Grange, Jacques , Zhang, Mengchao and Culling, John 2022. The role of efferent reflexes in the efficient encoding of speech by the auditory nerve. Journal of Neuroscience 42 (36) , pp. 6907-6916. 10.1523/JNEUROSCI.2220-21.2022

Publishers page: https://doi.org/10.1523/JNEUROSCI.2220-21.2022

Title page

Title: The Role of Efferent Reflexes in the Efficient Encoding of Speech by the Auditory Nerve

Abbreviated title: efferent reflexes in efficient encoding of speech

Authors: Jacques Grange, Mengchao Zhang (张梦超), John Culling

Affiliation: School of Psychology, Cardiff University, United Kingdom, CF10 3AT

Corresponding author: grangeja@cardiff.ac.uk

Number of pages: 33

Number of figures: 7

Number of tables: 2

Number of words for abstract: 232

Number of words for introduction: 701

Number of words for discussion: 1450

Conflict of interest statement: the authors declare no competing financial interests.

Acknowledgments: This research was funded by an EPSRC project grant (Grant No.: EP/R010722/1; P.I.: John Culling).

ABSTRACT

To avoid information loss, the auditory system must adapt the broad dynamic range of natural sounds to the restricted dynamic range of auditory nerve fibers. How it solves this dynamic range problem is not fully understood. Recent electrophysiological studies showed that dynamic-range adaptation occurs at the auditory-nerve level, but the amount of adaptation found was insufficient to prevent information loss. We used the physiological MATLAB® Auditory Periphery model to study the contribution of efferent reflexes to dynamic range adaptation. Simulating the healthy human auditory periphery provided adaptation predictions that suggest that the acoustic reflex shifts rate-level functions towards a given context level and the medial olivo-cochlear reflex sharpens the response of nerve fibers around that context level. A simulator of hearing was created to decode model-predicted firing of the auditory nerve back into an acoustic signal, for use in psychophysical tasks. Speech reception thresholds in noise obtained with a normal-hearing implementation of the simulator were just 1 dB above those measured with unprocessed stimuli. This result validates the simulator for speech stimuli. Disabling efferent reflexes elevated thresholds by 4 dB, reaching thresholds found in mild-to-moderately hearing-impaired individuals. Overall, our studies suggest that efferent reflexes may contribute to overcoming the dynamic range problem. Because specific sensorineural pathologies can be inserted in the model, the simulator can be used to obtain the psychophysical signatures of each pathology, thereby laying a path to differential diagnosis.

SIGNIFICANCE STATEMENT

The saturation of auditory nerve fibers at moderate sound levels seen in rate-level functions challenges our understanding of how sounds of wide dynamic range are encoded. Our physiologically inspired simulations suggest that efferent reflexes may play a major role in dynamic range adaptation, with the acoustic reflex moving auditory-nerve rate level function towards a given context level and the medial olivocochlear reflex increasing fiber sensitivity around that context level. A psychophysical task employing advanced simulations showed how the existence of the efferent system could prevent unrecoverable information loss and severe impairment of speech-in-noise intelligibility. These findings illustrate how important the precise modeling of peripheral compression is to both simulations and understanding of normal and impaired hearing.

INTRODUCTION

The dynamic range of an auditory neuron is the portion of its rate-level function (RLF), where its firing rate increases with the input level. Most sounds important to humans, such as speech and music, are highly modulated in amplitude by nature. Changes in firing rate, combined with frequency tuning, is the most straightforward mechanism by which these spectro-temporal modulations in the stimulus might be encoded on the auditory nerve (AN). However, traditional physiological measurements of AN rate-level functions (RLFs) indicate that most AN fibers are already saturated at moderate sound levels (Liberman, 1978; Winter et al., 1990), prompting some researchers to look for alternative codes based on phase-locking, such as the average, localized synchronized rate (Young and Sachs, 1979). Such a timing mechanism, now known to be essential for firing-rate cues at the cortical levels, seemed to be the only way to explain why mammals can continue to process spectral information over a wide dynamic range while AN fibers become saturated. However, recent work has suggested that processes of adaptation lead to a shift of the dynamic range of AN fibers in response to the prevailing sound level in the environment (termed ‘dynamic range adaptation’, DRA), potentially providing a degree of reprieve for firing-rate mechanisms.

