuli, the MTF type was BE, and the modulating frequency of the stimulus matched the BMF of the cell.

The median correlation between the model and physiological responses during the analysis time window was calculated for models with and without the efferent feedback system (Fig. 2.9). The low median correlation of 0.025 between data and the model without efferent response (Fig. 2.9-A) demonstrates that the model without efferents did not simulate the increasing trend in rate over time in the physiological dataset; instead, the response of that model showed a decrease in rate consistent with neural adaptation that arises in the AN model responses (Zilany et al., 2014). Simulations using a MOC efferent model with parameters identified for each neuron resulted in a correlation of 0.98 between

the data and model responses (Fig. 2.9-B). The correlation reduced to 0.68 after replacing the individual values with the median values of the distributions estimated for each parameter, and some of the correlations in this condition are negative, consistent with the wide distribution of the optimum parameters for each neuron (Fig. 2.9-C). This result shows that the model was sensitive to the parameter values, but still simulated the increased rate in response to AM noise using median parameters values.

Figure 2.10 shows responses to AM noise for three representative BE IC cells (left column), for models without MOC efferents (middle column), and for the model with efferents, using the mean values for the model parameters (right column). The models had BMFs and CFs that matched the example neurons. The model with MOC efferent feedback simulated the increase in rate over time, consistent with the hypothesis that a model with

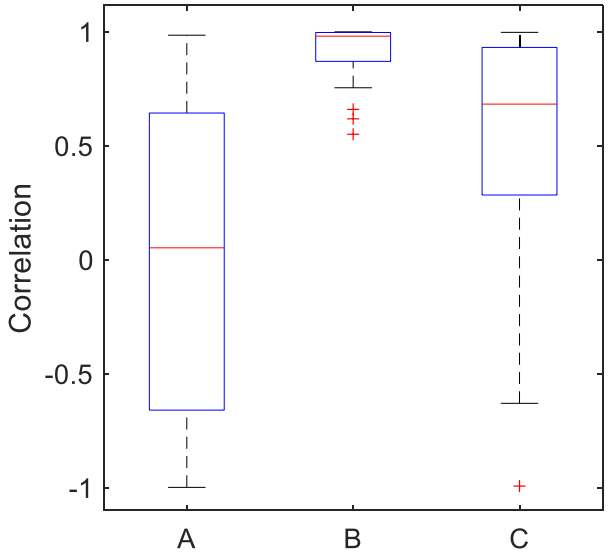

FIG. 2.9. (color online) Correlations between the physiology and models A) without efferents, B) with efferents, with all parameters individually identified for each neuron, and C) with efferent parameters fixed at the median of the distributions.

the MOC efferent system could simulate the response to AM noise more accurately than the model without efferents.

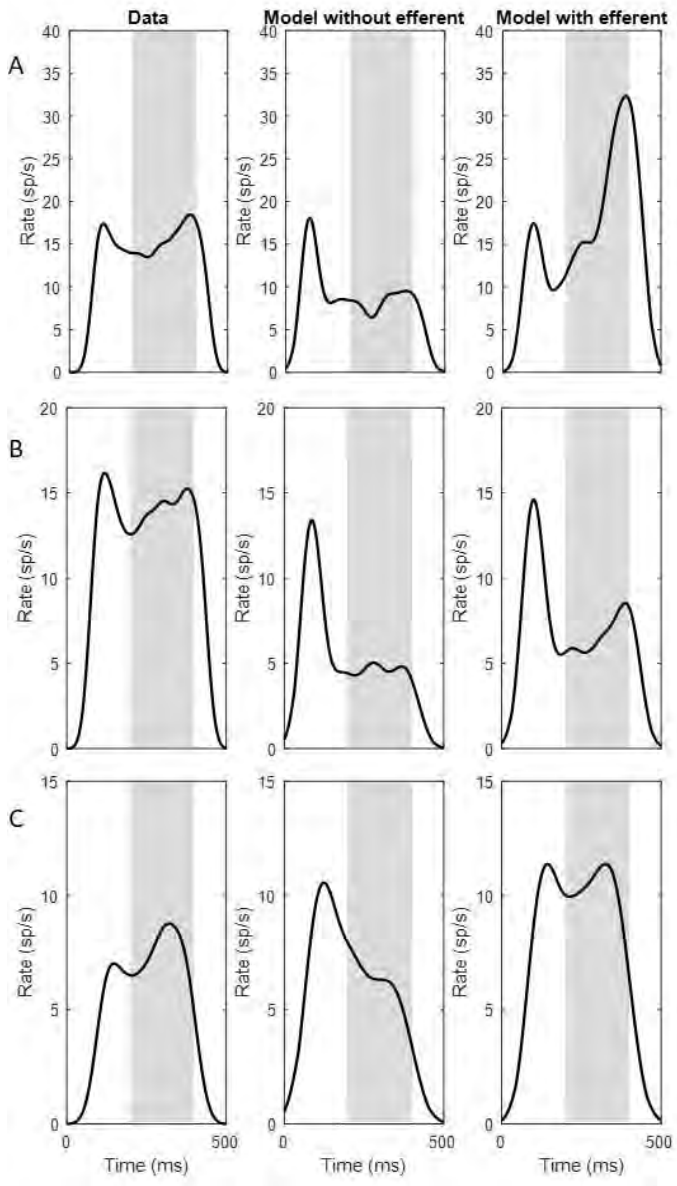

FIG. 10. Physiological data from a BE IC cell in awake rabbit, and IC model responses to AM noise with modulation depth of -10 dB without and with efferent activation for three example neurons (A,B, and C). The CF and BMF of the model were matched to the example neuron (BMF is 50 Hz and CF is 3650 Hz in this example neuron) the median values for the model parameters were used (Fig. 9-C).

2.4.3. Validating the proposed model responses

Model responses to tone stimuli were compared to responses of the previous model to verify that introducing the efferent gain-control pathways did not adversely affect the model's ability to predict AN responses to basic stimuli. It is important to note that the physiological data used for comparison were recorded from ANs of anesthetized animals, in which the MOC efferent system was likely suppressed (Guitton et al., 2004). Thus, complete agreement between the physiological responses of AN fibers in anesthetized animals and the responses of the AN stage of the model with MOC efferents was not expected, especially for complex stimuli (such as AM tones). However, for pure-tone stimuli, Rhode and Kettner (1987) and May and Sachs (1992) showed that anesthesia does not significantly affect post-stimulus time histograms (PSTHs) or rate-level functions of primary-like neurons in the cochlear nucleus (and one presumed AN fiber). Therefore, we validated the proposed model by comparing simulated responses to some baseline stimuli in physiological examples.

Model Response to pure tones and recovery of spontaneous rate

Figure 2.11 shows data and simulated responses to a 500-ms tone at CF followed by 500-ms of silence. The purpose of this task was to study both the tone response and the AN model's recovery to spontaneous rate. Recovery of spontaneous activity depends on stimulus level and on the spontaneous rate of the fiber (Kiang, 1965). Here we compare model responses to physiological recordings from two HSR fibers with CFs of 1.82 and 10.34 kHz from Kiang (1965) (Fig. 2.11). The model response during the tone is comparable to the physiological examples, and the model shows the same initial pause and

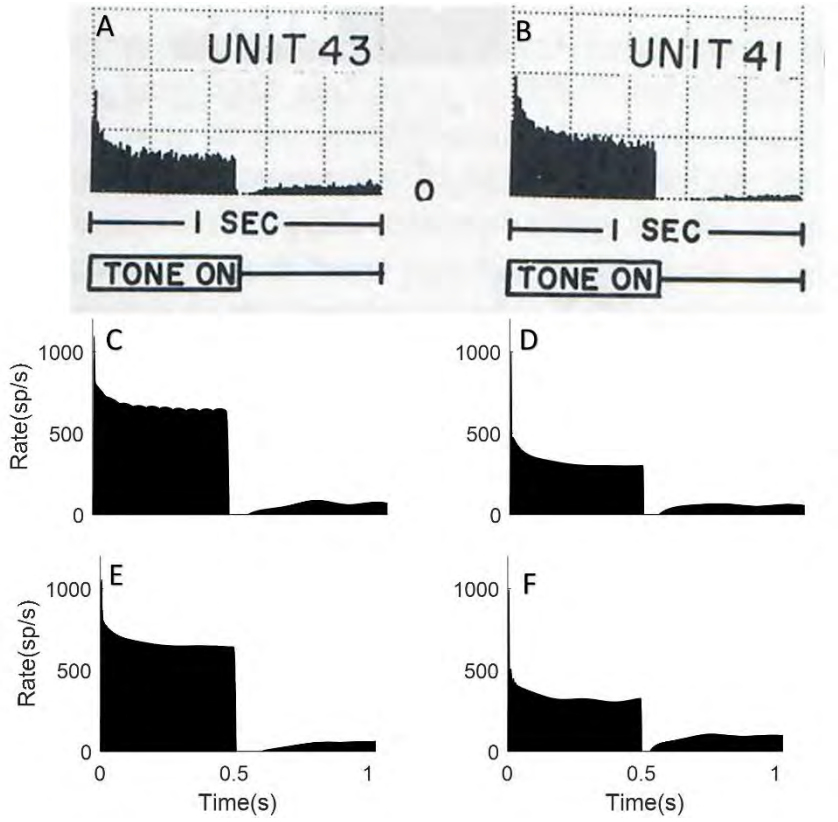

FIG. 2.11. AN and model responses to tones at CF: A) Unit 43 : CF = 1.82 kHz, 55 dB SPL, SR = 60.6 s/sec; B) Unit 41: CF = 10.34 kHz, 30 dB SPL, SR = 51.7 s/sec; C, D) Responses to the same tones for the model without efferents, and E, F) with MOC efferents. For all model simulations (C to F) the AN model type was HSR with CFs matching the AN fibers in the physiological study. The stimulus sound level was matched to stimuli for the physiological data from Kiang (1965, permission requested.)

time course of recovery after the stimulus as in the physiological responses. Note that the response to a pure tone would primarily activate the CN input to MOC efferent pathway. The IC stage of the model would mainly respond at the onset and offset of the tone, when the onset/offset ramps form a fluctuation in the stimulus envelope and therefore elicit a response from the fluctuation-sensitive IC neurons.

Model response to an increment in tone level

One of the paradigms tested in Zilany et al. (2009) was the model response to an increment in tone level. The purpose of this example is to evaluate the response to an increment in a tonal pedestal. In the physiological data (Westerman & Smith, 1987) the AN fiber response to a tone at the fiber's CF (5.99 kHz, HSR) was recorded in anesthetized cat. Over the first 300 ms of the stimulus, the tone level was 5, 10, 15, or 20 dB above threshold, and for the second 300 ms, the tone level increased to 43 dB above threshold. In the physiological responses, the transient response to the pedestal increased with pedestal level, but the transient response to the increment decreased with increasing pedestal level. The same trend was observed in the response of the AN model with and without MOC efferents. The increase in rate over time that followed the first transient response was observed for the model with MOC efferents (Fig. 2.12C, for initial tone levels of +5 and +10 dB), but not for physiological responses in the anesthetized gerbil (Fig. 2.12A). This discrepancy between the MOC model predictions and the physiological data may be due to the effects of anesthesia on the empirical AN recordings. The increase in rate during the low-level tones (Fig. 2.12-c, +5 and +10 dB) reflects the recovery of cochlear gain after an initial decrease in response to the onset of the stimulus. Responses in unanesthetized animals are not available for these stimuli, therefore, it is not possible to confirm this interpretation at this time.

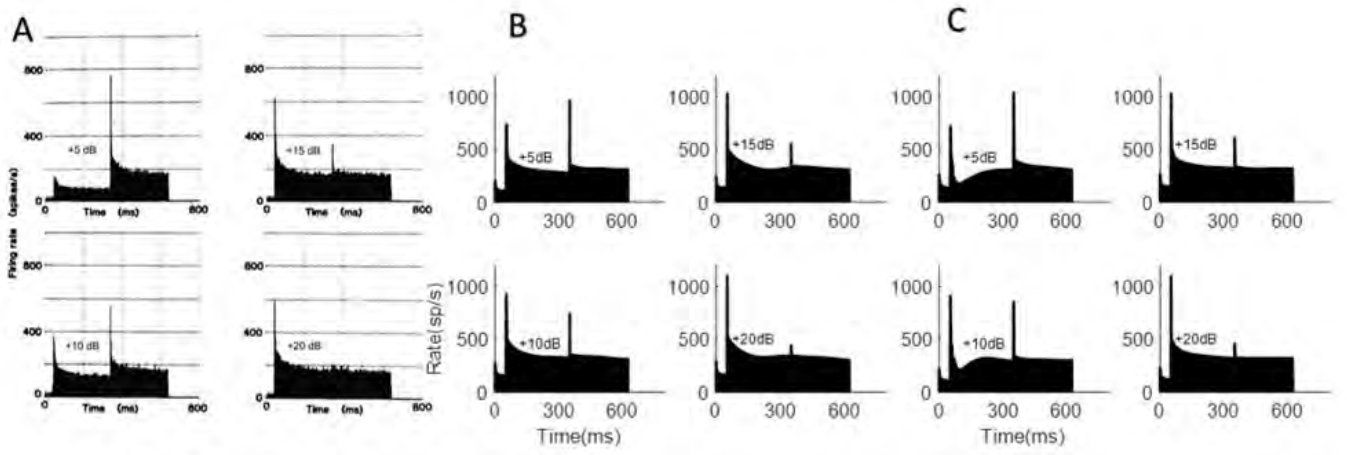

FIG. 2.12. A) Actual neuron physiological data from gerbil from Westerman and Smith (1987, with permission) in response to increase in tone level. The AN model response to the amplitude increment response paradigm B) without efferent system C) with MOC efferent system

Model response to AM stimuli

Here the model response was compared to the response of an HSR AN fiber with CF of 20.2 kHz to an AM tone with a carrier frequency at CF (Joris & Yin, 1992). The effect of modulation depth on the synchrony to modulation frequency was studied for modulation depths ranging from 0 to 1. The modulation frequency was 100 Hz and level was 49 dB SPL. A comparison of synchronization to AM frequency as a function of modulation depth showed that the model with MOC efferents had increased synchrony with modulation depth. However, the synchrony of the model response was weaker than that observed physiologically because of the reduction in cochlear gain as a result of MOC efferent activity. Complications for comparison exist due to the difference between dynamic ranges

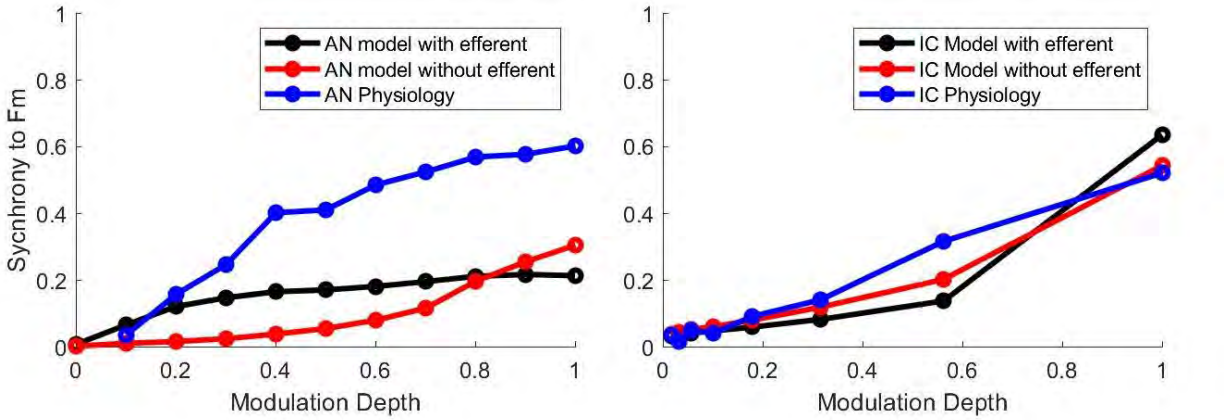

FIG. 2.13. (color online) A) Synchrony to the envelope of an AM tone at CF for an AN fiber with CF = 20 kHz (Joris & Yin, 1992; permission requested) and for an HSR model fiber with matching CF with and without efferent activation. Modulation frequency of 100 Hz; level = 49 dB SPL. B) IC model with and without activation of the MOC efferent system and IC physiology synchronization-modulation depth plots for responses to AM noise stimulus at CF 5330 Hz, BMF 80 Hz, modulation frequency 64 Hz and stimulus level 36 dB SPL.

of the model and this physiological example, as noted by Zilany et al. (2009) (Fig. 2.13-A). Furthermore, model responses to AM tones, which consist of both carrier and sideband frequencies, may be shaped by the effective bandwidths of MOC projections, discussed below. AM responses of the model with efferents, however, cannot be directly compared to physiological responses, which are not available for AN fibers in unanesthetized animals. However, we do have access to IC responses to AM noise in awake rabbit. Figure 13 B compares the synchrony to AM noise for the IC stage of the model with MOC efferents to actual responses from an IC cell in awake rabbit (Carney et al., 2014). This result confirms that, for the model with efferents, the increase in synchrony in the AN

model with increasing modulation depth results in appropriate synchrony at the level of the model IC.

2.5. Discussion

In this work, we modified an existing model of the peripheral auditory system to simulate both the ascending pathway and MOC efferent feedback with both CN and IC inputs. The physiological data and simulation results are consistent with the hypothesis that the MOC efferent system influences neural fluctuations over time (Carney, 2018), as observed here by the increasing responses of BE IC cells over the time course of an AM noise. The simulation results support the hypothesis that a sub-cortical model with MOC efferent gain control better simulates IC responses to AM noise compared to a model without efferents.

One of the components of the proposed model is a LPF that was implemented to simulate the dynamics of the MOC efferent system. The time constant (or cut-off frequency) of this filter was adjusted based on the physiological dataset. The dynamics of the MOC efferent system based on the fit parameter agree with the reported dynamics in the literature for the MOC efferent fast effect. Backus and Guinan (2006) used stimulus-frequency otoacoustic emission (SFOAE) measurements to estimate the human MOC-reflex time constant. They reported an overall onset time constant of 277±62 ms and decay time constant of 159±54 ms. Warren and Liberman (1989) estimated the MOC response to have a time constant of 250 ms in barbiturate-anesthetized cats. Roverud and Strickland (2010) used an optimized psychophysical technique to estimate the time course of cochlear-gain reduction. They estimated an average exponential decay of 100

ms, which is faster than the time constant estimated by Backus and Guinan. Yasin et al. (2020) used a computational model of human hearing that included MOC efferent feedback and evaluated the model's speech recognition for different MOC efferent time constants. Their results suggest the improvement in recognition of speech in AM noise is highest for a MOC efferent time constant of 200 ms. In general, all of the time constants suggested in the literature for the MOC efferent system are in the same range as our results based on rabbit midbrain physiology. However, the MOC efferent system time constant could differ between rabbits and humans. Using the proposed MOC efferent model to simulate human psychoacoustic experiments could better inform our understanding of the time constant for human listeners.

The dynamics of the efferent system support its potential role in improving speech perception (Yasin et al., 2018). Specifically, the latency and time constant of the MOC efferent system (Backus & Guinan, 2006; Roverud & Strickland, 2010; Warren III & Liberman, 1989) are comparable to typical phonemic durations (Jacewicz et al., 2007; Umeda, 1977). Therefore, the MOC efferent system would presumably affect auditory neural responses, both within the time course of phonemes and also in running speech, for which the perception of a phoneme would be influenced by the previous phoneme (co-articulation) (Chambers et al., 2017; Recasens, 1984, 2018).

Another important component of the proposed model MOC efferent system was the input-output function that maps the MOC input to a gain factor that scales the cochlear-gain coefficient between 0 and 1. The second-order rational function used in this model was chosen based on its asymptotic behavior at low and high input levels. The direct

input-output function of MOC neurons is not known, but several related functions have been reported. The shape of the proposed input-output function was similar to the MOC efferent attenuation as a function of input sound level shown in Russell and Murugasu (1997). Also, the asymptotic trend of this function for higher values is similar to the asymptotic responses shown in Liberman (1988) for MOC efferent activity as a function of input sound level. Finally, the proposed function is similar to the relationship between stimulus level and OHC power gain, which would be modulated by MOC signals, in an electrical-mechanical model for OHCs (Rabbitt & Bidone, 2023; Rabbitt, 2023).

Comparing the Proposed model with previous AN models that included the efferent system

Several previous AN models include efferent signals driven by the periphery (Clark et al., 2012; Giguere & Woodland, 1994; Kwan et al., 2019; Smalt et al., 2014; Yasin et al., 2020). In some of these previous implementations, cochlear gain was set to remain constant over time, omitting the dynamics of MOC efferents (Brown et al., 2010; Ferry & Meddis, 2007; Giguere & Woodland, 1994; Jennings et al., 2011). Smalt et al. (2014) proposed a model based on the Zilany et al. (2014) AN model, but with an efferent system modeled by a time-varying gain-control feedback system. However, in this model, a local gain-control system was implemented based on the output of model OHCs, which is inconsistent with the overall structure of the MOC efferent system, which depends on outputs from the OHCs, IHC, AN, CN, IC, and other structures (Smalt et al., 2014; Schofield, 2011). None of these models simulated the descending input to the MOC efferent system from the midbrain. The focus of the current study was to include both

peripheral and midbrain inputs to a dynamic MOC efferent system in a model for AN and BE IC responses.

Limitations and Future directions

Other physiological and psychoacoustic studies can be used to further fine tune the model's parameters and structure. For example, forward-masking studies with long-duration maskers would be expected to activate the MOC efferent system (Krull and Strickland, 2008; Brennan et al., 2023). At the level of the IC, further animal studies of the temporal properties of responses to complex sounds would contribute to this modeling effort, such as investigating the changes in synchrony of IC responses to the envelope of AM stimuli over time. Another direction for future work is to study the influence of efferent activity on the responses of other MTF types in the IC, such as BS or Hybrid MTFs. Here we focused on BE MTFs, as their increase in their rate over the time course of AM stimuli cannot be confused with rate adaptation.