Wen et al. (2009) showed such DRA in AN fibers of Cat. As seen in early electrophysiological studies, classical firing-rate adaptation is a decrease in firing rate to a steady tone or repeated stimulation (Kiang et al., 1965; Smith and Zwislocki, 1975; Harris and Dallos, 1979; Smith, 1979; Chimento and Schreiner, 1991). The RLF shows proportional reductions in firing rate at all stimulus levels. In contrast, DRA is defined as a horizontal shift of RLFs toward the sound levels with the highest frequency of occurrence. Somewhat stronger DRA is seen in the inferior colliculus (Dean et al., 2005) and auditory cortex (Watkins and Barbour, 2008). By shifting RLFs so that AN fibers respond best around the context level, DRA enables fibers to encode short-term amplitude changes with variations in response rate across a wide range of sound levels without saturation. The absence of such adaptation is thus expected to weaken an individual’s ability to process normal-level speech in noise.

The mechanisms underlying DRA are unclear. Zilany and Carney (2010) used a phenomenological model of the auditory periphery. They showed that such adaptation could be simulated by applying power-law dynamics at the inner haircell/fiber junction, but this mechanism does not have a physiologically known source. Moreover, although Wen et al. showed DRA at the auditory nerve, the amount of RLF shift (about 0.27 dB/dB) was insufficient to prevent fiber saturation at moderate sound levels. Here, we hypothesize that the efferent reflexes in the auditory periphery, the acoustic reflex and medial olivocochlear reflex (MOCR), have the potential to contribute to DRA. The partial or complete anesthesia-related deactivation of efferent reflexes in electrophysiological studies may have led to substantial underestimation of the amount of adaptation that occurs in an awake state.

These efferent reflexes seem good candidates for DRA, because they both reduce acoustic sensitivity following mid- to high-level sounds. The acoustic reflex contracts the middle-ear muscles and reduces the amplitude of stapes vibrations transferred to the cochlea oval window for intense sound levels (Hung and Dallos, 1972). The MOCR reduces the displacement of the basilar membrane by reducing cochlear amplification by outer hair-cells from moderate sound levels upwards (Guinan and Gifford, 1988).

The current study explores the mechanisms underlying auditory-nerve-level DRA through computational modelling and simulation based on a computer model of the human auditory periphery (Meddis et al., 2013). First, emergent DRA properties of the human model were compared to previous RLF findings in small mammals. The model reveals the distinct role of each efferent reflex, providing a full picture that had previously been partially hidden by anesthesia in electrophysiological studies. Second, a simulator that decodes the modelled auditory-nerve activity back into sound was used to present reconstructions of the stimulus based on the pattern of AN firing to human listeners. Simulations for which the two reflexes were disabled tested how important the reflexes are for speech perception. Human listeners achieved near-normal speech reception thresholds in noise when listening to simulations that included the efferent reflexes.

MATERIALS AND METHODS

A simulator of normal and impaired hearing was created, based on the MATLAB® Auditory Periphery (MAP) model (Meddis et al., 2013). Coined ‘MAPsim’, the simulator employs two modules (Figure 1). The first module is the MAP model, used to encode stimuli at the auditory nerve level. The second module is a decoder that regenerates an acoustic signal based on MAP-encoded auditory-nerve activity. MAP is used to generate RLF predictions and estimate the contribution of efferent reflexes to DRA. MAPsim is used to simulate normal hearing and illustrate the impact of knocking out efferent reflexes on speech-in-noise intelligibility.

Simulation of auditory nerve activity

The stimuli were encoded into simulated auditory nerve activity using the MAP model. MAP is a physiologically inspired computational model of the auditory periphery with a detailed modular structure that has been parameterized to replicate many physiological and psychophysical data sets (e.g. Panda et al., 2014). As shown in the left-hand section of Figure 1 ('MAP/Encoder'), MAP includes: (1) the outer and middle ear filtering, which outputs the stapes displacement, (2) the dual-resonance non-linear (DRNL) model of basilar membrane displacement (Lopez-Poveda and Meddis, 2001), (3) stereocilia flexing and inner hair cell transduction, (4) inner-hair-cell receptor potential, ion currents and neurotransmitter processing, (5) release of neurotransmitter vesicles at the synaptic cleft between inner hair cells and AN fibers, (6) resulting spiking activity of the fibers, (6) two layers of coincidence-detecting MacGregor neurons (MacGregor, 1987) that represent a simplified auditory brainstem network, and (7) the efferent pathways, including a broadband acoustic reflex signal that modulates the stapes displacement and a frequency-specific MOCR signal that differentially modulates the basilar membrane displacement within each best frequency (BF) channel at the DRNL stage.