The peripheral model studied here is a monaural model, with “one-sided” brainstem and midbrain stages. A binaural model would be a powerful tool for studying the crossed and uncrossed effects of the efferents. The proposed model in this study consists of a single CF channel, and MOC controls for each frequency influence the gain of the related CF channel. However, studies have shown that some projections from MOC neurons to OHCs have a broad distribution across CF channels (Brown, 2014; Lee & Lewis, 2023). To move towards a more comprehensive model, a future direction of research could be to explore a cross-frequency projection of the MOC control system that affects a range of CFs. A broader distribution of MOC projection could benefit

simulations of the neural response to complex stimuli such AM noise and also speech signals. Another direction for future work on this model is inclusion of a more advanced IC model, instead of the simple bandpass filter used for modulation filtering in BE IC neurons in the current model.

In general, recording from AN fibers requires an anesthetized animal preparation, but anesthesia suppresses MOC efferent activity (Guitton et al., 2004). As a result, there are no examples in the AN physiological literature for comparison to the AN model response with active MOC efferents. However, recording from more accessible levels of the auditory pathway, such as the IC in awake animals, could provide further information regarding the MOC efferent system. Recording from the CN in awake animals or recording from AN fibers in decerebrate animals could also potentially be beneficial for studying the MOC efferent system (Kim et al., 1990a, 1990b; Kawase et al., 1993).

Other phenomena could explain the increase in IC response rates over time in response to AM noise, either in addition to or instead of efferent activity. For instance, adaptation of inhibition (Viemeister & Bacon, 1982) could be an alternative explanation for the increase in rate we observed in IC responses to AM noise. Adaptation of inhibition has been suggested as an explanation for auditory enhancement, in which a target signal embedded in a masker is preceded by the same masker but without the target frequency. Adaptation of inhibition could explain an improvement in the threshold of detecting the tone (Byrne et al., 2011). However, neural mechanisms underlying adaptation of inhibition are not clear. Enhancement effects and its properties, such as time constant and level-dependence, may be further examined using the MOC efferent

model in the future to see if such effects are consistent with MOC feedback. Preliminary results suggest that the subcortical model with efferent feedback can also explain several aspects of auditory enhancement (Farhadi et al., 2023).

Acknowledgements

This work was supported by NIH-R01-DC010813 (AF, LHC), NIH-R01-DC008327 (EAS), and K23 DC014752 (SGJ).

Bibliography

Aedo, C., Tapia, E., Pavez, E., Elgueda, D., Delano, P. H., & Robles, L. (2015). Stronger efferent suppression of cochlear neural potentials by contralateral acoustic stimulation in awake than in anesthetized chinchilla. Frontiers in Systems Neuroscience, 9, 21.

Backus, B. C., & Guinan Jr, J. J. (2006). Time-course of the human medial olivocochlear reflex. The Journal of the Acoustical Society of America, 119(5), 2889–2904.

Bharadwaj, H. M., Verhulst, S., Shaheen, L., Liberman, M. C., & Shinn-Cunningham, B. G. (2014). Cochlear neuropathy and the coding of supra-threshold sound. Frontiers in Systems Neuroscience, 8, 26.

Blackburn, C. C., & Sachs, M. B. (1990). The representations of the steady-state vowel sound/e/in the discharge patterns of cat anteroventral cochlear nucleus neurons. Journal of Neurophysiology, 63(5), 1191–1212.

Brennan, M. A., Svec, A., Farhadi, A., Maxwell, B. N., & Carney, L. H. (2023). Inherent envelope fluctuations in forward masking: Effects of age and hearing loss. The Journal of the Acoustical Society of America, 153(4), 1994-2005.

Brown, M. C. (2014). Single-unit labeling of medial olivocochlear neurons: the cochlear frequency map for efferent axons. *Journal of neurophysiology*, *111*(11), 2177-2186.

Brown, G. J., Ferry, R. T., & Meddis, R. (2010). A computer model of auditory efferent suppression: Implications for the recognition of speech in noise. The Journal of the Acoustical Society of America, 127(2), 943–954. https://doi.org/10.1121/1.3273893

Byrne, A. J., Stellmack, M. A., & Viemeister, N. F. (2011). The enhancement effect: evidence for adaptation of inhibition using a binaural centering task. The Journal of the Acoustical Society of America, 129(4), 2088–2094.

Carney, L. H. (2018). Supra-threshold hearing and fluctuation profiles: implications for sensorineural and hidden hearing loss. Journal of the Association for Research in Otolaryngology, 19(4), 331–352.

Carney, L. H., Li, T., & McDonough, J. M. (2015). Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations. Eneuro, 2(4).

Carney, L. H., Zilany, M. S. A., Huang, N. J., Abrams, K. S., & Idrobo, F. (2014). Suboptimal use of neural information in a mammalian auditory system. Journal of Neuroscience, 34(4), 1306–1313.

Chambers, A. R., Hancock, K. E., Maison, S. F., Liberman, M. C., & Polley, D. B. (2012). Sound-evoked olivocochlear activation in unanesthetized mice. Journal of the Association for Research in Otolaryngology, 13(2), 209–217.

Chambers, C., Akram, S., Adam, V., Pelofi, C., Sahani, M., Shamma, S., & Pressnitzer, D. (2017). Prior context in audition informs binding and shapes simple features. Nature Communications, 8(1), 1–11.

Clark, N. R., Brown, G. J., Jürgens, T., & Meddis, R. (2012). A frequency-selective feedback model of auditory efferent suppression and its implications for the recognition of speech in noise. The Journal of the Acoustical Society of America, 132(3), 1535–1541. https://doi.org/10.1121/1.4742745

Delgutte, B., & Kiang, N. Y. (1984). Speech coding in the auditory nerve: I. Vowel-like sounds. The Journal of the Acoustical Society of America, 75(3), 866-878.

Delano, P. H., Elgueda, D., Hamame, C. M., & Robles, L. (2007). Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas. Journal of Neuroscience, 27(15), 4146-4153.

Deng, L., Geisler, C. D., & Greenberg, S. (1987). Responses of auditory-nerve fibers to multiple-tone complexes. The Journal of the Acoustical Society of America, 82(6), 1989-2000.

Farhadi, A., Agarwalla, S. & Carney, L. H. (2023). A Subcortical Auditory Model With Efferent Gain Control Explains Perceptual Enhancement. Abstract, Association for Research in Otolaryngology, SU198, pg.449.

Farhadi, A., Jennings, S. G., Strickland, E. A., & Carney, L. H. (2021). A Closed-Loop Gain-Control feedback model for the medial efferent system of the descending auditory pathway. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 291-295). IEEE.

Fekete, D. M., Rouiller, E. M., Liberman, M. C., & Ryugo, D. K. (1984). The central projections of intracellularly labeled auditory nerve fibers in cats. Journal of Comparative Neurology, 229(3), 432-450.

Ferry, R. T., & Meddis, R. (2007). A computer model of medial efferent suppression in the mammalian auditory system. The Journal of the Acoustical Society of America, 122(6), 3519–3526. https://doi.org/10.1121/1.2799914

Giguere, C., & Woodland, P. C. (1994). A computational model of the auditory periphery for speech and hearing research. II. Descending paths. Journal of the Acoustical Society of America, 95(1), 343–349. https://doi.org/10.1121/1.408367

Guinan Jr, J. J. (2006). Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans. Ear and Hearing, 27(6), 589–607.

Guinan Jr, J. J. (2018). Olivocochlear efferents: Their action, effects, measurement and uses, and the impact of the new conception of cochlear mechanical responses. Hearing Research, 362, 38–47.

Guitton, M. J., Avan, P., Puel, J.-L., & Bonfils, P. (2004). Medial olivocochlear efferent activity in awake guinea pigs. Neuroreport, 15(9), 1379–1382.

Gummer, M., Yates, G. K., & Johnstone, B. M. (1988a). Modulation transfer function of efferent neurones in the guinea pig cochlea. Hearing Research, 36(1), 41–51. https://doi.org/10.1016/0378-5955(88)90136-0

Gummer, M., Yates, G. K., & Johnstone, B. M. (1988b). Modulation transfer function of efferent neurones in the guinea pig cochlea. Hearing Research, 36(1), 41–51.

Huffman, R. F., & Henson Jr, O. W. (1990). The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus. Brain Research Reviews, 15(3), 295–323.

Jacewicz, E., Fox, R. A., & Salmons, J. (2007). Vowel duration in three American English dialects. American Speech, 82(4), 367–385.

Jamos, A. M., Kaf, W. A., Chertoff, M. E., & Ferraro, J. A. (2020). Human medial olivocochlear reflex: contralateral activation effect on low and high frequency cochlear response. Hearing Research, 389, 107925.

Jamos, A. M., Hosier, B., Davis, S., & Franklin, T. C. (2021). The role of the medial olivocochlear reflex in acceptable noise level in adults. Journal of the American Academy of Audiology, 32(03), 137-143.

Jennings, S. G., & Aviles, E. S. (2023). Middle ear muscle and medial olivocochlear activity inferred from individual human ears via cochlear potentials. The Journal of the Acoustical Society of America, 153(3), 1723-1732.

Jennings, S. G. (2021). The role of the medial olivocochlear reflex in psychophysical masking and intensity resolution in humans: a review. Journal of Neurophysiology, 125(6), 2279–2308.

Jennings, S. G., Heinz, M. G., & Strickland, E. A. (2011). Evaluating adaptation and olivocochlear efferent feedback as potential explanations of psychophysical overshoot. JARO - Journal of the Association for Research in Otolaryngology, 12(3), 345–360. https://doi.org/10.1007/s10162-011-0256-5

Joris, P X, Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. Physiological Reviews, 84(2), 541–577.

Joris, Philip X, & Yin, T. C. T. (1992). Responses to amplitude-modulated tones in the auditory nerve of the cat. The Journal of the Acoustical Society of America, 91(1), 215–232.

Kawase, T., Delgutte, B., & Liberman, M. C. (1993). Antimasking effects of the olivocochlear reflex. II. Enhancement of auditory-nerve response to masked tones. Journal of neurophysiology, 70(6), 2533-2549.

Kiang, N. Y. S., Watanabe, T., Thomas, E. C., & Clark, L. F. (1965). *Discharge patterns of single fibers in the cat's auditory nerve*. Cambridge, MA: MIT Press.

Kim, D. O., Carney, L., & Kuwada, S. (2020). Amplitude modulation transfer functions reveal opposing populations within both the inferior colliculus and medial geniculate body. Journal of Neurophysiology, 124(4), 1198–1215.

Kim, D. O., Zahorik, P., Carney, L. H., Bishop, B. B., & Kuwada, S. (2015). Auditory distance coding in rabbit midbrain neurons and human perception: monaural amplitude modulation depth as a cue. Journal of Neuroscience, 35(13), 5360-5372.

Kim, D. O., Sirianni, J. G., & Chang, S. O. (1990). Responses of DCN-PVCN neurons and auditory nerve fibers in unanesthetized decerebrate cats to AM and pure tones: analysis with autocorrelation/power-spectrum. Hearing research, 45(1-2), 95-113.

Kim, D. O., Chang, S. O., & Sirianni, J. G. (1990). A population study of auditory-nerve fibers in unanesthetized decerebrate cats: Response to pure tones. The Journal of the Acoustical Society of America, 87(4), 1648-1655.

Krishna, B. S., & Semple, M. N. (2000). Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. Journal of Neurophysiology, 84(1), 255–273.

Krull, V., & Strickland, E. A. (2008). The effect of a precursor on growth of forward masking. The Journal of the Acoustical Society of America, 123(6), 4352-4357.

Kwan, T. J. M., Zilany, M. S. A., Davies-Venn, E., & Abdul Wahab, A. K. (2019). Modeling the effects of medial olivocochlear efferent stimulation at the level of the inferior colliculus. Experimental Brain Research, 237(6), 1479–1491. https://doi.org/10.1007/s00221-019-05511-4

Lehky, S. R. (2010). Decoding poisson spike trains by gaussian filtering. Neural Computation, 22(5), 1245–1271.

Lauer, A. M., Jimenez, S. V., & Delano, P. H. (2022). Olivocochlear efferent effects on perception and behavior. Hearing Research, 419, 108207.

Lee, D., & Lewis, J. D. (2023). Inter-Subject Variability in the Dependence of Medial-Olivocochlear Reflex Strength on Noise Bandwidth. Ear and Hearing, 44(3), 544-557.

Liberman, M. C. (1988). Response properties of cochlear efferent neurons: monaural vs. binaural stimulation and the effects of noise. Journal of neurophysiology, 60(5), 1779-1798.

Liberman, M. C. (1978). Auditory-nerve response from cats raised in a low-noise chamber. The Journal of the Acoustical Society of America, 63(2), 442–455.

Liberman, M. C., & Guinan Jr, J. J. (1998). Feedback control of the auditory periphery: anti-masking effects of middle ear muscles vs. olivocochlear efferents. Journal of Communication Disorders, 31(6), 471–482.

Lichtenhan, J. T., Wilson, U. S., Hancock, K. E., & Guinan Jr, J. J. (2016). Medial olivocochlear efferent reflex inhibition of human cochlear nerve responses. Hearing research, 333, 216-224.

Lopez-Poveda, E. A. (2018). Olivocochlear efferents in animals and humans: from anatomy to clinical relevance. Frontiers in neurology, 9, 197.

Mao, J., Vosoughi, A., & Carney, L. H. (2013). Predictions of diotic tone-in-noise detection based on a nonlinear optimal combination of energy, envelope, and fine-structure cues. The Journal of the Acoustical Society of America, 134(1), 396–406.

May, B. J., & Sachs, M. B. (1992). Dynamic range of neural rate responses in the ventral cochlear nucleus of awake cats. Journal of neurophysiology, 68(5), 1589-1602.

Nelson, P. C., & Carney, L. H. (2004). A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. The Journal of the Acoustical Society of America, 116(4), 2173-2186.

Oertel, D., Wright, S., Cao, X.-J., Ferragamo, M., & Bal, R. (2011). The multiple functions of T stellate/multipolar/chopper cells in the ventral cochlear nucleus. Hearing Research, 276(1–2), 61–69.

Olsen, W. O. (1998). Average speech levels and spectra in various speaking/listening conditions. American Journal of Audiology, 7:21-25.

Rabbitt, R. (2023) The MOC efferent system acts as a control parameter for outer hair cell dynamic stability and power output. Abstract, ARO, 46:544.

Rabbitt, R., and Bidone, T.C. (2023) A parametric blueprint for optimum cochlear outer hair cell design, Journal of the Royal Society Interface, 20(199) 20220762.

Rahman, M., Willmore, B. D. B., King, A. J., & Harper, N. S. (2019). A dynamic network model of temporal receptive fields in primary auditory cortex. PLoS Computational Biology, 15(5), e1006618.

Recasens, D. (1984). Vowel-to-vowel coarticulation in Catalan VCV sequences. The Journal of the Acoustical Society of America, 76(6), 1624–1635.

Recasens, D. (2018). Coarticulation. In Oxford Research Encyclopedia of Linguistics. https://doi.org/10.1093/acrefore/9780199384655.013.416.

Rhode, W. S., & Kettner, R. E. (1987). Physiological study of neurons in the dorsal and posteroventral cochlear nucleus of the unanesthetized cat. Journal of neurophysiology, 57(2), 414-442.

Romero, G. E., & Trussell, L. O. (2021). Distinct forms of synaptic plasticity during ascending vs descending control of medial olivocochlear efferent neurons. Elife, 10, e66396.

Roverud, E., & Strickland, E. A. (2010). The time course of cochlear gain reduction measured using a more efficient psychophysical technique. The Journal of the Acoustical Society of America, 128(3), 1203–1214.

Russell, I. J., & Murugasu, E. (1997). Medial efferent inhibition suppresses basilar membrane responses to near characteristic frequency tones of moderate to high intensities. The Journal of the Acoustical Society of America, 102(3), 1734-1738.

Ryugo, D. K. (2008). Projections of low spontaneous rate, high threshold auditory nerve fibers to the small cell cap of the cochlear nucleus in cats. Neuroscience, 154(1), 114–126.

Schofield, B. R. (2011). Central descending auditory pathways. In Auditory and vestibular efferents (pp. 261–290). Springer.

Smalt, C. J., Heinz, M. G., & Strickland, E. A. (2014). Modeling the time-varying and level-dependent effects of the medial olivocochlear reflex in auditory nerve responses. JARO - Journal of the Association for Research in Otolaryngology, 15(2), 159–173. https://doi.org/10.1007/s10162-013-0430-z

Smith, S. B., Lichtenhan, J. T., & Cone, B. K. (2017). Contralateral inhibition of click-and chirp-evoked human compound action potentials. Frontiers in neuroscience, 11, 189.

Sun, X.-M. (2008). Contralateral suppression of distortion product otoacoustic emissions and the middle-ear muscle reflex in human ears. Hearing Research, 237(1–2), 66–75.

Tan, Q., & Carney, L. H. (2003). A phenomenological model for the responses of auditory-nerve fibers. II. Nonlinear tuning with a frequency glide. The Journal of the Acoustical Society of America, 114(4), 2007–2020.

Umeda, N. (1977). Consonant duration in american english. The Journal of the Acoustical Society of America, 61(3), 846–858.

Viemeister, N. F., & Bacon, S. P. (1982). Forward masking by enhanced components in harmonic complexes. The Journal of the Acoustical Society of America, 71(6), 1502–1507.

Warr, W. B. (1992). Organization of olivocochlear efferent systems in mammals. The Mammalian Auditory Pathway: Neuroanatomy, 410–448.

Warren III, E. H., & Liberman, M. C. (1989). Effects of contralateral sound on auditory-nerve responses. I. Contributions of cochlear efferents. Hearing Research, 37(2), 89–104.

Westerman, L. A., & Smith, R. L. (1984). Rapid and short-term adaptation in auditory nerve responses. Hearing Research, 15(3), 249–260.

Westerman, L. A., & Smith, R. L. (1987). Conservation of adapting components in auditory-nerve responses. The Journal of the Acoustical Society of America, 81(3), 680–691.

Yasin, I., Drga, V., Liu, F., Demosthenous, A., & Meddis, R. (2020). Optimizing Speech Recognition Using a Computational Model of Human Hearing: Effect of Noise Type and Efferent Time Constants. IEEE Access, 8, 56711–56719. https://doi.org/10.1109/ACCESS.2020.2981885

Yasin, I., Liu, F., Drga, V., Demosthenous, A., & Meddis, R. (2018). Effect of auditory efferent time-constant duration on speech recognition in noise. The Journal of the Acoustical Society of America, 143, EL112-EL115

Ye, Y., Machado, D. G., & Kim, D. O. (2000). Projection of the marginal shell of the anteroventral cochlear nucleus to olivocochlear neurons in the cat. Journal of Comparative Neurology, 420(1), 127–138.

Young, E. D., & Sachs, M. B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. The Journal of the Acoustical Society of America, 66(5), 1381-1403.

Zhang, X., Heinz, M. G., Bruce, I. C., & Carney, L. H. (2001). A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. The Journal of the Acoustical Society of America, 109(2), 648-670.

Zilany, M. S. A., & Bruce, I. C. (2006). Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. The Journal of the Acoustical Society of America, 120(3), 1446–1466.

Zilany, M. S. A., Bruce, I. C., & Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. The Journal of the Acoustical Society of America, 135(1), 283–286.

Zilany, M. S. A., Bruce, I. C., Nelson, P. C., & Carney, L. H. (2009). A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. The Journal of the Acoustical Society of America, 126(5), 2390–2412.

Chapter 3. Predicting thresholds in an auditory overshoot paradigm using a computational subcortical model with efferent feedback

This chapter is a reformatted proceedings paper that was accepted for presentation at in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 2023 (WASPAA2023).

3.1 Abstract

The threshold for detecting a brief tone in a long-duration gaussian-noise masker is elevated for a tone presented near masker onset compared to a tone later in the masker. This phenomenon, referred to as overshoot, has been hypothesized to reflect the dynamic control of cochlear gain by the medial olivocochlear (MOC) component of the efferent system. In this study, we used a subcortical computational model of the auditory system that included the MOC efferent system to predict listener thresholds in the overshoot paradigm. Our simulation results matched listener thresholds, including the trend of decreasing thresholds as the delay of the tone with respect to noise onset increased from 2 to 200 ms. A model without MOC efferents had the opposite trend, with thresholds increasing in longer delay conditions. These findings support the hypothesis that the MOC efferents are involved in the overshoot effect.

3.2 Introduction

Human listeners with normal hearing are known to outperform both individuals with hearing loss and automatic speech recognition (ASR) devices in hearing speech in noise

[1-3]. Understanding the mechanisms involved in hearing in background noise for the normal-hearing auditory system has significant implications for improving the performance of ASR devices and hearing aids.

Simultaneous masking, for which the presence of one sound interferes with the perception of another sound, is a common focus of hearing research [4-5]. Studying the mechanisms underlying simultaneous masking can lead to more insights into how the auditory system processes complex acoustic stimuli, such as speech in noise [6-7]. The overshoot effect is an aspect of tone detection in simultaneous masking [8]. In the overshoot paradigm, listeners are asked to detect a brief tone in a longer duration noise; the tone occurs either immediately after the onset of the noise or after a varying delay. Detection thresholds show an improvement in performance for tones that are delayed after the onset of the noise masker [8-12]. The improvement in detection threshold for this task is called the overshoot effect. The overshoot effect is measured in different stimulus and hearing conditions.

Despite much research on the overshoot effect, the neural mechanisms underlying overshoot are not fully understood [13]. One suggested mechanism is adaptation of auditory-nerve (AN) fibers [14]. When the masker noise is presented, the AN response adapts, or decreases, while the onset response to an increment or added tone remains strong. Thus, there is a higher proportional response to the tone when it is added to the masker noise after a delay [14]. However, Bacon and Healy [15] argue that neural adaptation of AN fibers cannot fully explain overshoot, as it cannot account for spectral features of the overshoot effect, such as frequency dependency for detection of tones in tonal maskers.

Also, for listeners with normal hearing the amount of enhancement is about 10-15 dB, whereas the magnitude of the temporal effect that can be explained based on neural adaptation is only 3-5 dB [8, 16].

One suggested mechanism for overshoot is the MOC reflex [16-20], a component of the efferent system in the descending auditory pathway [21]. The efferent system comprises various projections that extend from locations spanning from the cochlea to the auditory cortex [22]. Here, we focus on two excitatory projections to the MOC efferent neurons, one from wide-dynamic-range cells in the brainstem that convey level-dependent stimulus information, and the other from IC cells in the midbrain that transmit information about fluctuations in stimuli [21,23-25]. These two signals combine to adjust the activity of MOC neurons, which project directly to the outer hair cells (OHCs) and modify cochlear gain [21]. Previous studies suggest that the time constant of the MOC reflex ranges from 100 to 200 ms [26-28].

Several studies have provided evidence for the effect of the MOC reflex on overshoot. One study [29] investigated the timing that results in the strongest overshoot effect and found that the delay of 100-200 ms resulted in the maximum overshoot effect, consistent with MOC efferent dynamics [26-28]. This finding supports the hypothesis that MOC efferent activity plays a role in the overshoot effect. Another study [30] show that the overshoot effect varies with changes in the noise-masker sound level, with the most significant impact observed for mid-level maskers. Because MOC neurons receive level-dependent information from wide-dynamic-range (WDR) brainstem cells, the level dependence of the overshoot effect could be explained by MOC activity. On the other hand,

the dependence of overshoot on spectral properties reported in other studies [16, 30, 31] could be explained by efferent activity through IC projections to MOC neurons. IC neurons are sensitive to AN response fluctuations. Another interesting study [32] found that cutting the efferent bundle decreased the overshoot effect. Furthermore, several studies [35,22,21] found that hearing loss, whether permanent [19, 20, 33] or temporary, caused by factors such as aspirin intake [35] and noise exposure [36], can reduce overshoot. Although these findings could be related to different mechanisms that are disrupted by hearing loss, they could also indicate the inefficiency of the MOC reflex in listeners with hearing-loss [33] and support its role in overshoot and its effect on cochlear gain.

A computational model that simulates neural responses to arbitrary stimuli is a valuable tool for further understanding the effect of the MOC reflex on the overshoot effect. In this study, we replicated a psychoacoustic overshoot experiment [34] and compared the temporal effect for a peripheral model with and without the MOC reflex. The MOC efferent model used in this work, proposed by Farhadi [37, 38], includes a projection from WDR cells in the cochlear nucleus in the auditory brainstem, as well as a projection from IC cells in the midbrain. The WDR projection is a negative feedback loop that decreases cochlear gain as the stimulus level increases, keeping the inner hair cell (IHC) near the knee point of its nonlinear transduction function. IC cells carry fluctuation-dependent information and work as a positive feedback system that increases the fluctuation depth in AN fiber channels that are fluctuating and flattens AN responses that are less fluctuating [37-39].

The model parameters were optimized by fitting the response of IC band enhanced (BE) cells in awake rabbits to amplitude-modulated noise stimuli. The model was adjusted

to replicate the observed trend of increasing firing rate throughout an AM stimulus, which was not seen in the model without efferent [37,40,41]. This optimized model has been demonstrated to improve the ability to predict human performance in psychophysical tasks [42]. The dynamics of the efferent system were modeled using a low-pass filter with a time constant of 200 ms, which is consistent with studies of efferent system dynamics and falls within the range suggested by psychoacoustic studies as being most relevant to the overshoot effect [26-28].

3.3 Methods

3.3.1 Model

A computational model of the auditory system was used to simulate the neural response of the auditory system to a classic overshoot stimulus. The auditory model used in this study not only simulated the neural pathway in the ascending auditory periphery but also included the MOC efferent feedback control system. The MOC efferent system implemented here was comprised of two projections to MOC neurons: 1) from WDR cells in the brainstem and 2) from IC cells in the auditory midbrain. These two projections both increased the MOC efferent activity which suppressed the cochlear gain for each frequency channel independently. The two feedback signals to the MOC system were summed after scaling, and then were lowpass filtered to simulate the dynamics of the efferent system. Finally, the low-pass-filtered sum of the input signals was the input to the MOC input-output function. The input to the MOC at each time sample was mapped to an output gain factor between zero and one. With this implementation, an increase in the input signal to the MOC resulted in a decrease in the gain factor. This gain factor modulated the cochlear

gain directly. The MOC efferent system adjusted the gain dynamically for each time sample according to the previous inputs to MOC block (For more specifics and details of the peripheral model, see [37]).

The responses of eight characteristic frequencies (CFs), ranging from 2.5 to 6.5 kHz and centered around the target frequency were simulated. Based on the simulated neural responses, the thresholds of the models with and without efferent feedback were estimated. An IC model with a band-enhanced (BE) modulation transfer function (MTF) and best modulation frequency (BMF) of 100 Hz was used to estimate the model threshold. The BMF of 100 was selected based on the median BMF of IC cells in physiological studies [43] and is typical of values used in previous modeling studies using similar methods.

3.3.2. Stimuli

The target stimulus was a 10-ms tone probe with a frequency of 4 kHz. The masker was a wideband noise with a bandwidth of 20 to 40 kHz; the duration of the masker was 400 ms with 5-ms on/off cos2 ramp durations. The estimated thresholds were compared for two conditions: one in which the tone was added to the noise after a short (2-ms) delay, and the other in which the tone was added after a long (200-ms) delay. These two delay conditions were examined in this study to replicate the psychoacoustic task [44]. The tone level was fixed at one value, which varied from 40 to 90 dB SPL, with 10 dB increments, and the masker level was tracked to find the detection threshold.

3.3.3 Procedure

A two-interval task was used to estimate the threshold for detecting the target stimulus. Each trial consisted of two intervals, separated by a 500-ms silent period. One interval contained the target stimulus and the masker noise, and the other interval contained only noise. The decision variable was based on the increase in rates of IC neurons with BE responses during an analysis-time window of 20 ms, temporally centered around the probe tone. The IC responses were summed over CFs. A 70.7% correct criterion was used for threshold. The procedure was repeated for 50 trials to estimate the threshold signal-to-noise ratio (SNR) for each of the two models: one with efferent pathways included and one without efferent pathways.

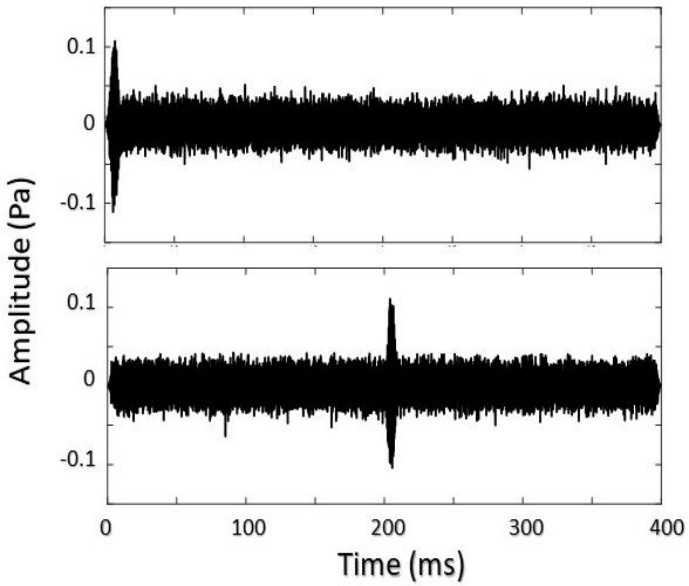

Figure 3.1. Overshoot stimuli with short (top- 2ms) and long (bottom- 200ms) delay of tone onset from noise masker onset

3.4 Results

Figure 3.2 displays the structural schematic of the subcortical model, featuring the medial olivocochlear (MOC) efferent system. The simulated neural responses at various levels of the auditory pathway are presented, including the IC cells (green), high-spontaneous-rate AN fibers (blue), and cochlear gain factor (orange), in response to the stimulus (black). By simulating neural responses at different stages of the auditory pathway, the model enables us to trace the peripheral and central processes that shape the responses to the overshoot stimulus. This representation provides a comprehensive and clear understanding of how the MOC efferent system may influence auditory processing of the overshoot stimulus. The simulated neural responses are shown for the models with and without the efferent system, and for stimuli with short (2-ms) and long (200-ms) delays after the onset of noise. For the short-delay condition, the responses at all levels were similar because the model with efferents did not have enough time to adjust the cochlear gain before tone onset. Therefore, there are no significant differences between the efferent and non-efferent models at the IC level, where the decision variable was applied. Note that the illustrated responses are for a single fiber, for one repetition, and with a CF matched to the tone frequency, while a range of characteristic frequencies were used to evaluate the threshold. The IC rate response to the tone appeared to merge with the response to the onset of the noise for both models, which resulted in a strong peak at the beginning of the IC response. The neural responses to the stimulus with a tone added at a short delay agrees

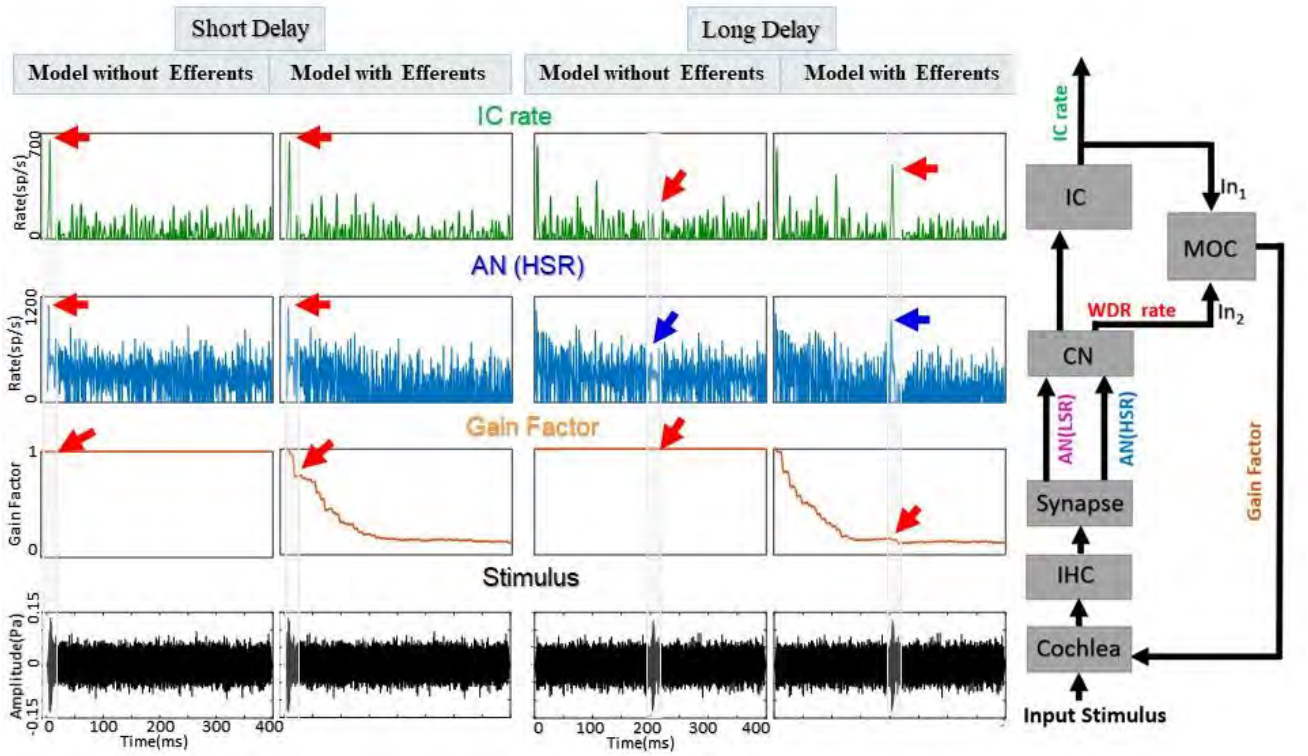

Figure 3.2. the Model structure used for this study with MOC efferent simulation and the simulated neural responses to overshoot stimuli (black), at different stages including cochlear gain factor (orange) AN HSR (blue) IC BE (green) in the subcortical model. For a magnified image of the delayed tone responses (blue arrows) see Figure 3.B.

with our hypothesis that the MOC reflex does not have a significant effect on the short-delay condition.

Figure 3.3 provides a more detailed illustration of the effect of the efferent model on the delayed tone response (cf. blue arrows in Fig. 3.2). In the long-delay condition, efferent activity reduces cochlear gain over the time course of the wideband noise. With the addition of the tone at 200 ms, the model with the efferent pathways responds with a strong peak in the IC rate as a result of the tone onset. For the model without efferent pathways this strong

peak is not observed because the maximum cochlear gain remains constant during the noise masker in the model without efferents. This relatively high cochlear gain in response to a peak in the stimulus results in saturation of the IHC. As a result of IHC saturation, the response of the high-spontaneous-rate AN fiber is flattened. The flattened response observed in the HSR fiber response occurs, despite the modulation in the stimulus, because of the IHC saturation. The flat AN HSR response results in a lower rate in the BE IC response to the tone, as these BE IC cells are excited by low-frequency fluctuation in AN HSR fibers. In contrast, for the model with efferents, since the gain is reduced in response to noise, the IHC is pulled away from saturation, and the modulation in the stimulus results in modulation of the AN fiber response, resulting in a peak in IC BE response.

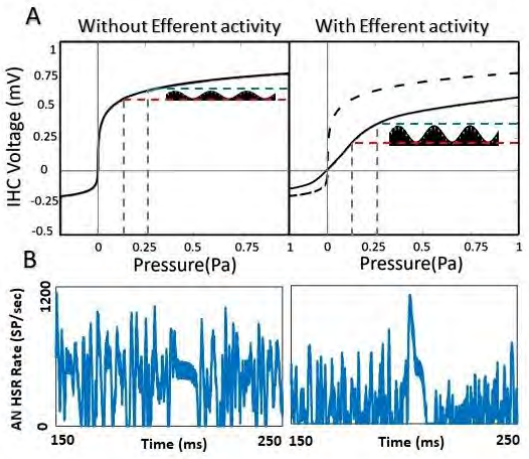

Figure 3.3. A) IHC influence on the effective modulation depth changes with efferent activity. B) Effect of IHC saturation in AN HSR Response to delayed tone without efferent activity on left and with efferent activity on right.

The simulated neural responses, and the intuition behind them, matches the estimated thresholds for both models - with and without efferents. In the short-delay condition, the threshold for psychoacoustic data and both models - with and without efferents - are the same. For the long-delay condition, there is an increase in the noise level at threshold in the physiological data, indicating that listeners can tolerate a higher level of noise, or in other words, their performance is improved. For the model with efferents, the same trend was observed: the noise level at the estimated threshold was higher for the long-delay condition compared to the short delay. This result shows that better performance was estimated using the model with efferents. On the other hand, for the model without efferents, the trend was opposite, such that for the longer-delay stimuli, the noise level at threshold was lower compared to the short delay. This result indicated worse detection performance for the long delay compared to short delay, which is opposite to the threshold trend observed for human listeners in psychoacoustic experiments.

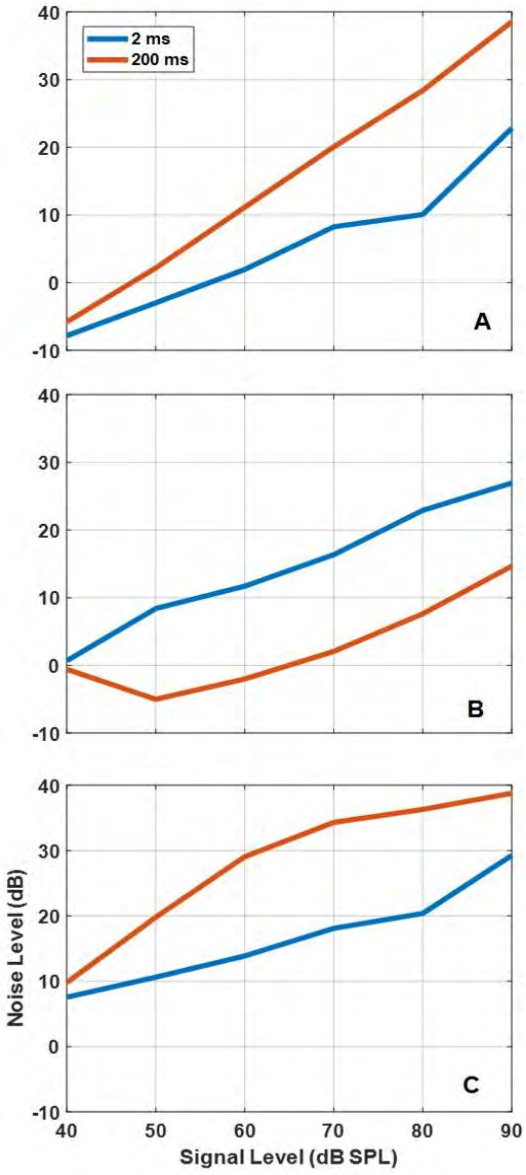

Figure 3.4. A- psychoacoustic thresholds in human listeners thresholds are from Strickland (2004) [44] B) Estimated thresholds for model without efferent. C) estimated thresholds for the model with efferents.

3.5. Discussion

A comparison between simulations using models with and without an efferent system supports the hypothesis that the dynamic gain control by the MOC reflex plays a significant role in simultaneous masking, especially for the overshoot effect. Moreover, the threshold predictions (Fig. 3.4) demonstrated better accuracy for the model with efferents compared to that without. These results show that the model with the MOC reflex, which agrees with anatomical and physiological information, can also be more comprehensive and accurate in simulating psychoacoustic tasks and estimating human behavioral thresholds. This new model can serve as a powerful tool for testing different hypotheses in hearing and lead to more insight about auditory mechanisms.

A future direction for this work could be to use the model with efferents to study the effects of different factors in the overshoot paradigm, such as modulation of the masker (i.e., flat or fluctuating masker), changes in the delay between the masker and tone onsets, and hearing loss, to further test the hypothesis that overshoot is affected by the efferent system.

Acknowledgements

This work was supported by NIH-R01-DC010813.

Bibliography

V. A. Trinh and M. Mandel, "Directly comparing the listening strategies of humans and machines," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 312-323, 2020.

C. Spille, B. Kollmeier, and B. T. Meyer, "Comparing human and automatic speech recognition in simple and complex acoustic scenes," Computer Speech and Language, vol. 52, pp. 123-140, 2018.

O. Scharenborg, "Reaching over the gap: A review of efforts to link human and automatic speech recognition research," Speech Communication, vol. 49, no. 5, pp. 336-347, May 2007.

R. A. Lutfi, "Additivity of simultaneous masking," Journal of the Acoustical Society of America, vol. 73, no. 1, pp. 262-267, Jan. 1983.

A. J. Oxenham and C. A. Shera, "Estimates of human cochlear tuning at low levels using forward and simultaneous masking," Journal of the Association for Research in Otolaryngology, vol. 4, pp. 541-554, 2003.

P. Balazs, B. Laback, G. Eckel, and W. A. Deutsch, "Time-frequency sparsity by removing perceptually irrelevant components using a simple model of simultaneous masking," IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 1, pp. 34-49, Jan. 2009.

K. K. Paliwal and B. T. Lilly, "Auditory masking based acoustic front-end for robust speech recognition," in Proceedings of IEEE TENCON'97. IEEE Region 10 Annual

Conference. Speech and Image Technologies for Computing and Telecommunications, Brisbane, Australia, Dec. 1997, vol. 1, pp. 165-168.

E. Zwicker, "Temporal effects in simultaneous masking and loudness," Journal of the Acoustical Society of America, vol. 38, pp. 132-141, 1965.

E. Zwicker, "Temporal effects in simultaneous masking and loudness," Journal of the Acoustical Society of America, vol. 38, pp. 132-141, 1965.

L. L. Elliott, "Changes in the simultaneous masked threshold of brief tones," Journal of the Acoustical Society of America, vol. 38, no. 5, pp. 738-746, May 1965.

L. L. Elliott, "Development of Auditory Narrow-Band Frequency Contours," Journal of the Acoustical Society of America, vol. 42, no. 1, pp. 143-153, July 1967.

H. Fastl, "Temporal masking effects: I. Broad band noise masker," Acta Acustica united with Acustica, vol. 35, no. 5, pp. 287-302, Sept./Oct. 1976.

S. G. Jennings, "The role of the medial olivocochlear reflex in psychophysical masking and intensity resolution in humans: a review," Journal of Neurophysiology, vol. 125, no. 6, pp. 2279-2308, Jun. 2021.

R. L. Smith, "Adaptation, saturation, and physiological masking in single auditory-nerve fibers," in The Journal of the Acoustical Society of America, vol. 65, no. 1, pp. 166-178, 1979.

S. P. Bacon and E. W. Healy, "Effects of ipsilateral and contralateral precursors on the temporal effect in simultaneous masking with pure tones," in The Journal of the Acoustical Society of America, vol. 107, no. 3, pp. 1589-1597, Mar. 2000.

S. P. Bacon and M. A. Smith, "Spectral, intensive, and temporal factors influencing overshoot," The Quarterly Journal of Experimental Psychology Section A, vol. 43, no. 3, pp. 373-399, 1991.

S. G. Jennings, M. G. Heinz, and E. A. Strickland, "Evaluating adaptation and olivocochlear efferent feedback as potential explanations of psychophysical overshoot," Journal of the Association for Research in Otolaryngology, vol. 12, pp. 345-360, 2011.

S. Schmidt and E. Zwicker, "The effect of masker spectral asymmetry on overshoot in simultaneous masking," in The Journal of the Acoustical Society of America, vol. 89, no. 3, pp. 1324-1330, 1991.

B. S. Bacon and G. A. Takahashi, "Overshoot in normal-hearing and hearing-impaired subjects," in The Journal of the Acoustical Society of America, vol. 91, no. 5, pp. 2865-2871, 1992.

E. A. Strickland and L. A. Krishnan, "The temporal effect in listeners with mild to moderate cochlear hearing impairment," in The Journal of the Acoustical Society of America, vol. 118, no. 5, pp. 3211-3217, 2005.

J. J. Guinan Jr, “Olivocochlear efferents: Their action, effects, measurement and uses, and the impact of the new conception of cochlear mechanical responses,” Hearing Research, vol. 362, pp. 38-47, 2018.

D. L. Weedman and D. K. Ryugo, "Pyramidal cells in primary auditory cortex project to cochlear nucleus in rat," Brain research, vol. 706, no. 1, pp. 97-102, 1996.