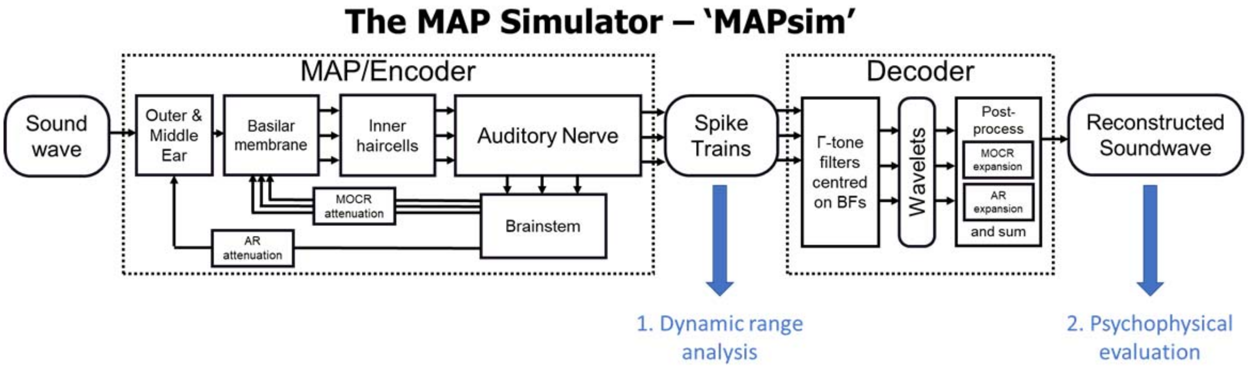

Figure 1. Schematic processing stages of the MAP simulator (MAPsim). Rectangles: signal processing modules of the simulator. Rounded rectangles: input or output signals. One arrow: broadband processing. Three arrows: frequency-specific processing within each BF channel. MAP predicts the AN spike trains of ~30,000 auditory-nerve fibers across 30 BFs and 3 SRs.

The closest model implementation to the current study is in Panda et al. (2014). The parameters to simulate the normal-hearing condition for this study are provided in Table 1. A total of 29,970 AN fibers were arranged over 30 BFs (equally spread on an ERB scale between 56 and 8000 Hz) and 3 levels (low, medium and high) of spontaneous rate (SR), rendering 333 fibers per BF and SR combination. The role of efferent reflexes in efficient coding of sound intensity was first examined through a dynamic range analysis of the encoder.

Module component and Parameter name	**Value**
OME – Outer & Middle ear - two resonance filters [gain order lowpass highpass (Hz)]	[10 1 1000 4000; 25 1 2500 7000]
high-pass stapes filter [order cut-off frequency (Hz)]	[1 600]
air to stapes displacement scalar	45e-9
Acoustic Reflex - latency and smoothing time constant (ms)	10 250
low SR stream threshold (spks/s)	40
broadband rate-to-attenuation factor applied to low SR IC firing rate	5e-3
DRNL - number of best frequencies (BFs)	30
frequency range [low high] over which BFs are ERB-spaced (Hz)	[56 8000]
gain ***a*** & compression exponent ***c***	6e3 0.25
compression knee-point (dB relative to ref. displacement) ***ct***	25
non-linear bandwidth parameters [***p, q***]	[0.14, 180]
linear gain ***g*** and bandwidth parameters [***p, q***]	500 [0.2,235]
linear CFs parameters [***p, q***] with CF = *p**BF + *q*	[0.6,266]
order of all gammatone filters	3
DRNL MOC efferent - latency (ms)	10
maximum attenuation (dB)	35
time constants (s) and weights	[0.055 0.4 1] [0.9 0.1 0]
BF-specific rate-to-attenuation factor applied to IC firing rate	[6e3 0 0]
Inner hair-cell (IHC) cilia/basilar membrane time constant (ms)	1.2
basilar membrane/cilia displacement scalar	0.01
maximum and resting conductance (nS)	6 0.1
displacement sensitivity [***s0 s1***] (nm^{-1}) and offset [***u0 u1***] (nm)	[6 1] [0.3 1]
IHC - endocochlear potential ***Et*** and potassium reversal potential ***Ek*** (mV)	100 -80
potassium conductance ***Gk*** (nS)	21
IHC capacitance ***Cab*** (pF.cm^{-2})	1
combined resistance ***Rpc*** (ohm.cm^{-2})	0.04
IHC Pre-synaptic calcium current - reversal potential ***ECa*** (mV)	66
channel opening parameters [**β*Ca* γ*Ca***]	[400 100]
membrane time constant **τ*M*** (µs)	50
spontaneous-rate fiber types	low, medium and high SR
calcium clearance [low- medium- high-SR] time constants **τ*Ca*** (µs)	[80 120 200]
[Ca^{2+}]^{3} to probability scalar ***z***	45e12
maximum Ca^{2+} conductance (nS)	25
IHC-AN neurotransmitters (3-store model) – maximum vesicles available for release ***M***	17
rate of vesicle replenishment from depletion ***y*** (s^{-1}) and loss from the cleft ***l*** (s^{-1})	8 5
rate of vesicle reprocessing from re-uptake to ready-release pools ***x*** (s^{-1})	100
rate of vesicle re-uptake from cleft to cell ***r*** (s^{-1})	4
AN fiber refractory period (ms)	0.75
Brainstem 1st (CN) & 2nd (IC) order MacGregor chopper cells	
number of input AN/CN fibers per CN/IC cell and of CN cells per BF	10 10
current per spike [CN IC] (nC)	[29 150]
dendritic low-pass cutoff [CN IC] (Hz)	[50 100]
membrane capacitance ***Cm*** (nF)	16.7
potassium recovery time constant ***tauGk*** (ms)	0.5
increment in conductance **Gk** *after spike* (µS)	133
equilibrium threshold ***Th0*** (mV)	10
variable threshold ***tauTh*** and membrane time constant ***tauM*** (ms)	20 2
potassium reversal potential ***Ek***, resting potential ***Er*** and spike height ***Eb*** (mV)	-10 -60 60

Table 1. Parameters for the MAP (v.1_14j_2017) model of the normal auditory periphery.

Dynamic range analyses

The role of the efferent system in DRA at the AN level was examined by comparing the output of the encoder under four efferent conditions. These included the normal-hearing condition (‘normal’) and conditions disabling the acoustic reflex (‘noAR’), the MOCR (‘noMOCR’) and both efferent reflexes (‘noEff’). The parameters in the MAP model to create different efferent-disabled conditions are described below:

(1) To disable the acoustic reflex in MAP, the parameters that determine the minimum number of spikes to activate the reflex, was raised from 40 (‘normal’) to 10^6 spikes/sec so that no attenuation was applied to the stapes displacement.

(2) To disable the MOCR, the DRNL parameter that determines the attenuation strength applied to the basilar membrane displacement in the non-linear path of the DRNL module (DNRLa) was changed from 1 (‘normal’) to 0, effectively deactivating the MOCR.

Based on physiological findings (e.g. Wen et al., 2009), RLFs exhibit DRA when firing rates are probed at various levels along a continuous and silent-free stimulation that sets a *context* level. We expected RLFs to shift closer to the context level when both efferent reflexes are activated (under ‘normal’ simulation). Following the analyses in Wen et al. (2009), our measures included RLFs, normalized RLFs, level at 50% of normalized RLFs, firing rate slope, and sensitivity index δ’.

The RLFs were based on the mean firing at the BF and SR of interest, as a function of probe level. The RLFs were fitted with a four-parameter logistic function:

$$R(L) = R_{min} + \frac{(R_{max} - R_{min})}{1+\exp(-S\cdot(L-\theta_e))} \quad \text{(EQ.1)}$$

where L is the input level; R_{min} and R_{max} are the minimum and the maximum firing rates, respectively; θ_e is the response threshold in dB SPL; S is the slope of the RLF. The least-squared method was used to determine the parameters. The firing rates under each condition were also normalized between 0 and 1 using the following equation:

$$R_{norm} = \frac{R - R_{min}}{(R_{max} - R_{min})} \quad \text{(EQ.2)}$$

The horizontal shift of RLFs was quantified by measuring the increase in the threshold parameter θ_e, the level at which the function reaches half its maximum. Wen et al. (2009) also used rate slope and sensitivity index δ' to examine the impact of rate *variabilities* on the precision of intensity coding along the RLF. The rate slope is the slope of the RLF at a given probe level. Sensitivity index δ', developed by Colburn et al. (2003), is defined as the ratio of the rate slope to the standard deviation (SD) of the rates.

In order to observe the change of RLF shift under various efferent activation conditions, three experimental paradigms were implemented and compared.

A 'baseline' paradigm was used to generate predictions of human RLFs without DRA. This paradigm was similar to those traditionally used in small-mammal electrophysiological studies, where a silent gap preceded each probe, thereby resetting efferent reflexes and hair cells to resting states prior to each measure of firing rate. The probe signal was either a pure tone pip (of frequency matching the fiber's BF) or a broadband noise burst, each 50 ms in duration, with 2-ms rise/fall times and preceded by a 200-ms silence. The probe level spanned 0 – 80 dB SPL for tones and 20 – 100 dB SPL for broadband noise in 4-dB steps. At each probe level, the 50-ms probe was processed through the encoder model, and the mean firing rates were averaged from the activities of all 333 fibers of the same SR and BF.