M. Gummer, G. K. Yates, and B. M. Johnstone, "Modulation transfer function of efferent neurones in the guinea pig cochlea," Hearing research, vol. 36, no. 1, pp. 41-51, Oct. 1988.

R. F. Huffman and O. W. Henson Jr, “The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus,” Brain research reviews, vol. 15, no. 3, pp. 295-323, Dec. 1990.

B. R. Schofield, "Central descending auditory pathways," in Auditory and Vestibular Efferents, Springer, 2011, pp. 261-290.

B. C. Backus and J. J. Guinan Jr, “Time-course of the human medial olivocochlear reflex,” in The Journal of the Acoustical Society of America, vol. 119, no. 5, pp. 2889-2904, 2006.

W. B. Salloom, H. Bharadwaj, and E. A. Strickland, "The effects of broadband elicitor duration on a psychoacoustic measure of cochlear gain reduction," J. Acoust. Soc. Am., vol. 153, no. 4, pp. 2482-2482, 2023.

E. H. Warren III and M. C. Liberman, “Effects of contralateral sound on auditory-nerve responses. I. Contributions of cochlear efferents,” Hearing Research, vol. 37, no. 2, pp. 89-104, 1989.

D. McFadden, K. P. Walsh, E. G. Pasanen, and E. M. Grenwelge, "Overshoot using very short signal delays," J. Acoust. Soc. Am., vol. 128, no. 4, pp. 1915-1921, Oct. 2010.

S. P. Bacon and S. Savel, "Temporal effects in simultaneous masking with on-and off-frequency noise maskers: Effects of signal frequency and masker level," The Journal of the Acoustical Society of America, vol. 115, no. 4, pp. 1674-1683, Apr. 2004.

S. P. Bacon and N. F. Viemeister, "The temporal course of simultaneous tone-on-tone masking," The Journal of the Acoustical Society of America, vol. 78, no. 4, pp. 1231-1235, Oct. 1985.

F. G. Zeng, K. M. Martino, F. H. Linthicum, and S. D. Soli, "Auditory perception in vestibular neurectomy subjects," Hearing Research, vol. 142, no. 1-2, pp. 102-112, Nov. 2000.

S. G. Jennings, J. B. Ahlstrom, and J. R. Dubno, "Effects of age and hearing loss on overshoot," The Journal of the Acoustical Society of America, vol. 140, no. 4, pp. 2481-2493, Oct. 2016.

S. G. Jennings, M. G. Heinz, and E. A. Strickland, "Evaluating adaptation and olivocochlear efferent feedback as potential explanations of psychophysical overshoot," Journal of the Association for Research in Otolaryngology, vol. 12, pp. 345-360, Sep. 2011.

D. McFadden and C. A. Champlin, "Reductions in overshoot during aspirin use," The Journal of the Acoustical Society of America, vol. 87, no. 6, pp. 2634-2642, Jun. 1990.

C. A. Champlin and D. McFadden, "Reductions in overshoot following intense sound exposures," The Journal of the Acoustical Society of America, vol. 85, no. 5, pp. 2005-2011, May 1989

A. Farhadi, S. G. Jennings, E. A. Strickland, and L. H. Carney, "Auditory-nerve Model including Efferent Dynamic Gain Control with Inputs from Cochlear Nucleus and Inferior Colliculus," bioRxiv, pp. 2022-10, Oct. 2022.

A. Farhadi, S. G. Jennings, E. A. Strickland, and L. H. Carney, "A Closed-Loop Gain-Control feedback model for the medial efferent system of the descending auditory pathway," in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 291-295, IEEE, Jun. 2021.

L. H. Carney, "Supra-threshold hearing and fluctuation profiles: implications for sensorineural and hidden hearing loss," Journal of the Association for Research in Otolaryngology, vol. 19, no. 4, pp. 331-352, 2018.

L. H. Carney, M. S. Zilany, N. J. Huang, K. S. Abrams, and F. Idrobo, "Suboptimal use of neural information in a mammalian auditory system," Journal of Neuroscience, vol. 34, no. 4, pp. 1306-1313, 2014.

M. S. Zilany, I. C. Bruce, and L. H. Carney, “Updated parameters and expanded simulation options for a model of the auditory periphery “Updated parameters and expanded simulation options for a model of the auditory periphery” The Journal of the Acoustical Society of America, 135(1), 283-286, 2014.

M. A. Brennan, A. Svec, A. Farhadi, B. N. Maxwell, and L. H. Carney, "Inherent envelope fluctuations in forward masking: Effects of age and hearing loss," The Journal of the Acoustical Society of America, vol. 153, no. 4, pp. 1994-2005, 2023.

D. O. Kim, L. H. Carney, and S. Kuwada, “Amplitude modulation transfer functions reveal opposing populations within both the inferior colliculus and medial geniculate body” Journal of Neurophysiology, 124(4), 1198-1215, 2020.

E. A. Strickland, "The temporal effect with notched-noise maskers: analysis in terms of input-output functions," J. Acoust. Soc. Am., vol. 115, no. 5, pp. 2234-2245, May 2004.

Chapter 4. Effect of MOC efferent system on neural coding of vowel-liked sounds using a subcortical computational auditory model

4.1 Abstract

Improving speech intelligibility in background noise is a significant challenge in hearing research. Individuals with hearing loss face difficulties in understanding speech in noise mostly due to dysfunction in their auditory peripheral system. Understanding the neural coding to speech sound in the healthy auditory system can be an important guide for improving speech intelligibility and inspire new algorithms for improving hearing devices. In this study, we employed computational modeling to investigate how the healthy hearing auditory system encodes speech sounds. Our focus was on exploring the hypothesis that the medial olivocochlear (MOC) efferent system enhances the neural coding of speech by analyzing neural fluctuation profiles as a robust coding at conversational-level speech. We utilized a subcortical auditory model that integrated within channel IC-driven projections conveying fluctuation information and WDR-driven cross frequency projections incorporating sound level information. We specifically examined vowel-like sounds with single and double formants and assessed the impact of MOC activity on neural coding at different sound and noise levels. Additionally, we compared the robustness of the modulation transfer function (MTF) shape across varying sound levels with and without MOC activity. Our simulation results demonstrate that the MOC efferent system can dynamically enhance neural coding of vowel-liked stimuli by sharpening the contrast in neural fluctuation profiles during 300 ms stimuli at supra-thresholds sound and noise levels. Furthermore, MOC activity helped maintain the MTF shape in the IC model when

sound levels varied. These findings emphasize the potential significance of MOC efferent activity in understanding speech perception in noisy environments.

4.2 Introduction

Understanding speech in background noise is difficult for listeners with hearing loss, even when they use advanced hearing aids in comparison with listeners with healthy hearing. Individuals with normal hearing perform better than even the most advanced speech recognition devices when it comes to dealing with background noise (Spille et al., 2018; Humes et al., 2013, Chou et al., 2019). These differences in speech perception among healthy hearing, impaired hearing and machine hearing, show that the healthy auditory system is developed for processing speech in challenging real-life environments with multiple sources of noise. Understanding how the auditory system accomplishes understanding speech in noise becomes particularly intriguing, as it can inspire designing more impactful hearing-aid devices.

This chapter focuses on neural coding of vowel-like stimuli. Vowels are periodic waveforms produced by vibration of the vocal cords and shaped by resonances of the vocal tract. The vowel spectrum includes peaks that are associated with the harmonics of the fundamental frequency (F0), which relates to the pitch of an individual's voice. Additionally, there are peaks known as vocal formants that are specific to each vowel sound. Frequencies of the first two formants and their relationship, can be used to classify vowels (Fant, 1960). Studying vowels is crucial for understanding the robustness of the healthy auditory system in background noise, as vowels have relatively high energy and their periodicity remains robust in noise.

The neural fluctuation code, proposed by Carney in 2018, provides a robust framework for understanding the neural coding of speech. The neural fluctuating coding relies on the differences in amplitude of low-frequency neural fluctuations in the responses of auditory-nerve (AN) fibers across different frequency channels. Interestingly, despite other codes that rely on average-discharge-rate information versus sound level (Sachs and Young, 1979; Young and Sachs, 1979; Delgutte and Kiang, 1984), neural fluctuations are maintained despite AN saturation at conversational levels (Liberman, 1978; Costalupes et al., 1984; Carney, 2018). The spectral information of vowel sounds can be coded by the neural-fluctuation profile (Carney et al., 2015). In response to a vowel, AN fibers have lower-amplitude neural fluctuations in channels near the formant frequency (Deng et al., 1987; Delgutte and Kiang, 1984; Young and Sachs, 1979) and higher amplitudes in other frequency channels because the inner hair cells (IHCs) in the auditory system become saturated at the formant-frequency channels, which are local maxima in the spectrum. Other channels are not saturated. This contrast in neural fluctuations is then carried to higher levels in the auditory pathway, the brainstem and the auditory midbrain. IC cells in the midbrain are sensitive to the low-frequency fluctuations in their inputs and can code the differences in neural fluctuation amplitude with different average IC response rates (Langner and Schreiner 1988; Krishna and Semple 2000; Nelson and Carney 2004; review: Joris et al., 2004).

Neural fluctuations are a robust cue, and interestingly, can play a role not only in processing of sound in the ascending auditory pathway, but also in the feedback control loop that adjusts cochlear gain in the healthy auditory system (Guinan, 2018; Carney,

2018). The efferent system, consisting of various projections from different stages of the auditory system (Guinan, 2018), is proposed to play a crucial role in speech intelligibility, especially in noise (Brown et al., 2010; Yasin et al., 2020; Martes et al., 2019). In this study, we focus on the medial olivocochlear (MOC) efferent system, which projects to the outer hair cells and adjusts the cochlear gain across different frequency channels (Guinan, 2018). Previous models of the MOC efferent system (Clark et al., 2012; Giguere & Woodland, 1994; Kwan et al., 2019; Smalt et al., 2014; Yasin et al., 2020) have usually only included the projection from the cochlear nucleus (CN) to the MOC, and the MOC efferent system was considered as an automatic gain control system, reducing gain at higher sound levels and increasing gain at lower sound levels. However, studies of the MOC efferent system have shown that there are additional projections to MOC neurons carrying different types of information (Schofield, 2011; Guinan, 2018), thus gain control is not solely determined by sound levels. In this work we will use the Farhadi et al. (2022) subcortical auditory model that incorporates the MOC efferent system, including projections to MOC neurons from both wide dynamic range (WDR) cells in the CN and IC cells in the midbrain. The WDR cells provide sound-level information and the IC projection includes neural-fluctuation information in the AN fiber response, conveying spectral and temporal information about the stimulus.

Our hypothesis for this chapter is that the MOC efferent system's gain adjustment sharpens the neural fluctuation profile by maintain maximal gain at the spectral peaks (e.g., formants in vowel stimuli) and decreasing the gain at other frequencies. These processes enhance the contrast in the neural fluctuation profile, and consequently in the IC rate

profile, between the response to formant frequencies and non-formant frequencies. The rationale behind our hypothesis is that when the spectral peak corresponds to the formant frequency, the IHC becomes saturated, leading to a decrease in neural fluctuation amplitude and a subsequent decrease in the response of band-enhanced (BE) cells in the inferior colliculus (IC) (Fig. 4.1). These IC BE cells are a specific class of IC cells: their rates increase when there is low-frequency fluctuation of their input. IC BE cells are hypothesized to project to MOC neurons. The reason for this argument is that as MOC neurons receive descending projections from IC cells and also these cells have a band-pass modulation transfer functions (MTFs) and hence it is likely that the reason for their band-pass MTF functions be the projection from IC BE cells specifically to these neurons (Gummer et al., 1988a; Huffman & Henson, 1990; Schofield, 2011). The decrease in neural-fluctuation amplitudes in AN fibers tuned near formant frequencies results in a decrease in the IC BE firing rate for formant-frequency channels. Consequently, by decreasing the IC BE rate, the input to the MOC neurons is reduced. This reduction in MOC input decreases MOC activity, which in turn allows the cochlear gain to increase specifically at the formant frequencies, as MOC activity typically suppresses the cochlear gain. For other frequency channels, that are not tuned near spectral peaks, IHCs are not saturated and the temporal neural fluctuations increase the firing rates of IC BE cells. This increased rate further increases MOC activity, resulting in a decrease in the cochlear gain for these frequency channels. This process dynamically enhances the contrast in the neural-fluctuation profile. In summary, the reduction in cochlear gain in non-formant-frequency channels pushes the nonlinearity of the IHC away from saturation, amplifying the neural

fluctuation amplitude even more. As a result, the firing rates of IC BE cells and MOC

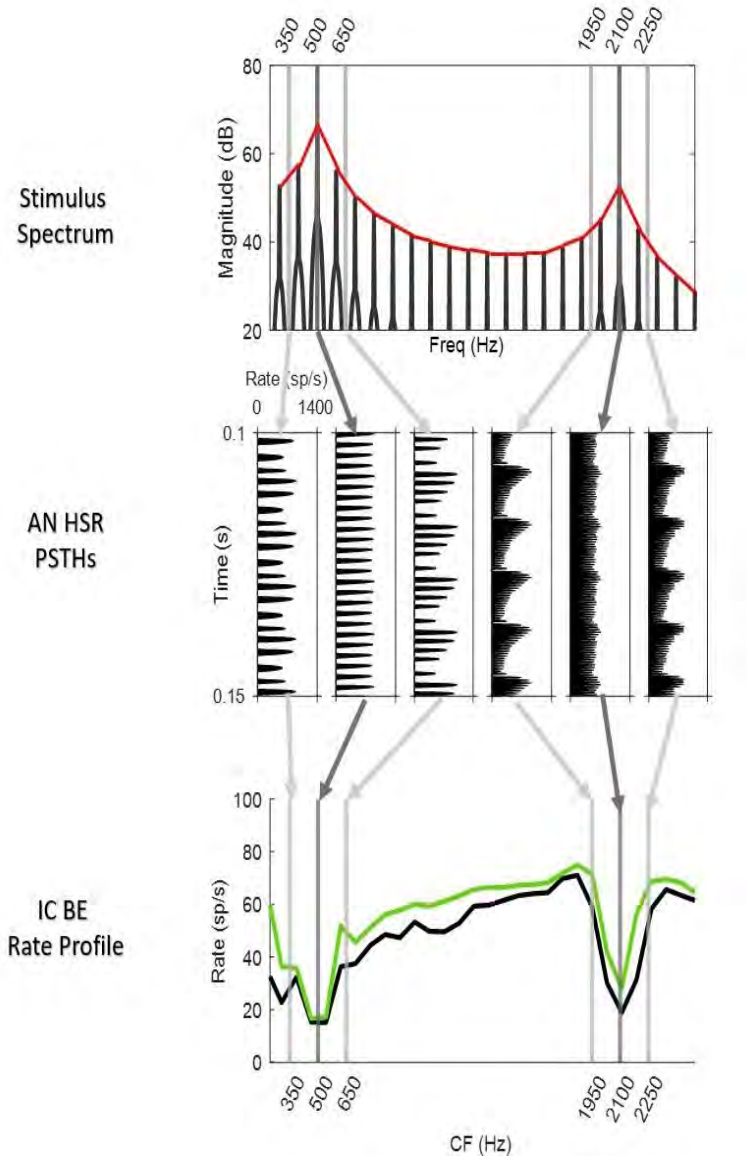

Fig.4.1. Neural fluctuation in responses to a vowel-liked sound of AN fibers and the neural coding of the fluctuations in the profile of response rates of IC BE cells. AN fibers tuned near formant frequencies have relatively flat temporal fluctuations, resulting in a dip in the IC BE response profile. The IC BE response profile is enhanced (green) with activation of MOC efferent.

activity increase, leading to further reduction in gain for these frequency channels. However, for the formant frequency channel, the increased gain pushes the IHC closer to saturation, resulting in even less neural fluctuation amplitude, lower IC BE firing rates, and less gain reduction.

The subcortical model proposed by Farhadi et al. (2022) has demonstrated better performance in predicting various psychoacoustic tasks than models without efferents (Brennan et al., 2023; Farhadi et al., 2023; Farhadi and Carney, 2023; Maxwell et al., 2023). This model was designed as a channel-by-channel system, where the adjustment of cochlear gain in each frequency channel depended only on the projections from the IC and WDR cells associated with that specific frequency. While this channel-by-channel implementation of the MOC system could account for trends in less complex psychoacoustic and physiological datasets, it poses challenges when applied to more complex wideband stimuli, such as speech. For wideband stimuli such as speech, a channel-by-channel automatic gain control based on the WDR projection can flatten the spectrum. A channel-by-channel WDR projection decreases gain for spectral peak frequencies and increases gain for other frequencies, resulting in neural responses with minimum contrast across frequency channels and diminishing the spectral information of the stimulus in the neural response similar to the challenges with multi-channel compressive amplification in hearing-aid devices (Holube et al., 2016; Bor et al., 2008; Plomp, 1988; Stone and Moore, 2008). The compression in the hearing aid is a nonlinearity that lowers the amplification for higher sound levels, and if it is implemented with a high number of channels, it can reduce the spectral information of the input sound.

Additionally, there is evidence suggesting the existence of two types of MOC projections with different bandwidth properties (Liberman & Brown,1986; Brown, 2014). Our hypothesis proposes that WDR-driven MOC projections could project to a wider band region of the cochlea compared to projections from the MOC that are driven by fluctuation (Fig. 4.2). In the proposed implementation of the MOC efferent system, WDR driven MOC

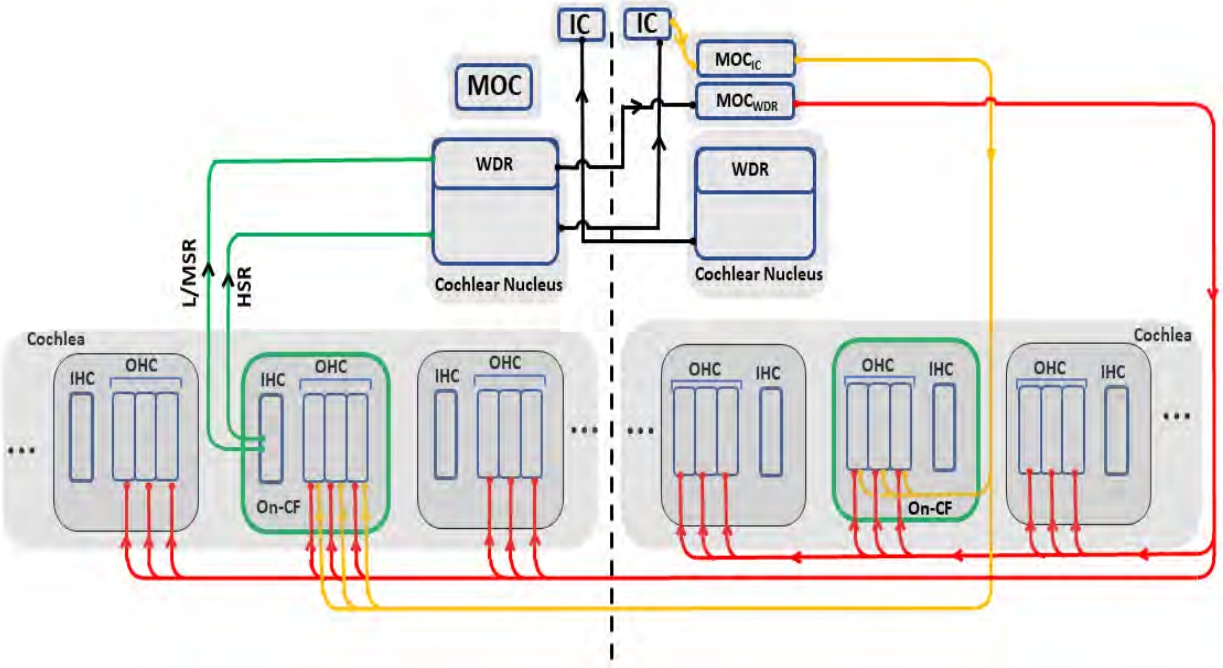

Fig 4.2. Schematic illustration of the peripheral afferent (black and green arrows) and MOC efferent (red and yellow arrows) pathways. This schematic shows the projections from MOC neurons to OHC of both on-CF (within-channel projections: hypothesized in this work to initiates from IC input to MOC neurons) and also projections to neighboring CFs (cross-frequency projections : hypothesized in this work to initiates from WDR input to MOC neurons). Here for a simple presentation, we have shown the afferent pathways on the left side and the efferent pathways on the right side, but anatomically both pathways exist on both sides of the brainstem.

signals project to a wider bandwidth of frequency channels in cochlea and fluctuation-driven MOC projection (that get their input from IC model) influence only the within channel cochlear gain. We propose that the WDR projection adjusts the cochlear gain within a range that ensures the IHC operates near its knee point. This WDR projection allows the higher stimulus amplitude near spectral peaks to saturate IHCs, while avoiding saturation for other frequencies. The IC projection, potentially with its finer-tuned characteristics, can adjust the cochlear gain more precisely.

In this study, we used the subcortical auditory model proposed by Farhadi et al. (2022), but with a broader bandwidth for the WDR projection. Our primary focus is to observe how the inclusion of the MOC efferent system enhances the coding of neural fluctuations over time. Initially, we investigate the neural response to single-formant stimuli and observe how the MOC efferent system sharpens the coding of neural fluctuations over time. Then, we transition to more complex vowel-like stimuli, two-formant stimuli or synthesized vowels. We illustrate the hypothesized effect of the MOC efferent system on sharpening neural-fluctuation coding across different sound levels and SNRs.

4.3. Methods

First, we provide an abstract explanation of existing models that do not include the efferent system and can simulate responses of AN fibers (Zilany et al., 2009, 2014) and IC cells (Nelson and Carney, 2004). Then we briefly introduce the subcortical model that includes the MOC efferents, with IC and CN projections to MOC neurons, proposed by Farhadi et al. (2022). Finally, we describe the modifications that are introduced in this study to the previous MOC efferent model to include a wider projection to the cochlea of

the WDR inputs and a fine-tuning projection to the cochlear of the IC inputs to the MOC. After introducing the model, we will explain the properties of vowel-like stimuli and the background noise that was used in this study.

4.3.1. Computational Model

In this section we will explain the peripheral model for the ascending auditory pathway and then explain the step-by-step modifications applied to this model to include more detail properties of the MOC efferent system.

Zilany AN model for the ascending periphery

The Zilany et al. (2014) model simulates the responses of auditory-nerve fibers in response to any desired input sound. In this model, the sound first goes through a middle-ear filter (Fig. 4.3). After that, there are 3 parallel filters to model cochlear tuning. These filters include a sharply tuned chirping filter at the characteristic frequency (CF), a broader bandpass filter tuned slightly higher than CF, and a control-path filter that in combination with the outer hair cell (OHC) model modifies the time constant, consequently affecting the gain and bandwidth of the chirping filter (Zhang et al., 2001; Tan & Carney, 2003; Zilany & Bruce, 2006). The output of the IHC model is the input to the synapse model, which simulates the spike rate of AN fibers for all three groups of AN fibers based on their spontaneous rates (high (HSR), medium (MSR) and low (LSR)). The output of the synapse model with HSR properties provides the input to the CN and IC models that will be described below.

Farhadi et al. (2022) model with inclusion of MOC efferent system

The subcortical model described above has only ascending projections. In Farhadi et al. (2022) the MOC efferent system was added to this model to simulate a part of the descending efferent system (Fig. 4.3). To implement the feedback loops in the MOC system, the model was reconfigured to a sample-by-sample process instead of the previous implementation where the output for the entire stimulus waveform was simulated and was the input to the IC stage. In the Farhadi et al. (2022) model, the output of the IC is one of the two inputs to the MOC block (Fig. 4.3), and the other input is from the low-spontaneous rate (LSR) fibers, which represent the WDR cell types in the CN that project to MOC. In the MOC stage of the model, these two excitatory projections combine and are low-pass filtered to simulate the dynamics of the MOC efferent system with a physiologically plausible time constant. The output of the LPF goes through an input-output function that maps the input to the MOC to a gain factor between 0 and 1. The gain factor modulates the cochlear gain. Farhadi et al. (2022) is a time-varying, within-channel, gain-control model based on the physiology and anatomical information of the MOC efferent system. The parameters of this model were adjusted using a physiological dataset (Carney et al., 2014) consisting of the neural response of IC cells in awake rabbits in response to Sinusoidal Amplitude Modulation (SAM) noise.

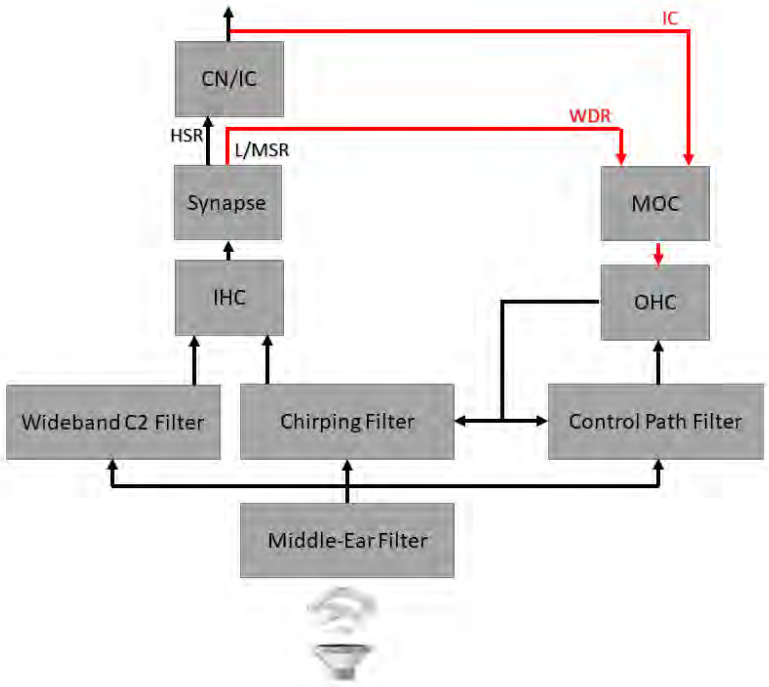

FIG 4.3. Schematic illustrating overall structure of the ascending auditory-pathway model including the Zilany et al. (2014) AN model and SFIE (Nelson and Carney 2004) IC/CN model (black). Addition of the MOC efferent system as a descending auditory pathway was implemented in Farhadi et al. (2022), including two feedback signals from IC and WDR cells to the MOC unit. (red)

Proposed MOC model with cross-frequency WDR projection

Farhadi et al. (2022) model showed a better performance compare to the model without the inclusion of the MOC efferent system in various psychoacoustic tasks in terms of predicting human behavior (Brennan et al., 2023; Farhadi et al., 2023; Farhadi and Carney, 2023; Maxwell et al., 2023). However, for more complex stimuli, such as speech signals in a background noise, a more detailed MOC model includes both within-channel and cross-channel projections. MOC projections based on what is suggested in

anatomy studies for MOC efferent system seems to be essential (Liberman & Brown,1986; Brown, 2014). The WDR projection in the MOC as a negative feedback loop in a within-channel automatic gain control system can result in a neural response with minimum contrast across frequency channels, regardless of the spectrum of the stimuli. This feedback, if implemented with a channel-by-channel projection, will lower the gain for spectral peaks and increase the gain for spectral dips, resulting in a flattened response profile across frequencies.

In this study, we change the implementation of the MOC system so that the IC projection is only affecting the cochlear gain of the CF-frequency channel, while the WDR projection affects gain over a wider frequency span. Therefore, the WDR signal associated with one specific frequency channel influences the gain factors associated with not only one CF but a range of frequencies (Fig. 4.4). To adjust the gain factor for each frequency, the IC signal associated with that specific frequency is low-pass filtered, scaled, and mapped to a value between 0 and 1 by an MOC input-output function. WDR signal associated with the specific CF associate with a neighboring of specific CF will go through the same process but with different properties for the input-output function and all outputting a value between 0 and 1. The geometric mean of the gain factor output of all WDR-driven was multiplied to IC-driven MOC output, resulting in a final gain factor between 0 and 1, which modulated the cochlear gain for the specific CF.

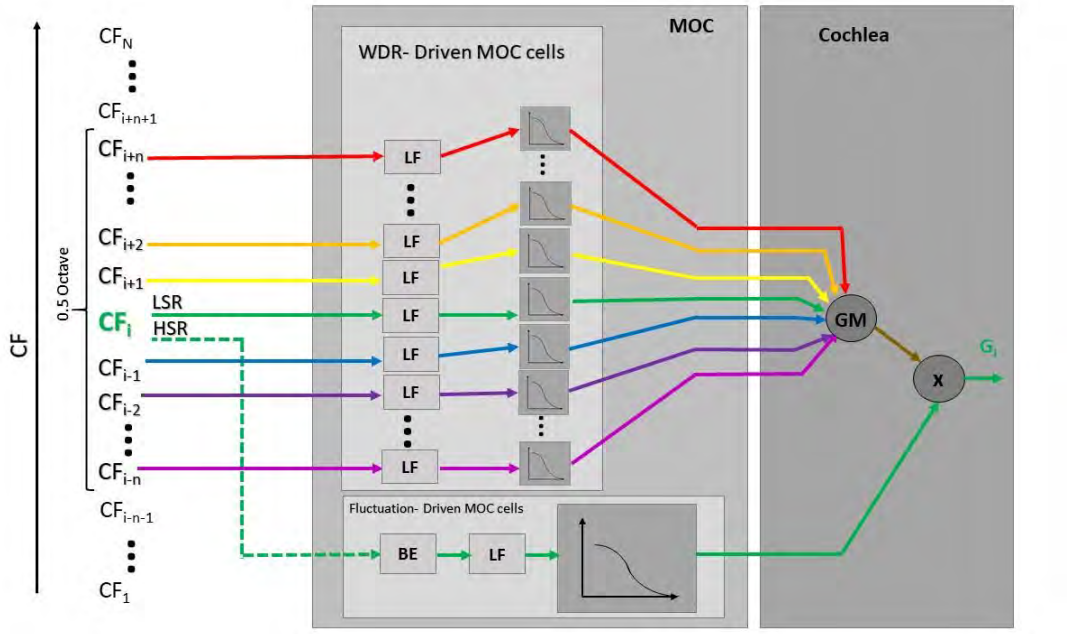

FIG. 4.4. Schematic of the proposed MOC model with fine-tuning IC projection from on-CF (green) frequency and the wide-tuned WDR projection. In this figure all the feedback signals influencing the gain for a sample frequency of "i" is showed. For WDR all the neighboring CFs within 0.5 Octave of the specific frequency are involved in gain adjustment for this frequency channel. Each WDR signal (both on-CF and off-CF) and the IC signal (on-CF) goes through a low pass filter with similar properties for all signals. WDR signal after low-pass filter will go through the WDR input-output nonlinearity and then the geometric mean (GM) of these gain factors are calculated as one value for WDR-driven MOC. The Lowpass-filtered IC signal goes through Fluctuation-driven input-output nonlinearity and the output is the IC-driven gain factor. At the end the WDR-driven gain factor and IC-driven gain factor, which have values between 0 and 1, are multiplied together, resulting in a final gain factor with a value between 0 and 1 for the specific frequency, "i". The same structure exists for other frequency channels.

The IC model used in the MOC feedback system in Farhadi et al. (2022) was a bandpass filter (Mao et al., 2013) but has been updated in this study to a more detailed IC model, SFIE model (Nelson and Carney, 2004). The SFIE model consists of two stages: CN and IC models, both functioning as coincidence detectors with inhibitory and excitatory inputs from the same frequency. The inputs to the CN model are the convolution of the AN fiber rates and appropriate alpha functions representing excitatory and inhibitory synaptic inputs, and the output of the CN model is convolved with a different set of alpha functions to create inputs to the IC model. The schematic diagram (Fig. 4.5) represents the SFIE model, with line thickness showing the relative strength of inhibitory and excitatory inputs. The alpha functions displayed in the figure have different time constants for excitatory and inhibitory inputs at each stage. Equation 1 describes the response of the IC model, including a delay (D) to account for the synaptic delay of the inhibitory response. The parameter A scales the output rate to a reasonable average rate, and S indicates the relative strength of inhibitory to excitatory input. The CN model's parameters are specified in Nelson and Carney (2004), and the IC model's parameters are listed in Table 4.1. Figure 4.5 illustrates the modulation transfer function (MTF) of the IC model.

$$r_{IC}(\mathrm{t}) = \mathrm{A}\left[\mathrm{t}e^{\frac{-t}{\tau_{exc}}} * \mathrm{r}_{cn}(\mathrm{t}) - S_{IC,inh}(\mathrm{t} - D_{IC}) \times e^{\frac{-(t-D_{IC})}{\tau_{inh}}} * \mathrm{r}_{cn}(\mathrm{t} - D_{IC})\right]$$

$(Eqn.\ 1)$

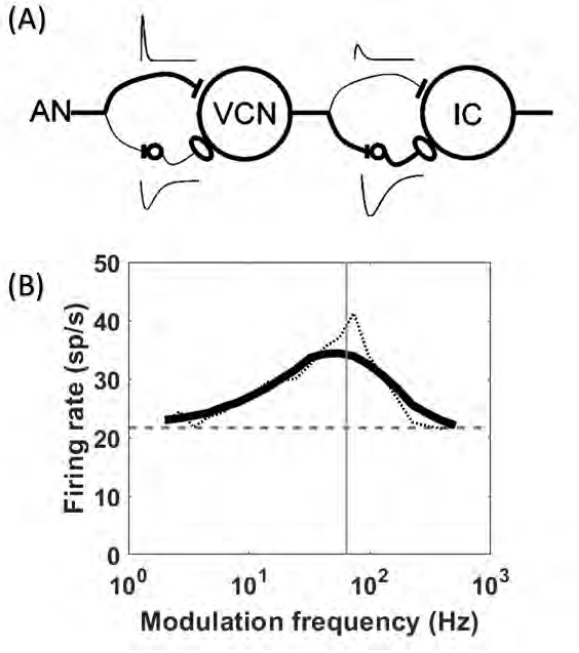

FIG.4.5. The IC model used in the MOC efferent feedback is the SFIE model (Nelson and Carney, 2004) A) Schematic of the CN/IC model (from Nelson and Carney, 2004) and alpha functions at each stage of the model. The thickness of the projection lines shows the relative strength of the projections B) The noise MTF of the SFIE model for a BE IC cell using the parameters in table 4.1 and with inclusion of MOC efferent activity at 65 dB SPL. The horizontal dotted line shows the unmodulated rate, the vertical line shows the BMF (64 Hz).

Parameter	Description	Value
τ_{ex}	Excitation time constant (S)	0.0016
τ_{inh}	Inhabitation time constant (S)	0.0023
D	Inhibitory delay (S)	0.0032
A	alpha function area	1
S_{inh}	Inhibitory strength	0.9

Table 4.1 Parameter's value and description for SFIE IC

The new parameter introduced to this model is the bandwidth of the WDR projections, while also other parameters such as scalers for WDR and IC and parameters associated with the input-output function of the WDR pathway and IC pathway were re-adjusted, as the implementation differed and the IC rates differed because of using SFIE IC model. We kept the low-pass filter property intact as it was physiologically reasonable and had been adjusted in the Farhadi et al. (2022) model. The other parameters were adjusted using speech-like stimuli to improve the sharpness of the response-rate profile for all different stimuli with different properties. However, we cannot be certain that these parameters accurately replicate a healthy human auditory system. This study investigates the feasibility of sharpening of the neural fluctuation profile using a model with MOC efferent feedback. While a more accurate adjustment of these parameters based on physiological datasets can be an interesting future work of this study, it is out of the scope of the current computational modeling study, as there is not enough information about the MOC efferent projection properties due to the various challenges that such physiological and anatomical studies face. The parameters that were used in the MOC model are described in Table 4.2. Equation 2 also shows the equation for calculating the gain factor for an example frequency (i). In this equation n indicates the number of CFs that falls within the band of BW_{WDR}.

The IC model used in this study for both models with and without efferents is the SFIE model with a best modulation frequency (BMF) of 64 Hz, which is the median value of population of rabbit IC neurons (Kim et al., 2020), and a BE MTF shape. A total of 100 CFs with a linear space in the range of 150 Hz and 3 KHz were simulated.

Analysis windows of 100-ms duration were used to find the average IC rate and gain factor over those windows to evaluate the effect of MOC efferent activity over the stimulus time. All the simulated responses were averaged over 10 repetitions.

Parameter	Description	Value
B_{WDR}	"beta" parameter in the MOC input-output nonlinearity for the WDR MOC pathway.	0.01
B_{IC}	"beta" parameter in the input-output function of the IC projection.	0.01
O_{WDR}	"offset" parameter in the MOC input-output nonlinearity for the WDR MOC pathway. (Spikes/Sec)	0
O_{IC}	"offset" parameter in the MOC input-output nonlinearity for the IC MOC pathway. (Spikes/Sec)	50
K_{WDR}	Scalar value multiplied with lowpass-filtered WDR signal before signal is passed through MOC input-output nonlinearity.	6
K_{IC}	Scalar value multiplied with lowpass-filtered IC signal before signal is passed through MOC input-output nonlinearity.	3
F_{cutoff}	Cutoff of the lowpass filter used in the MOC stage (Hz).	0.64
BW_{WDR}	Width of the WDR cross-channel spread (octaves).	0.5
BW_{IC}	Width of the IC cross-channel spread (octaves).	0

Table 4.2. Parameter's value and description for the proposed MOC efferent model.

$$G_i(t) = \left\{ \sqrt[2n]{\prod_{x=i-\frac{n}{2}}^{i+\frac{n}{2}} \frac{1}{1 + \left(B \times \left(LPF\left(WDR_x(t)\right) - O_{WDR}\right)^2\right)}} \right\} \times \frac{1}{1 + (B \times LPF(IC_i(t) - O_{ic})^2)} \qquad (Eqn.\ 2)$$

4.3.2. Stimuli

To maintain the proper function of the positive feedback loop that incorporates the IC fluctuation-driven projection, it was essential to ensure that the bandpass-shaped property of the IC model is maintained at the sound levels we are simulating the neural response. In this study, first we examined the changes in the modulation transfer function (MTF) shape of the IC model under different sound levels ranging from 10 dB to 90 dB, comparing the MTFs of the model with efferent activity and the model without efferent activity. Then we used the model to explore the neural coding to vowel-liked stimuli in silence and also in background noise.

SAM noise**:** For the sinusoidally amplitude-modulated (SAM) noise MTF analysis, the stimulus duration was set to 500 ms, and the CF was 1 kHz. The modulation frequency varied from 2 Hz to 512 Hz, with 21 linearly spaced modulation frequencies. The wideband noise had a bandwidth from 100 Hz to 8,000 Hz. The modulation depth was set to 0 dB (fully modulated noise). Each condition was repeated 50 times and the average smooth response was used for analysis.

Single-Formant Stimulus: The first speech-like stimuli that were used in this study were single-formant stimuli, a harmonic complex with a triangular spectral envelope over a log-frequency axis with a center spectral peak, similar to vowel spectrum. The peak frequency was 2 kHz and the fundamental frequency (Fo) was 100 Hz. The slope of the spectral envelope (G) was set to 50 dB per octave. The duration of the stimulus was 300 ms, with 25-ms $\cos^2$ on/off ramps.

Double-formant stimulus (synthesized-vowel stimuli): The double-formant stimulus was generated using the Klatt (1980) vowel-synthesizer software with resonators and impulse-train inputs. The formant frequencies were 500 Hz and 2100 Hz and BW for each formant is 70 and 90 Hz respectively. The duration of the stimulus was 300 ms, with 5 ms $\cos^2$ on/off ramps durations. The F_0 was 100 Hz for this stimulus.

Background Noise: In order to evaluate the robustness of the neural coding of vowel-like stimuli in background noise, we used long-term-average-speech-spectrum (LTASS) noise based on the spectrum described by Byrne et al. (1994). LTASS is a wideband noise with a spectrum similar to the long-term average spectrum of real speech. This noise has been used in several studies for speech-intelligibility experiments.

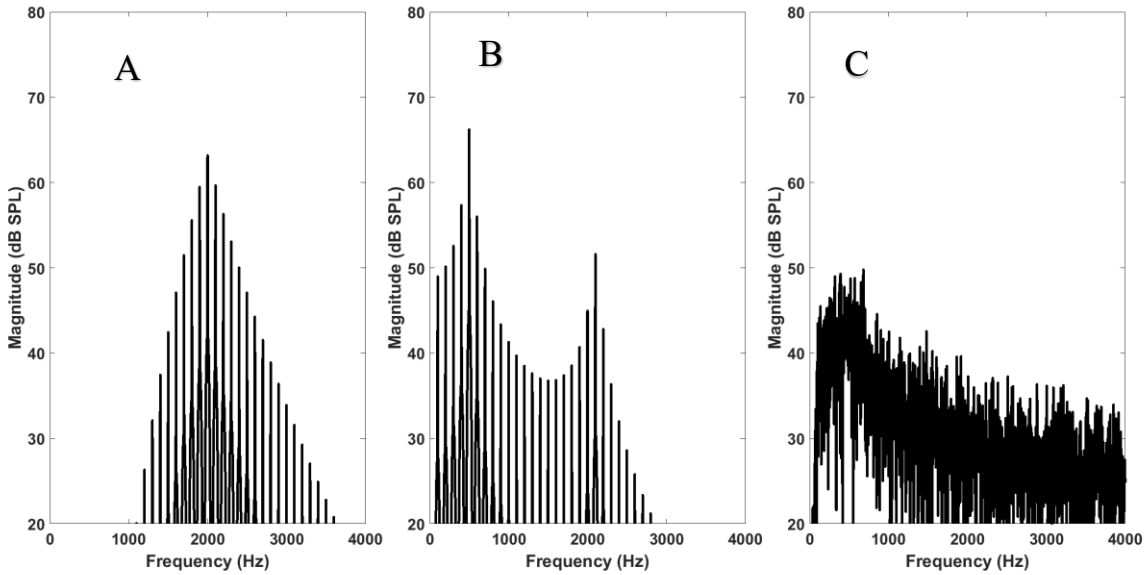

FIG.4.6. Spectrum of the stimuli used in this study. A) Single-formant stimulus with peak frequency at 2 kHz and F_0 of 100 Hz spectrum slope of 50 dB per octave. B) Double-formant (Formant Frequencies 500, 2100 Hz and F_0 of 100 Hz. C) LTASS speech-shaped background noise (Byrne et al., 1994). All three stimuli had overall level of 65 dB SPL.

4.4. Results

First, we will examine the impact of MOC efferent activity on the robustness of the MTF shape of the IC model at various sound levels. Then, we will investigate the neural response to a single-formant stimulus at different sound and noise levels, both with and without MOC efferent activation. The next section follows the same procedure but with a more vowel-like stimulus, a double-formant stimulus. We will analyze the dynamic effect of MOC activity using analysis time windows throughout the duration of the stimuli.

4.4.1. Effect of the MOC efferent activity on MTF shape

The IC model response is influenced by the MOC efferent system, which in turn affects the cochlear gain. However, the cochlear gain adjustments due to the MOC efferent activity also impact the response of the IC because adjustments in the cochlear gain change the modulation depth of the neural fluctuations in AN HSR response, which is the input to the CN and then IC model. Decreasing the gain enhances the effective modulation depth in neural fluctuation of the AN response. As a result, we anticipate a higher IC rate in response to AM noise with MOC activation, resulting in a stronger and more robust MTF shape. Preserving the MTF shape in the IC model is crucial for the effective operation of the IC positive feedback. The IC feedback's role is to enhance contrast in the neural fluctuation profile (see introduction of this chapter) and this process relies on the sensitivity of IC cells to low-frequency modulation in AN response.

In Fig 4.7, the MTFs of the IC model are presented for 5 different noise levels. At very low sound pressure levels (SPLs), the MTF shape remains relatively unchanged with

MOC activation. This is because the WDR input to MOC is minimal, resulting in low MOC activity and minimal gain reduction. Therefore, the model responses with and without MOC activation are expected to be similar. As the sound level increases, the MTF shape for the model without MOC activation quickly becomes flatter. This means that, for an average sound level of 50-70 dB SPL, the IC model response without MOC activation becomes less sensitive to modulation at the BMF frequency. In contrast, the model with MOC activity maintains the MTF shape over a wider range of sound levels. Considering the ability of human listeners with healthy hearing to discriminate vowels at 50-70 dB SPL, as well as the physiological response to AM noise in awake rabbits, the model with the

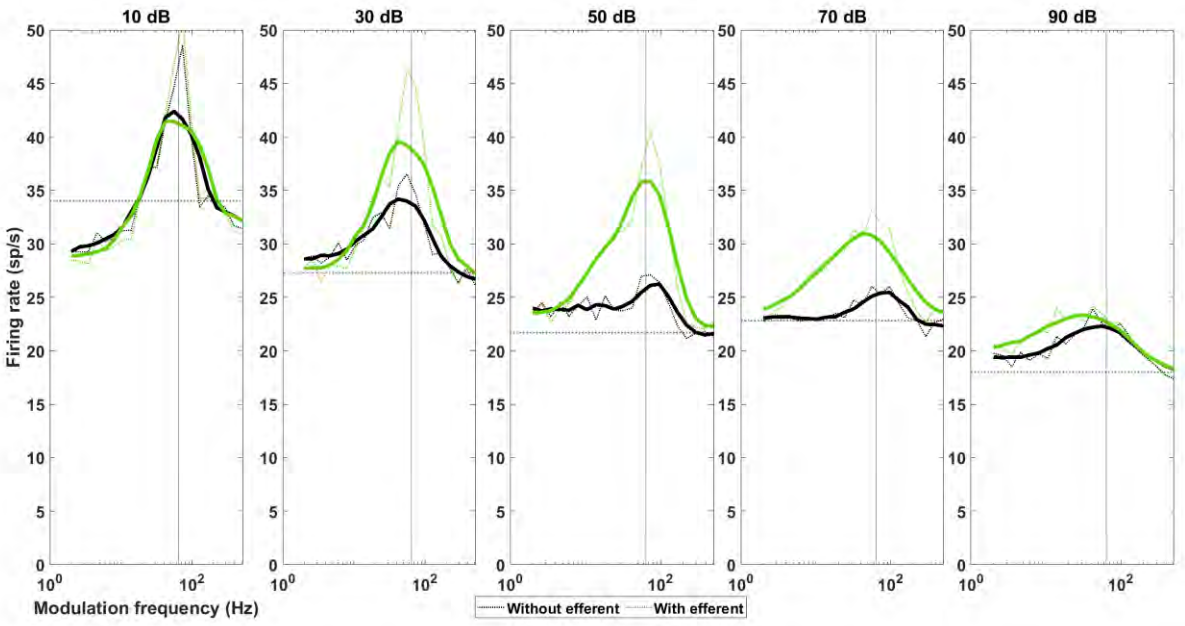

Fig. 4.7. MTF of the IC BE Model with (green) and without (black) a cross-channel MOC efferent activity. Noise was varied between 10 to 90 dB SPL. The horizontal dotted line shows the IC rate in response to an unmodulated noise, the vertical line shows the BMF (64 Hz).

cross-frequency MOC efferent system more accurately simulates the neural response to AM noise.

4.4.2. Effect of the across-frequency MOC efferent model on neural coding of single-formant stimuli

In this section, we investigated the impact of the cross-frequency MOC efferent system on neural coding of single formant stimuli, as a simple representation of a vowel-like stimulus. Our hypothesis suggests that MOC efferent activity enhances the neural-fluctuation profile throughout the duration of the stimulus. To facilitate better comparisons between the model's responses at different time windows and to compare them with the responses of the model without MOC efferent activation, we introduce two criteria that indicate an enhancement in neural fluctuation coding.

Figure 4.8 presents depth and sharpness as two measures for assessing the enhancement in neural fluctuation coding. Depth refers to the maximum difference between the rate at the formant frequency and the edge frequencies, while sharpness

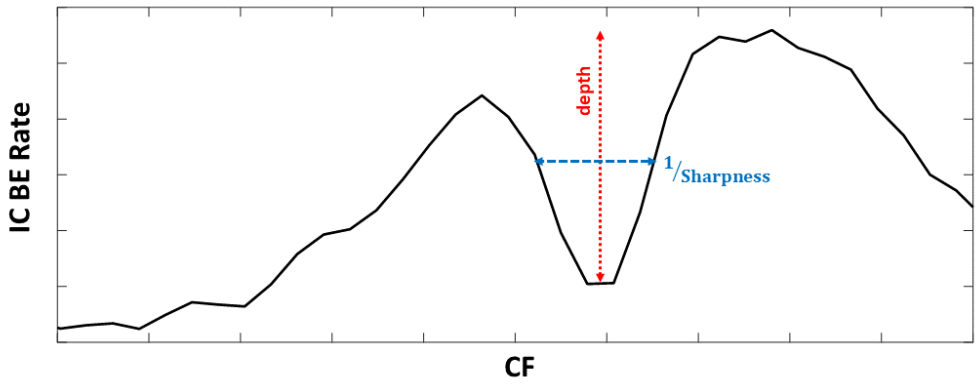

Fig.4.8. Enhancement of the neural fluctuating coding in IC BE rate can be measured by sharpness and depth of the response at the local maxima frequency the stimulus spectrum

represents the bandwidth of the response at half of the depth. Increased values in both depth and sharpness indicate an enhancement in the neural fluctuation profile.

Figure 4.9 displays the IC BE response of the proposed model with cross-channel MOC efferent activity to a single-formant stimulus with peak frequency of 2 KHz and spectral slope of 50 dB/Octave (see Figure 4.6.A). The top row illustrates the cochlear gain factor that multiplies with the cochlear gain. Each color in the plot shows the average gain factor over a different non-overlapping time analysis window to show the time-varying

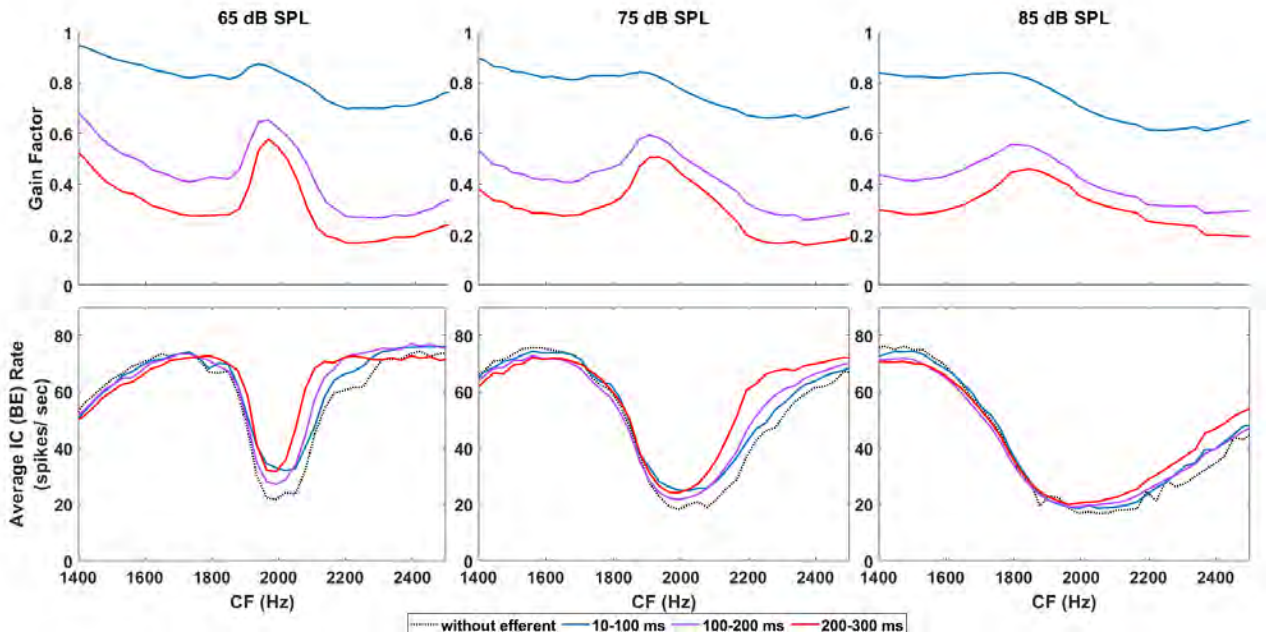

Fig.4.9. Top row: the averaged gain factor adjusting the cochlear gain for each CF frequency over time analysis windows. Bottom row: the average IC BE model response to a single formant centered at 2 kHz using three different time analysis windows, along with the average response of the model without MOC activation (dotted black line). Each column corresponds to one sound level (65 dB SPL, 75 dB SPL, and 85 dB SPL), showing the response and gain factor.

gain control of the MOC efferent system. Noticeably, in Fig. 4.9, the gain has sharper local maxima around the peak frequency of the stimulus, which is 2 kHz and expicifically for the time analysis window that are are later in time.The bottom row shows the average IC BE rate over the three different time windows, and the average response of the model without efferent activation. We observe that, across all sound levels, the response of the model with cross-frequency MOC activation has a sharper response at target frequency in comparison with the model without dynamic gain control and it gets sharper over time. The sharpening effect appears to be more distinct at the moderate sound level of 75 dB SPL. The depth of the response, which serves as another measure for evaluating the enhancement of the coding of the neural fluctuation profile, either increased in the model with MOC efferent activation compared to the model without efferent responses, or did not show significant changes throughout this experiment at different sound pressure levels (SPLs).

Figure 4.10 illustrates the response of proposed model with cross-frequency MOC system to a single-formant in background noise. Each column represents a different signal-to-noise ratio (SNR), ranging from 15 dB SNR and gradually increasing the noise to achieve 0 dB SNR. The stimulus level remained constant at 65 dB SPL, while the speech-shaped noise level varies.

The first row of Fig. 4.10 displays the spectrum of the stimulus, consisting of the single formant and the speech-shaped noise. The second column shows the averaged gain factor over the three different time analysis windows. Due to the presence of background noise, the peak at the target frequency (2 kHz) is not as pronounced as it was in Fig. 4.9 with the single formant in silence. The shape of the peak on the lower side is also less

distinct, mainly influenced by the spectral properties of the speech-shaped noise, as the noise level was higher at lower frequencies and rolled off at higher frequencies. The third row in Fig. 4.10 presents the average IC BE responses over the time analysis windows. Over time, not only does the response become sharper, but the depth of the response also increases. These observations indicate an enhancement in neural coding at the IC BE in response to the single-formant stimulus in the presence of background noise. However, this enhancement becomes weaker as the noise level increases.

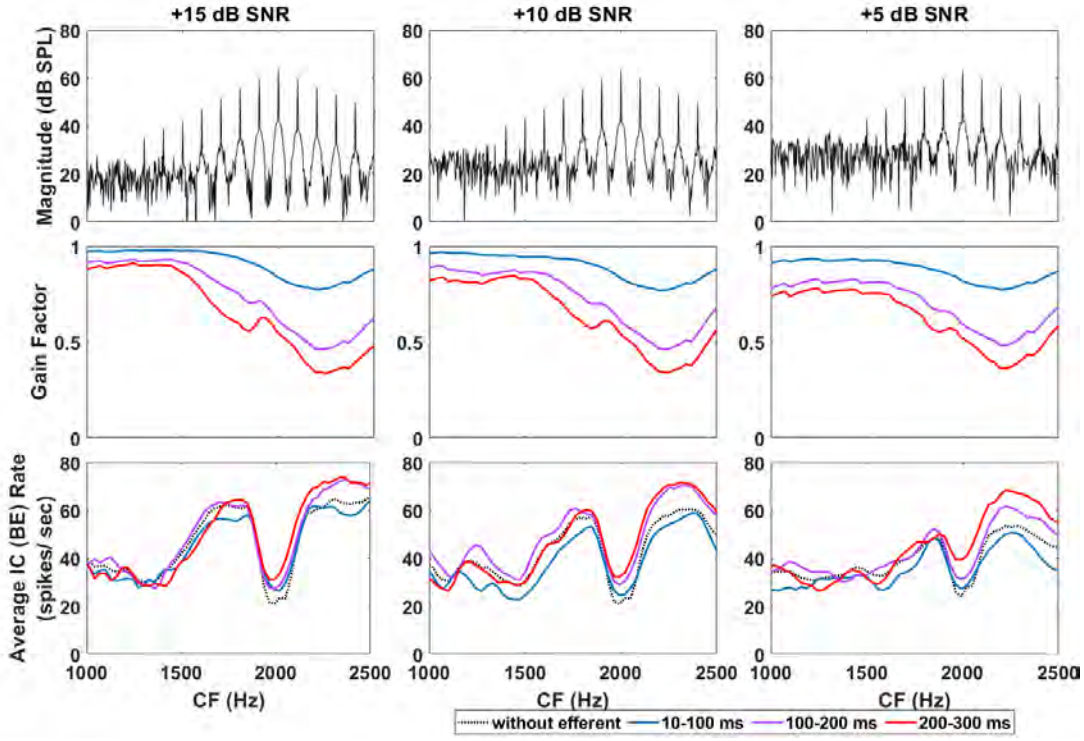

Fig.4.10: Top row: the stimulus spectrum comprising of a single formant stimulus and speech-shaped noise. Middle row: the averaged gain factor adjusting the cochlear gain for each CF frequency over the time analysis windows. Bottom row: the IC BE model response to the single formant stimulus centered at 2 kHz with a speech-shaped background noise, averaged over three different time analysis windows. The average response of the model without MOC activation is represented as a dotted black line. Each column corresponds to a different signal-to-noise ratio (SNR): 15 dB, 10 dB, and 5 dB.

4.4.3. Effect of the across-frequency MOC efferent system on neural coding of double-formant stimuli

In this section, we will examine the influence of the cross-frequency MOC efferent system on the neural coding of double formant stimuli, which more closely resembles a vowel stimulus. Using the two-formants stimuli in this study guide us to better adjust the constrains on the bandwidth of WDR projection so that considering the bandwidth of the typical formants in a vowel the WDR projection could still discriminate between the two peaks. The two formants in this stimulus are located at frequencies of 500 Hz and 2000 Hz. Figure 4.11 illustrates the gain factor (top row) and the average IC BE response (bottom row) over time windows for sound levels of 65 dB, 75 dB, and 85 dB SPL. In response to both formants, there is a peak in the gain factor plots at all sound levels. However, the peak becomes shallower as the sound level increases for both formants. The top row of Fig. 4.11 also shows that the cross-frequency MOC efferent activity sharpens the peak at the formant frequencies over time.

In the bottom row of Fig. 4.11, the average IC rate of the model with cross-frequency MOC implementation exhibits a sharper response at the high-frequency formant compared to the model response without efferent activation. This sharpness further increases over time, as observed by comparing the different colored plots in each figure of the bottom row. The depth of the response either improves or remains unchanged when compared to the model without efferent activation, across all sound levels. Similar to the single formant response (Fig. 4.9), the enhancement in neural fluctuation coding is most pronounced in the middle column, where the sound level is in the middle range.

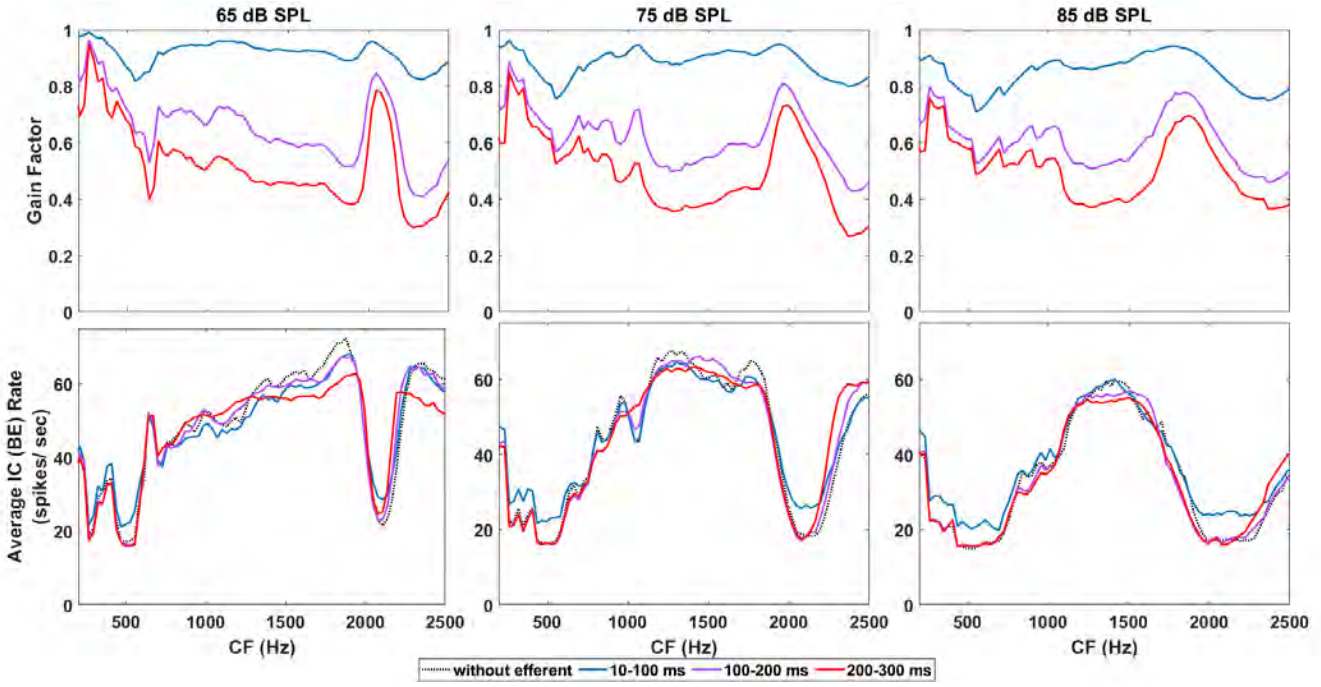

Fig.4.11: Top row: the averaged gain factor adjusting the cochlear gain for each CF frequency over time analysis windows. Bottom row: the average IC BE model response to a double formant with formant frequencies at 500 and 2100 Hz over three different time analysis windows, along with the average response of the model without MOC activation (dotted black line). Each column corresponds to one sound level (65 dB SPL, 75 dB SPL, and 85 dB SPL), showing the response and gain factor.

Responses to double-formant stimuli in the presence of background speech-shaped noise are depicted in Fig. 4.12. Similar to Fig. 4.10, the top row displays the stimulus spectrum, which in this case represents the double-formant in noise. The stimulus spectrum level remains constant at 65 dB SPL. In Fig. 4.12, each column represents a different spectrum level of the noise, resulting in various signal-to-noise ratios (SNRs) ranging from

15 dB to 0 dB. Compared to the single-formant stimulus, the effect of noise was more pronounced in Fig. 4.12, particularly for the higher-frequency formants, due to the lower energy of the higher-frequency formants when compared to the lower-frequency formants. In Fig. 4.12, the second row displays the averaged gain factor. It is evident that at lower sound levels, there is a local peak in the gain factor at both formant frequencies. However, as the noise level increases, the local maximum at the second formant becomes weaker, while remaining robust for the first formant. The effect of these gain adjustments can be observed in the third row of Fig. 4.12, which shows the average IC BE rates for the three analysis windows. The dip in response to the first formant remains strong and consistent across different noise levels. However, for the second formant with the higher frequency, the depth of the response is initially strong at lower noise levels but gradually decreases with increasing noise level. The enhancement of neural coding in the IC BE response, considering both sharpness and depth criteria, increases over time and is more pronounced compared to the model without efferents. However, this enhancement effect diminishes as the noise level increases, eventually becoming indistinguishable in the response to the 0 dB SNR stimulus.

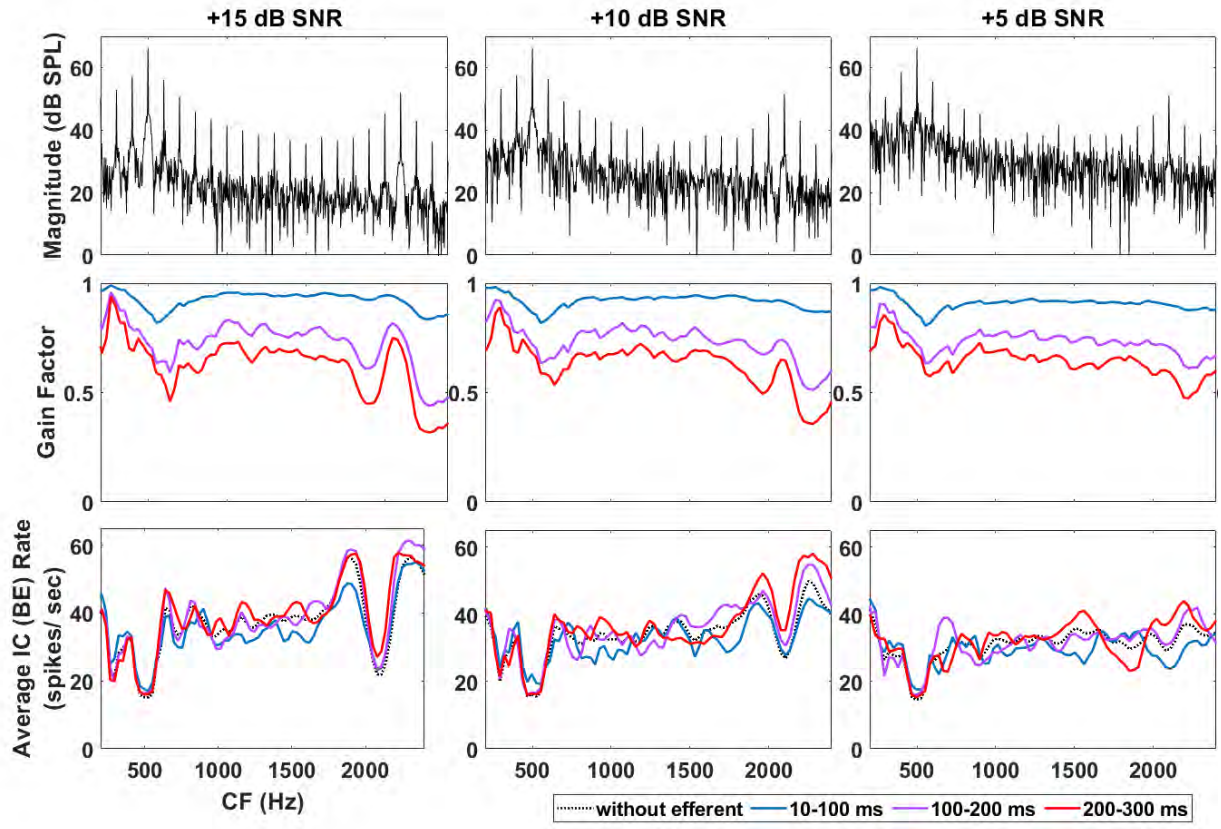

Fig.4.12. Top row: The stimulus spectrum consists of double-formant stimulus and speech-shaped noise. Middle row: the gain factor adjusting the cochlear gain for each CF frequency averaged over time analysis windows. Bottom row: the IC BE model response to double formant with formant frequencies at 500 and 2100 Hz in a speech-shaped background noise. These responses averaged over three different time analysis window and shown with different shade of blue colors. Also the average response of the model without MOC activation is plotted (dotted black). Each column corresponds to a different SNR (15, 10, and 5 dB).

Figure 4.13 provides a quantitative overview of the results related to neural fluctuation (NF) enhancement, considering the depth and sharpness of the representation of F2 in IC BE response profile. In the absence of background noise (silent condition, Fig. 4.13, top), the sharpness of the dip in the IC BE response profile increased from the model without efferents to the model with efferents. In addition, for the model with efferents the sharpness increased over time. However, the depth either decreased or remained similar in both models and over time. Consequently, the combined effect, which was the product of depth and sharpness, improved for both single-formant and double-formant stimuli in the silent condition when compared to the model without efferents, particularly during the later time-window analysis. This improvement was not as prominent at the 85-dB SPL sound level and became more noticeable at medium sound levels (ranging from 65- to 75-dB SPL), which align with conversational levels (Olsen, 1998).

In the results with background noise (Fig. 4.13, bottom), an increase in the depth of the dip of the IC BE response profile was observed when comparing the model with efferents to the model without efferents. However, the improvement in depth was relatively smaller compared to the enhancement in sharpness. When considering the combined effect, for both single-formant and double-formant stimuli in the presence of noise, the enhancement effect was stronger in the model with efferents compared to the model without efferents, particularly during later time windows. While there were examples where the enhancement effect was not consistently present across all time windows, at least one time window had a higher enhancement compared to the model without efferents. For instance, in the case of a 5-dB SNR, the last window had a lower enhancement compared

to the model without efferents, but the other two windows had higher enhancement effects. For this result, the most noticeable difference between the models with and without efferents was observed at the medium noise level (10-dB SNR).

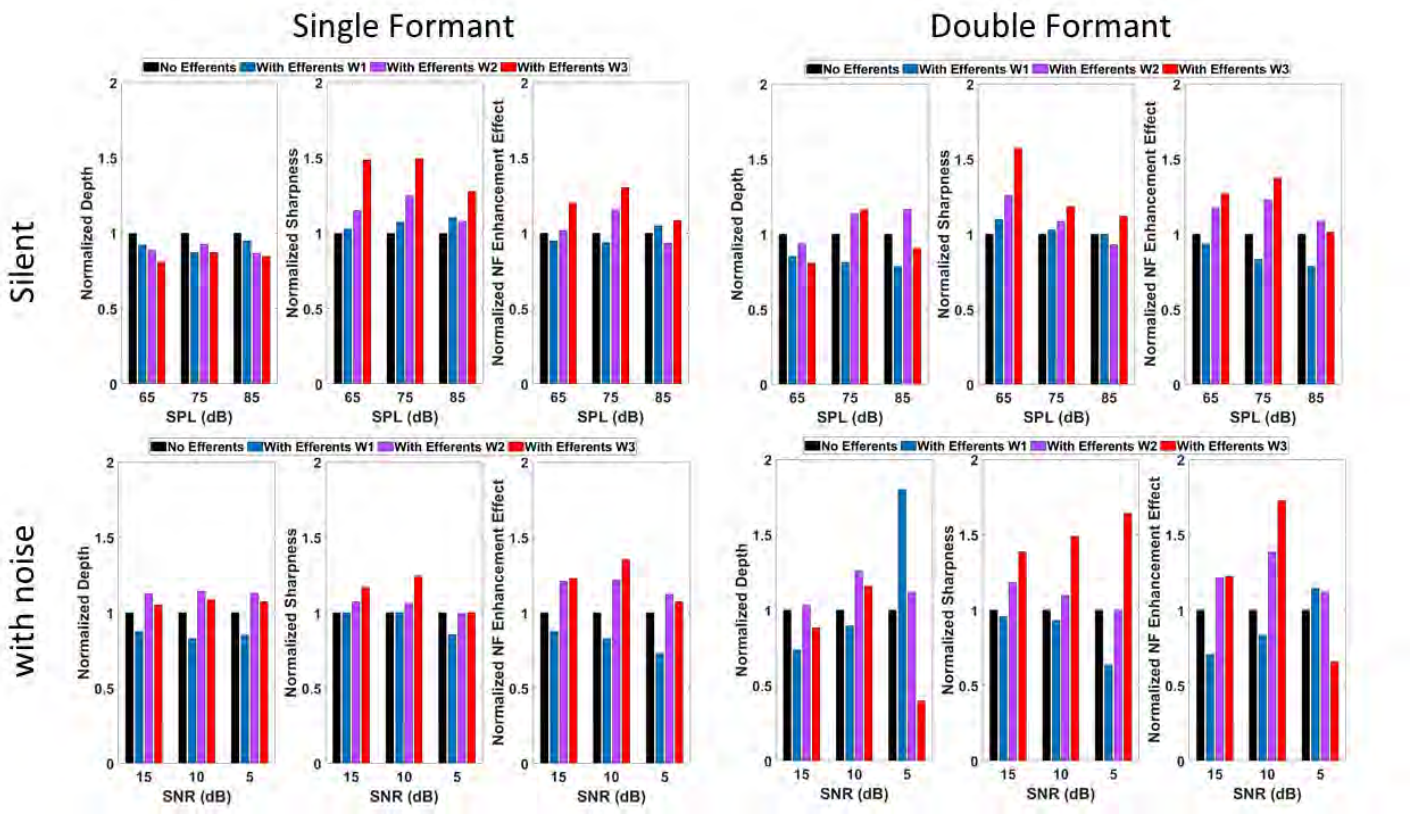

Fig.4.13. Summary of the Neural Fluctuation (NF) enhancement criteria for both stimuli (left: single formant, right: double formant) in silence (top) and in noise (bottom). All values are normalized to the criteria value for the model without efferents. Black bar is associated with the model without efferents and the colored bars are associated with the model with efferent for different time analysis windows. (Blue: W1 (10-100ms), Purple: W2 (100-200 ms), Red: W3 (200-300 ms). For each combination of the stimuli and noise/silence there are three subplots. The first subplot shows the depth of the response, the second subplot shows the sharpness of the response and the last subplot shows the combination of the two by calculating the depth multiplied by sharpness.

4.5. Discussion

In this study, we used a subcortical auditory model that incorporated the MOC efferent system to investigate the role of the MOC efferent system in neural coding of speech sound. We made modifications to the MOC model structure, so that it includes both within-channel and cross-channel MOC projections to the cochlea. The WDR-driven MOC projection served as a dynamic automatic gain control system, maintaining neural fluctuation at different sound levels by keeping the cochlear gain within a range by maintaining the operating point of inner hair cells (IHCs) near the knee point of their input-output functions. The cross-frequency implementation of WDR-driven projection ensured that the spectral information of the stimulus was preserved and not removed by flattening of the response spectrum by narrowband automatic gain control of this feedback. On the other hand, the fluctuation-driven MOC projection was implemented using within-channel projections, allowing the fluctuation amplitude of each frequency channel to influence the cochlear gain at that specific frequency. This finer adjustment of gain, based on the peaks and valleys in the stimulus spectrum, enhanced the contrast in the neural-fluctuation profile over time.

This new implementation of the subcortical auditory model was then used to explore the enhancement in neural-fluctuation coding using simulated neural responses at the level of IC BE neurons to SAM, single-formant, and double-formant stimuli. Our study explored the effects of sound level and speech-shaped noise on neural coding, both with and without efferent activation, analyzing various time windows to investigate the dynamic effects of the MOC efferent system.

The simulation results presented in this chapter showed an enhancement in neural fluctuation coding for different test stimuli. This enhancement was evaluated based on two criteria in the IC BE rate profile: response sharpness and response depth. When considering various sound levels, the simulation results revealed that the model with efferent activation exhibited sharper responses to stimuli compared to the model without efferent activation. Furthermore, the sharpness of the response increased over time in the efferent model. In the case of stimuli presented in noise, the model with efferent activation consistently simulated a sharper response over time, outperforming the model without efferent activation for both single-formant and double-formant stimuli in terms of enhancement of neural-fluctuation coding. Notably, these results not only demonstrated sharper responses but also had improved response depth, as the model with efferent activation had higher simulated rates at the edges of spectral peaks. This improvement was observed because of the increase in neural-fluctuation amplitude due to the addition of noise. The increased neural fluctuations contributed to a positive feedback loop that enhanced the contrast between the responses at the peak and the edges of spectral peaks, as discussed in the introduction.

It is noteworthy that the enhancement in neural-fluctuation coding for vowels in noise was mainly observed for average sound levels and lower SNR values, rather than more challenging conditions such as very low or very high sound levels or negative SNRs. This observation aligns with the findings of Jennings, 2021, that suggested that the MOC efferent activity is more beneficial in suprathreshold conditions. This observation suggests

that MOC efferent activity may be more important in everyday life conditions, rather than near-threshold situations.

Considering the feedback control system associated with the MOC efferents, there is a requirement for some level of initial contrast in neural fluctuation for the fluctuation-driven feedback to further improve the contrast through a positive feedback loop. For very high levels of noise, the contrast may not be sufficient for this positive feedback to enter the enhancement loop effectively. This intriguing finding provides valuable insight into understanding individual differences with hearing loss. While the audiometric thresholds of human listeners may be similar across individuals, variations in the performance of the MOC efferent system, whether due to misfunctioning of the system itself or dysfunction of peripheral signals at the input to the system, can contribute to differences in speech comprehension in noisy environments, particularly in more complex but supra-threshold conditions.

It is important to note that the parameters and structure of the proposed MOC model may not definitively represent an accurate model of this complex system. Many aspects of the anatomy and physiology of the MOC efferent system remain unknown (Guinan et al, 2018). However, in this study, our focus was primarily on evaluating the potential of such a model to improve neural contrast through simulation and modeling. This work can serve as inspiration and guidance for future experiments and studies aimed at better understanding the mechanisms involved. Additionally, it can inspire new ideas for enhancing the signal-processing capabilities of hearing-aid devices and other automatic

hearing devices by considering the improvement of neural-fluctuation contrast as a coding mechanism for speech comprehension and learning from the processes of normal hearing.

Lastly, in this study, our focus was on how the MOC system may enhance neural-fluctuations as a robust coding mechanism, particularly at conversational sound levels. This coding benefits from the inherent nonlinearities in the normal auditory system for coding information conveyed to central and, eventually, cortical levels for the final goal of sound perception. However, the extent to which the enhancements observed in the simulation results of this study can contribute to improved speech perception remains an intriguing question that falls outside the scope of this study. The results presented here encourage further investigation as an important future direction. A potential initial step towards achieving this goal could involve predicting formant-discrimination accuracy for suprathreshold noise levels, which would require additional data collection with physiological and psychoacoustical measurements. In summary, our work supports the hypothesis regarding the potential of the MOC efferent system to enhance neural coding of speech sounds. Our work also suggests the need for further exploration and consideration of the role of the MOC efferent system in speech perception in noise, as well as the improvement of computational models relating to MOC efferent system.

Acknowledgements

This work was supported by NIH-R01-DC010813.

Bibliography

Spille, C., Kollmeier, B., & Meyer, B. T. (2018). Comparing human and automatic speech recognition in simple and complex acoustic scenes. Computer Speech & Language, 52, 123-140.

Humes, L. E., Kidd, G. R., & Lentz, J. J. (2013). Auditory and cognitive factors underlying individual differences in aided speech-understanding among older adults. Frontiers in systems neuroscience, 7, 55.

Chou, K. F., Dong, J., Colburn, H. S., & Sen, K. (2019). A physiologically inspired model for solving the cocktail party problem. Journal of the Association for Research in Otolaryngology, 20, 579-593.

Fant, G. (1960). Acoustic theory of speech production (mouton, the hague, the netherlands). Google Scholar, 169-185.

Carney, L. H. (2018). Supra-threshold hearing and fluctuation profiles: implications for sensorineural and hidden hearing loss. Journal of the Association for Research in Otolaryngology, 19(4), 331-352.

Sachs, M. B., & Young, E. D. (1979). Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. The Journal of the Acoustical Society of America, 66(2), 470-479.

Young, E. D., & Sachs, M. B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. The Journal of the Acoustical Society of America, 66(5), 1381-1403.

Delgutte B, Kiang NY (1984A) Speech coding in the auditory nerve: I. Vowel-like sounds. J Acoust Soc Am 75:866–878

Liberman MC (1978) Auditory-nerve response from cats raised in a low-noise chamber. J Acoust Soc Am 63:442–455

Costalupes, J. A., Young, E. D., & Gibson, D. J. (1984). Effects of continuous noise backgrounds on rate response of auditory nerve fibers in cat. Journal of neurophysiology, 51(6), 1326-1344.

Carney, L. H., Li, T., & McDonough, J. M. (2015). Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations. Eneuro, 2(4).

Deng, L., Geisler, C. D., & Greenberg, S. (1987). Responses of auditory-nerve fibers to multiple-tone complexes. The Journal of the Acoustical Society of America, 82(6), 1989-2000.

Langner, G., & Schreiner, C. E. (1988). Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. Journal of neurophysiology, 60(6), 1799-1822.

Krishna, B. S., & Semple, M. N. (2000). Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. Journal of neurophysiology, 84(1), 255-273.

Nelson, P. C., & Carney, L. H. (2004). A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. The Journal of the Acoustical Society of America, 116(4), 2173-2186.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. Physiological reviews, 84(2), 541-577.

Guinan Jr, J. J. (2018). Olivocochlear efferents: Their action, effects, measurement and uses, and the impact of the new conception of cochlear mechanical responses. Hearing research, 362, 38-47.

Brown, G. J., Ferry, R. T., & Meddis, R. (2010). A computer model of auditory efferent suppression: implications for the recognition of speech in noise. The Journal of the Acoustical Society of America, 127(2), 943-954.

Yasin, I., Drga, V., Liu, F., Demosthenous, A., & Meddis, R. (2020). Optimizing speech recognition using a computational model of human hearing: effect of noise type and efferent time constants. IEEE Access, 8, 56711-56719.

Mertes, I. B., Wilbanks, E. C., & Leek, M. R. (2018). Olivocochlear efferent activity is associated with the slope of the psychometric function of speech recognition in noise. Ear and hearing, 39(3), 583.

Clark, N. R., Brown, G. J., Jürgens, T., & Meddis, R. (2012). A frequency-selective feedback model of auditory efferent suppression and its implications for the recognition of speech in noise. The Journal of the Acoustical Society of America, 132(3), 1535–1541. https://doi.org/10.1121/1.4742745

Giguere, C., & Woodland, P. C. (1994). A computational model of the auditory periphery for speech and hearing research. II. Descending paths. Journal of the Acoustical Society of America, 95(1), 343–349. https://doi.org/10.1121/1.408367

Kwan, T. J. M., Zilany, M. S. A., Davies-Venn, E., & Abdul Wahab, A. K. (2019). Modeling the effects of medial olivocochlear efferent stimulation at the level of the inferior colliculus. Experimental Brain Research, 237(6), 1479–1491. https://doi.org/10.1007/s00221-019-05511-

Smalt, C. J., Heinz, M. G., & Strickland, E. A. (2014). Modeling the time-varying and level-dependent effects of the medial olivocochlear reflex in auditory nerve responses. JARO - Journal of the Association for Research in Otolaryngology, 15(2), 159–173. https://doi.org/10.1007/s10162-013-0430-z

Yasin, I., Drga, V., Liu, F., Demosthenous, A., & Meddis, R. (2020). Optimizing speech recognition using a computational model of human hearing: effect of noise type and efferent time constants. IEEE Access, 8, 56711-56719.

Schofield, B. R. (2011). Central descending auditory pathways. In Auditory and vestibular efferents (pp. 261–290). Springer.

Farhadi, A., Jennings, S. G., Strickland, E. A., & Carney, L. H. (2022). Subcortical Auditory Model including Efferent Dynamic Gain Control with Inputs from Cochlear Nucleus and Inferior Colliculus. bioRxiv, 2022-10.

Gummer, M., Yates, G. K., & Johnstone, B. M. (1988a). Modulation transfer function of efferent neurones in the guinea pig cochlea. Hearing Research, 36(1), 41–51. https://doi.org/10.1016/0378-5955(88)90136-0

Huffman, R. F., & Henson Jr, O. W. (1990). The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus. Brain Research Reviews, 15(3), 295–323.

Brennan, M. A., Svec, A., Farhadi, A., Maxwell, B. N., & Carney, L. H. (2023). Inherent envelope fluctuations in forward masking: Effects of age and hearing loss. The Journal of the Acoustical Society of America, 153(4), 1994-1994.

Maxwell, B. N., Farhadi, Brennan, M. A., Svec, & Carney, L. H. (2023). Simulating Physiological and Psychoacoustic Forward Masking in a Subcortical Model with Efferent Gain Control. Abstract, Association for Research in 553 Otolaryngology, SA200, pg.301.

Farhadi, A., Agarwalla, S. & Carney, L. H. (2023). A Subcortical Auditory Model With Efferent 552 Gain Control Explains Perceptual Enhancement. Abstract, Association for Research in 553 Otolaryngology, SU198, pg.449.

Farhadi, A., & Carney, L. H. (2023). Predicting thresholds in an auditory overshoot paradigm using a computational subcortical model with efferent feedback. In WASPAA

2023. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY.

Liberman, M. C., & Brown, M. C. (1986). Physiology and anatomy of single olivocochlear neurons in the cat. Hearing research, 24(1), 17-36.

Brown, M. C. (2014). Single-unit labeling of medial olivocochlear neurons: the cochlear frequency map for efferent axons. Journal of neurophysiology, 111(11), 2177-2186.

Zilany, M. S., Bruce, I. C., Nelson, P. C., & Carney, L. H. (2009). A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. The Journal of the Acoustical Society of America, 126(5), 2390-2412.

Zilany, M. S., Bruce, I. C., & Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. The Journal of the Acoustical Society of America, 135(1), 283-286.

Zhang, X., Heinz, M. G., Bruce, I. C., & Carney, L. H. (2001). A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. The Journal of the Acoustical Society of America, 109(2), 648-670.

Tan, Q., & Carney, L. H. (2003). A phenomenological model for the responses of auditory-nerve fibers. II. Nonlinear tuning with a frequency glide. The Journal of the Acoustical Society of America, 114(4), 2007-2020.

Zilany, M. S., & Bruce, I. C. (2006). Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. The Journal of the Acoustical Society of America, 120(3), 1446-1466.

Carney, L. H., Zilany, M. S., Huang, N. J., Abrams, K. S., & Idrobo, F. (2014). Suboptimal use of neural information in a mammalian auditory system. Journal of Neuroscience, 34(4), 1306-1313.

Mao, J., Vosoughi, A., & Carney, L. H. (2013). Predictions of diotic tone-in-noise detection based on a nonlinear optimal combination of energy, envelope, and fine-structure cues. The Journal of the Acoustical Society of America, 134(1), 396–406.

Kim, D. O., Carney, L., & Kuwada, S. (2020). Amplitude modulation transfer functions reveal opposing populations within both the inferior colliculus and medial geniculate body. Journal of Neurophysiology, 124(4), 1198-1215.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. the Journal of the Acoustical Society of America, 67(3), 971-995.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., ... & Ludvigsen, C. (1994). An international comparison of long-term average speech spectra. The journal of the acoustical society of America, 96(4), 2108-2120.

Chapter 5. Summary and Future Work

This thesis proposed an auditory subcortical model that simulated both ascending and descending neural projections in the auditory system by including a model for the MOC efferent system. The impact of incorporating the MOC efferent system into an existing ascending subcortical model was evaluated in predicting human behavior in different psychoacoustic forward and simultaneous masking tasks. The simulation results showed better performance for the model that included efferent feedback in terms of accuracy in estimating human subjects' thresholds for the studied psychoacoustic tasks. Additionally, this study explored the effect of the MOC efferent system activity on the neural coding of speech-shaped signals. We tested the hypothesis that the efferent system enhances the representation of vowel-like sounds in neural responses. Simulation results of neural response the speech shape stimuli aligned with our hypothesis.

5.1. Summary and Novel results

The first chapter of this thesis provides background information on the auditory system to aid in understanding the subsequent chapters. It briefly discusses the nonlinearities found in a healthy auditory system and their potential benefits for speech perception in healthy auditory system. Chapter one also emphasizes the significance of the MOC efferent system in hearing research and the importance of incorporating this feedback in auditory modeling. The remainder of the chapter provides a brief overview of the three main aims of the thesis. Aim 1 focuses on modeling the MOC efferent system, aim 2 involves testing and refining the model using psychoacoustic data, and aim 3 examines the potential impact of efferent feedback on the neural coding of speech.

Chapter 2 provides a detailed explanation of the proposed MOC efferent system model. It discusses the functionality and importance of the two feedback loops, one from IC cells in the midbrain and the other from WDR cells in CN, that are integrated into the MOC model. Additionally, the chapter explains how the model's structure was modified to enable the inclusion of these feedback loops. The modification involved transitioning from processing the entire stimulus at once to implementing it sample-by-sample. The chapter also describes how the MOC activity was considered when reevaluating an existing physiological study. In this study, neural recordings from IC cells in the midbrain were analyzed in response to AM noise, focusing on changes in the response over time instead of average rate. This reevaluation reveals an interesting trend of increasing neural response over time. We hypothesized that this trend may be attributed to MOC activity. To adjust the model's parameters, we utilized this dataset and compared the performance of the proposed model with the previous model in simulating the observed trend from the physiological study. Our simulation results support the hypothesis that the observed trend of increasing rate over time in the physiological dataset may be a result of MOC efferent activity. Additionally, the results demonstrate that the model incorporating the MOC efferent system more accurately simulates the physiological responses tested in this study. Furthermore, tests were conducted on the proposed model to ensure that it can accurately simulate the neural properties of the auditory system predicted by the previous model.

The proposed model was evaluated to predict human listeners' performance in various psychoacoustic tasks. The inclusion of the MOC efferent system in the model resulted in improved accuracy compared to the previous model when estimating listeners' thresholds.

Some of these findings have been published in collaborative works, as noted in the appendix. Chapter 3 of the thesis is dedicated to an independent study that focuses on examining the model's performance, specifically with the MOC efferent system, in predicting human listeners' thresholds in a simultaneous tone in noise detection task. In this task, a brief target tone is introduced either immediately after the noise onset or with a delay. The model that incorporates efferent feedback successfully predicts the observed enhancement in human listeners' performance when the target tone is delayed, while the model lacking efferent feedback fails to exhibit this trend.

Chapter 4 investigated the impact of the efferent system on speech coding. Initially, the chapter explored the neural coding of speech using the auditory model, focusing on the quality of neural fluctuation cue as a measure for speech perception. Subsequently, the effect of MOC activity on the neural fluctuation coding was investigated. In line with our hypothesis, it was found that the efferent system enhanced speech perception by sharpening the neural fluctuation coding in response to speech-shaped stimuli. This project specifically examined single formant stimulus and synthesized vowel. We analyzed the impact of the efferent system on the robustness of neural fluctuation profile in the presence of background noise and at various sound levels. The model incorporating efferent system exhibited greater robustness compared to the model without efferent under challenging conditions, including different noise backgrounds, sound levels, and signal-to-noise ratios (SNRs). As a result, the model with efferent more accurately predicted the performance of the normal hearing system in such conditions compared to the model without efferent.

In summary, compared to previous models for the auditory system, the proposed model demonstrates better performance in simulating the trends observed in both physiological and psychoacoustic datasets. This model is a physiologically plausible tool for investigating the role of the MOC efferent system and has significant potential for predicting various physiological and behavioral hypotheses facilitating a deeper understanding of the auditory system and its functionality. Finally, this model has the potential to contribute to advancements in auditory devices and hearing aids.

5.2. Future work

The research in the area of auditory efferent system is highly intriguing, as it has been largely unexplored compared to the ascending auditory pathway. There are still so many questions unanswered regarding the anatomy, functionality and role of the efferent system on perception of sounds. Studying such a complicated system with complex circuitry, including projections from various stages of the auditory pathway, is very challenging especially since some of these projections are not accessible for physiologically recording in awake animals (Guinan, 2018, Schofield, 2011). However, despite the challenges, there is ongoing excitement about the potential avenues of research in this field (Jennings, 2021). Exploring the efferent system further can provide a better understanding of the human auditory system as a whole, including both the ascending and also the descending pathways.

5.2.1. Employing the proposed auditory model for studying the MOC efferent system

An auditory model with inclusion of the MOC efferent system can be a powerful tool for studying the efferent system (Meddis et al., 2010). The predicted responses from the auditory model can be insightful for designing physiological and psychoacoustical experiments that can lead into more information about the efferent system. Recognizing the role of the efferent system in auditory processing can lead to a new way of analyzing and thinking about the existing physiological and psychoacoustic studies specifically in unanesthetized measurements (Aedo et al., 2015; A. R. Chambers et al., 2012; Guitton et al., 2004). As an example, we started adjusting the parameter of the proposed efferent model by re-analyzing the physiological dataset of IC response to AM noise that was previously used and analyzed for different purposes (Carney et al., 2014). Keeping the efferent system activity in our mind, we managed to look at this dataset in a different way, which was to look at the responses over the stimulus time that are longer than the suggested sluggishness suggested for efferent system (Backus & Guinan, 2006; Salloom et al, 2023; Roverud & Strickland, 2010; Warren & Liberman, 1989). This method of physiological analysis, and keeping the efferent activity in perspective, can be applied to other existing physiological studies specially for studies in awake animal and can bring in more insight about the auditory system and also efferent system functionality, specifically. It could be a very interesting future direction to reevaluate some existing dataset, but by considering the effect of the efferent system and observing the changes over time in the neural responses.

Auditory models incorporating the efferent system can offer valuable insights into the impact of the efferent system on various psychoacoustic tasks. Through the modeling work presented in this study, along with our anatomical and physiological knowledge, we now understand that temporal fluctuations in the stimulus strongly influence efferent system activity (Brennen et al., 2023). The modulation present in the stimulus can translate into modulation in neural responses of auditory nerve fibers, ultimately exciting the IC BE cells and increasing efferent activity (Carney, 2018, Farhadi et al 2023, Kim et al., 2015, 2020). With this understanding, we can design psychoacoustic studies to gain a better understanding of the efferent system and the auditory system as a whole. By using the effect of stimulus fluctuation on MOC efferent activity, we can stimulate MOC activity in a more natural way compared to electrical stimulation. In such experiments, altering the modulation depth and sound level may potentially control the extent of MOC stimulation. Computational models can help in testing whether human behavior aligns with model predictions, allowing us to evaluate our hypotheses and assumptions about the efferent system.

Furthermore, the modeling work can assist in the accurate design of noninvasive methods for studying the efferent system, such as investigating otoacoustic emissions (Guinan et al, 2003; Lee and Lewis, 2023), while ensuring that the measurements do not significantly impact the observations. In other words, by employing these models, we can comprehend how the stimulus itself can influence efferent activity when studying this system. For instance, when examining the efferent system using a pulse of clicks as a stimulus, it is crucial to consider that the modulation present in the stimulus, caused by the

timing between the clicks, may greatly impact the efferent activity rather than solely the nature of the clicks themselves. So, in summary one area of future work and contribution of this model could be to reevaluate the datasets and the experimental design, considering the efferent activity, and using the model for designing new experiments at different levels of the auditory pathway by first testing the hypotheses using the model.

5.2.2. Using proposed auditory model for understanding the mechanism of hearing in noise

Utilizing the proposed auditory model for investigating speech and speech-in-noise neural processing presents exciting possibilities for future research. By using the model with MOC efferents, along with the speech intelligibility tests (Zaar and Carney, 2022; Hamza et al., 2023), we can test various hypotheses related to speech intelligibility in noise and the relationship between speech intelligibility and efferent activity across different hearing conditions. This approach will aid in designing experiments that provide deeper insights into these effects and facilitate further exploration of this topic. Furthermore, age-related hearing loss can be explored using this platform considering the findings from our study (Brennen et al., 2023) ,which examined the impact of Gaussian noise (GN) and low-fluctuation noise (LFN) maskers on forward masking, considering the effects of age and sensorineural hearing loss (SNHL). The research involved three groups: younger listeners with normal hearing (YNH), older participants with normal or near-normal hearing (ONH), and older participants with SNHL (OSNHL). The findings suggested that older individuals with normal or near-normal hearing might be more vulnerable to the disturbing effects of masker-temporal fluctuations compared to younger individuals with normal hearing. Our

proposed computational model with inclusion of the efferent system showed that efferent feedback might have a significant influence on the fluctuation-related disruption. These findings have the potential to offer valuable insights into the underlying mechanisms of hidden hearing loss (Liberman, 2015). It is fascinating to utilize psychoacoustic studies to investigate the impact of age on efferent system functionality (Kim et al 2002). By observing whether the decline in efferent system functionality contributes to age-related hearing impairments, even when the audiogram and audibility appear to be within a healthy range, we can gain significant understanding. Interestingly, based on our simulation results, we observed that the efferent system could potentially degrade listeners' performance in simple auditory tasks like tone detection. However, when it comes to more complex tasks like speech in noise, the efferent system in a healthy auditory system can actually enhance listeners' performance. By considering all these aspects together, we can explore whether efferent system dysfunction is a potential factor, or at least one of the factors, underlying hidden hearing loss, where audibility is comparable to a healthy auditory system but speech intelligibility in noise is compromised. Studying the impact of efferent system dysfunction in hearing loss conditions can provide insights into individual differences. Since the efferent system functions differently in each person due to varying degrees of hearing impairment, it can potentially affect hearing abilities in more challenging situations compared to basic tone detection measured in an audiogram. While the audiogram may appear similar among individuals, the variation in efferent system operation can play a significant role in more difficult listening condition.

5.2.3. Application of the proposed auditory model for improving hearing devices

As emphasized in this study, incorporating detailed physiological information into models not only enhances our understanding of mechanism in hearing and allows for testing more complex hypotheses, but it also has the potential to improve the performance of applications that rely on such models. With increasing knowledge about efferent system functionality and anatomical details, we can refine physiologically based auditory models. These more accurate and advanced models can be used in various auditory domains, including hearing aids, cochlear implants, automatic speech recognition devices, and augmented reality/virtual reality applications. One potential future direction for this study could involve using the more physiologically accurate model to design auditory devices that simulate the functionality of a normal, healthy auditory system. This goal can be accomplished by optimizing the characteristics of these devices in such a way that the output sound, when processed through the model of the auditory system, generates neural responses that closely resemble those of a desired healthy auditory system. For instance, the gain of different frequency channels in a hearing aid could be adjusted when combined with a model representing the hearing-impaired system to achieve similar neural responses to responses of a normal hearing model at various stages of the auditory pathway (Drakopoulos and Verhulst, 2023) Similar approaches could also be applied to the design of other auditory devices, including cochlear implants, although this approach may pose additional challenges due to resource limitations in cochlear implant users. One problem with the configuration of the proposed model with the efferent pathways is computational efficiency. The speed of the model is not a major issue for research purposes; however, the

speed could be a challenging issue for developing a new signal-processing algorithm for hearing-aid systems. A real-time system with high processing speed is a requirement for this application. One possible direction for addressing this issue is to train a deep neural network using model responses to different challenging speech datasets (Drakopoulos et al., 2021). The response of the model to new stimuli could then be achieved by running a convolutional neural network, which is a fast approach that could potentially be included in specific hardware for a hearing-aid device

5.2.4 Improving, refining, and extending the proposed efferent model

We have proposed a subcortical auditory model that incorporates the MOC efferent system. While this model includes novel projections from the IC to the MOC (Guinan, 2018), it still lacks certain details of the efferent system. This initial generation of the model undoubtedly requires improvement. Several aspects of the proposed auditory model can be enhanced. Firstly, the model parameters can be more accurately optimized. Parameters such as the bandwidth of projections to MOC neurons, the weights or effects of different projections on MOC neurons based on their source (IC or CN), the time constant for these input projections, and the tuning of projections from MOC neurons to OHCs, along with the timing and strength of these projections, can all be adjusted by modifying the model's parameters. Although we have adjusted some of these parameters using physiological data, there may still be additional degrees of freedom in terms of the model's parameters. Furthermore, the physiological dataset used for parameter adjustment, while promising, is not as complex as real-world stimuli. Re-optimizing the model parameters using various and potentially more intricate studies, and exploring state-of-the-art optimization methods,

can further enhance the accuracy of the model. Neural responses to wideband tone in noise, as well as neural recordings of the MTF in awake animals, and noninvasive otoacoustic emissions (Guinan, 2006) in response to acoustic stimuli from human listeners, could all be employed for this purpose.

Aside from optimizing the parameters of the existing implementation, a future direction worth exploring is the inclusion of additional projections in the efferent system within the model. In the current proposed model, we incorporated projections from IC cells and brainstem to the MOC neurons (Guinan, 2018). However, there are also direct projections from the auditory cortex to the MOC neurons, as well as indirect projections through IC cells (Malmierca and Ryugo, 2010; Delano and Elgoyhen, 2016). These additional projections may play a significant role in cognitive aspects of hearing, such as the effects of attention (Garinis et al., 2011). The proposed model presented here focuses on subcortical processing and does not include cortical processing. Deep neural network models have been used to simulate the auditory cortex (Saddler et al., 2021; Francl and McDermott, 2022). but auditory periphery process has been overlooked in such studies. By combining cortical and subcortical models, we can aim for more accurate predictions of human behavior in auditory tasks. It is crucial to consider each of these components, as they play significant roles in human auditory perception. Integrating the subcortical model with a cortical model opens up possibilities to study effects like informational masking and learning process, providing a more comprehensive understanding of auditory processing. Additionally, incorporating cortical projections to MOC neurons within the auditory model could enable the investigation of attentional effects in human listeners. Although this

project poses challenges in both modeling and experimental aspects of studying attention, it holds the potential to provide valuable insights into this aspect of the auditory system. Furthermore, it could inspire the development of hearing aids with algorithms designed to enhance selective listening, potentially addressing auditory challenges such as the "cocktail party problem.

Another important component of the efferent system that is not currently included in this model is the lateral olivocochlear (LOC) system (Guinan, 2018). The LOC system consists of projections from LOC neurons to the inner hair cells. The role of this projection in the auditory system is still not well understood but reducing the acoustic-trauma and balancing the outputs from the left and right ears to enable binaural localization are the two roles suggested for this system suggested. Studying the LOC system poses additional challenges compared to the MOC system. As recording and stimulating LOC neurons is more difficult due to its specific characteristics such as thin and unmyelinated axons and as a result there are no recordings from LOC neurons in anesthetized, intact condition. (Guinan, 2011, Brown, 2011, Romero and Trussell, 2022). Despite these challenges, gaining more insights into the LOC system can lead to the inclusion of this system in the proposed subcortical auditory model, resulting in a more comprehensive model that incorporates both types of olivocochlear efferent systems. It is worth noting that apart from the discussed projections, there are other descending projections, both peripheral and central, that can be incorporated into the model as our understanding of the anatomy and physiology of the efferent system improves.

Another intriguing future project is the development of a binaural auditory model that incorporates both ipsilateral and contralateral MOC efferent projections. By including these projections, a more accurate model can be achieved, enabling the study of the specific effects of ipsilateral and contralateral efferent activity. While the contralateral projections have been more extensively studied due to their straightforward measurement, the ipsilateral projections are stronger but less explored (Guinan, 2018). Having a binaural auditory model can facilitate the testing of more complex hypotheses in hearing research and enable the design of experiments to investigate the contralateral and ipsilateral MOC activity separately. Furthermore, incorporating binaural cues into the model can enhance its accuracy in studying binaural phenomena such as sound localization and the effects of efferent activity in tasks like release from masking (Litovsky, 2012). Notably, the inclusion of binaural stages in the auditory model (Klug et al 2020; Cai et al 1998), including the projections from the lateral superior olive (LSO) and the medial superior olive (MSO) to the inferior colliculus (IC) cells (Park, 1998; Tollin, 2003, Nordeen et al., 1983), can provide valuable insights into studying the efferent system's functionality.

Bibliography

Guinan Jr, J. J. (2018). Olivocochlear efferents: Their action, effects, measurement and uses, and the impact of the new conception of cochlear mechanical responses. Hearing Research, 362, 38–47.

Schofield, B. R. (2011). Central descending auditory pathways. In Auditory and vestibular efferents (pp. 261–290). Springer.

Jennings, S. G. (2021). The role of the medial olivocochlear reflex in psychophysical masking and intensity resolution in humans: A review. Journal of Neurophysiology, 125(6), 2279-2308

Meddis, Ray, Lopez-Poveda, E. A., Fay, R. R., & Popper, A. N. (2010). Computational models of the auditory system. Springer.

Aedo, C., Tapia, E., Pavez, E., Elgueda, D., Delano, P. H., & Robles, L. (2015). Stronger efferent suppression of cochlear neural potentials by contralateral acoustic stimulation in awake than in anesthetized chinchilla. Frontiers in Systems Neuroscience, 9, 21.

Chambers, A. R., Hancock, K. E., Maison, S. F., Liberman, M. C., & Polley, D. B. (2012). Sound-evoked olivocochlear activation in unanesthetized mice. Journal of the Association for Research in Otolaryngology, 13(2), 209–217.

Guitton, M. J., Avan, P., Puel, J.-L., & Bonfils, P. (2004). Medial olivocochlear efferent activity in awake guinea pigs. Neuroreport, 15(9), 1379–1382

Carney, L. H., Zilany, M. S. A., Huang, N. J., Abrams, K. S., & Idrobo, F. (2014). Suboptimal use of neural information in a mammalian auditory system. Journal of Neuroscience, 34(4), 1306–1313.

B. C. Backus and J. J. Guinan Jr, “Time-course of the human medial olivocochlear reflex,” in The Journal of the Acoustical Society of America, vol. 119, no. 5, pp. 2889-2904, 2006.

W. B. Salloom, H. Bharadwaj, and E. A. Strickland, "The effects of broadband elicitor duration on a psychoacoustic measure of cochlear gain reduction," J. Acoust. Soc. Am., vol. 153, no. 4, pp. 2482-2482, 2023.

E. H. Warren III and M. C. Liberman, “Effects of contralateral sound on auditory-nerve responses. I. Contributions of cochlear efferents,” Hearing Research, vol. 37, no. 2, pp. 89-104, 1989.

Brennan, M. A., Svec, A., Farhadi, A., Maxwell, B. N., & Carney, L. H. (2023). Inherent envelope fluctuations in forward masking: Effects of age and hearing loss. The Journal of the Acoustical Society of America, 153(4), 1994-2005.

Carney, L. H. (2018). Supra-threshold hearing and fluctuation profiles: implications for sensorineural and hidden hearing loss. Journal of the Association for Research in Otolaryngology, 19(4), 331-352.

Farhadi, A., Jennings, S. G., Strickland, E. A., & Carney, L. H. (2022). Subcortical Auditory Model including Efferent Dynamic Gain Control with Inputs from Cochlear Nucleus and Inferior Colliculus. bioRxiv, 2022-10.

Kim, D. O., Carney, L., & Kuwada, S. (2020). Amplitude modulation transfer functions reveal opposing populations within both the inferior colliculus and medial geniculate body. Journal of Neurophysiology, 124(4), 1198–1215.

Kim, D. O., Zahorik, P., Carney, L. H., Bishop, B. B., & Kuwada, S. (2015). Auditory distance coding in rabbit midbrain neurons and human perception: monaural amplitude modulation depth as a cue. Journal of Neuroscience, 35(13), 5360-5372.

Guinan, J. J., Backus, B. C., Lilaonitkul, W., & Aharonson, V. (2003). Medial olivocochlear efferent reflex in humans: otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs. Journal of the Association for Research in Otolaryngology, 4, 521-540.

Lee, D., & Lewis, J. D. (2023). Inter-Subject Variability in the Dependence of Medial-Olivocochlear Reflex Strength on Noise Bandwidth. Ear and Hearing, 44(3), 544-557.

Zaar, J., & Carney, L. H. (2022). Predicting speech intelligibility in hearing-impaired listeners using a physiologically inspired auditory model. Hearing Research, 426, 108553.

Hamza, Y., Farhadi, A., Schwarz, D. M., McDonough, J. M., & Carney, L. H. (2022). Representations of fricatives in sub-cortical model responses: comparisons with human consonant perception. BioRxiv, 2022-10.

Liberman, M. C. (2015). Hidden hearing loss. *Scientific American*, *313*(2), 48-53.

Kim, S., Frisina, D. R., & Frisina, R. D. (2002). Effects of age on contralateral suppression of distortion product otoacoustic emissions in human listeners with normal hearing. Audiology and Neurotology, 7(6), 348-357.

Drakopoulos, F., & Verhulst, S. (2023). A neural-network framework for the design of individualized hearing-loss compensation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.

Drakopoulos, F., Van Den Broucke, A., & Verhulst, S. (2023, June). A DNN-based hearing-aid strategy for real-time processing: One size fits all. In ICASSP 2023-2023 IEEE

International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

Malmierca, M. S., & Ryugo, D. K. (2010). Descending connections of auditory cortex to the midbrain and brain stem. In The auditory cortex (pp. 189-208). Boston, MA: Springer US

Delano, P. H., & Elgoyhen, A. B. (2016). Auditory efferent system: new insights from cortex to cochlea. Frontiers in Systems Neuroscience, 10, 50.

Garinis, A. C., Glattke, T., & Cone, B. K. (2011). The MOC reflex during active listening to speech.

Saddler, M. R., Gonzalez, R., & McDermott, J. H. (2021). Deep neural network models reveal interplay of peripheral coding and stimulus statistics in pitch perception. Nature communications, 12(1), 7278.

Francl, A., & McDermott, J. H. (2022). Deep neural network models of sound localization reveal how perception is adapted to real-world environments. Nature human behaviour, 6(1), 111-133.

Guinan, J. J. (2011). Physiology of the medial and lateral olivocochlear systems. Auditory and vestibular efferents, 39-81.

Brown, M. C. (2011). Anatomy of olivocochlear neurons. Auditory and vestibular efferents, 17-37.

Romero, G. E., & Trussell, L. O. (2022). Central circuitry and function of the cochlear efferent systems. Hearing research, 425, 108516.

Litovsky, R. Y. (2012). Spatial release from masking. Acoust. Today, 8(2), 18-25.

Klug, J., Schmors, L., Ashida, G., & Dietz, M. (2020). Neural rate difference model can account for lateralization of high-frequency stimuli. The journal of the Acoustical Society of America, 148(2), 678-691.

Cai, H., Carney, L. H., and Colburn, H. S. (1998). "A model for binaural response properties of inferior colliculus neurons. II. A model with interaural time difference-sensitive excitatory and inhibitory inputs and an adaptation mechanism," J. Acoust. Soc. Am. 103, 494–506.

Park, T. J. (1998). IID sensitivity differs between two principal centers in the interaural intensity difference pathway: the LSO and the IC. Journal of Neurophysiology, 79(5), 2416-2431.

Tollin, D. J. (2003). The lateral superior olive: a functional role in sound source localization. The neuroscientist, 9(2), 127-143.

Nordeen, K. W., Killackey, H. P., & Kitzes, L. M. (1983). Ascending auditory projections to the inferior colliculus in the adult gerbil, Meriones unguiculatus. Journal of Comparative Neurology, 214(2), 131-143.