A second paradigm emulated that employed by Dean at al. (2005) and Wen et al. (2009). In each stimulus, a ‘high probability region’ (HPR) was specified where a range of probe levels occurred more frequently than other probe levels throughout a continuous and silent-free stimulation. The probe signals were the same tone pips or noise bursts as those used in the baseline paradigm. This HPR paradigm differed from that of Wen et al. (2009) in that they used continuous stimulation for 5 minutes, while the computational demands of the MAP model limited our stimuli to 8 seconds. The probe levels (each 50 ms in duration, with 2-ms rise/fall times) were randomly varied over the duration of stimulation, but the ongoing stimulation was always dominated by a range of sound levels centered on a given context level. Specifically, the probe level spanned 0 – 80 dB SPL for tones and 20 – 100 dB SPL for broadband noise in 4-dB steps, but the probe levels inside the HPR occurred 80% of the time while the levels outside of it occurred 20% of the time (Figure 2, left panel). The HPR mean levels were 36, 48, 60, and 72 dB SPL for tonal stimulation and 48, 60, 72, and 84 dB SPL for noise stimulation. Within a stimulation sequence, HPR levels spanned a 12-dB range. During our 8-second stimuli, 160 50-ms probes were presented continuously, and probe levels were assigned in a pre-determined random order (Figure 2, right panel). Ten continuous runs of different level randomizations were completed for each of the four efferent conditions. As in the Wen et al. (2009) studies, the response of a single fiber was recorded. The firing rate was averaged for each probe level and across the ten runs (over a total of 20 occurrences per probe level).

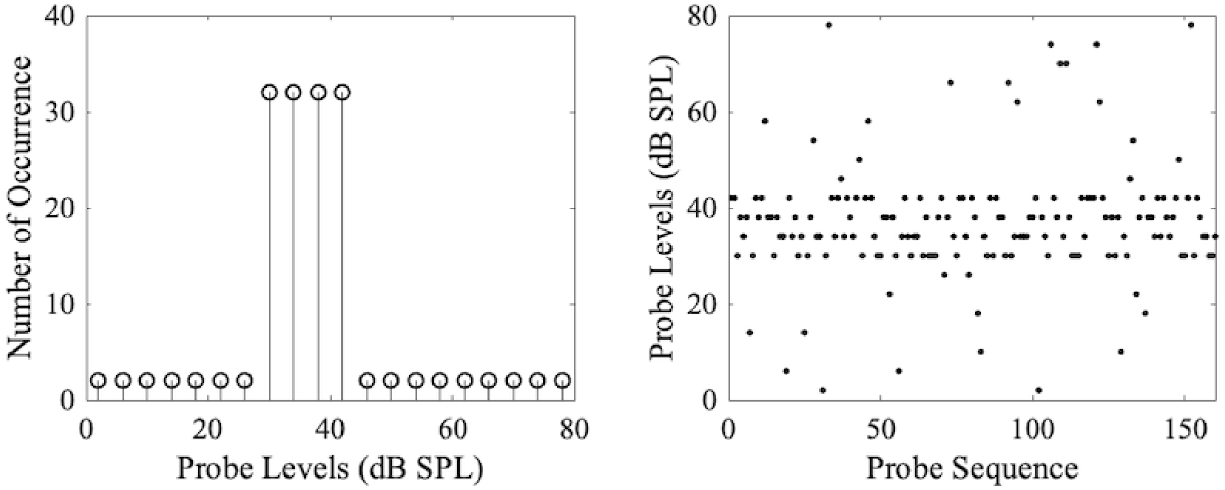

Figure 2. An HPR mean level of 36 dB: (Left) histogram of probe levels and (Right) example of probe level changes during a continuous, 8-second stimulation made of 160 x 50 ms pips/bursts.

The 'precursor' paradigm was employed as a more computationally efficient alternative to the HPR paradigm. The processing of the HPR paradigm at a given HPR level requires a continuous and prolonged signal, usually hundreds of seconds, in order to present a randomized sequence of probe levels to a single fiber. A disadvantage of such processing is that measuring the activity of one fiber among thirty thousand does not make computationally efficient use of the MAP model. Instead, the precursor paradigm employs a steady precursor signal of set duration that immediately precedes a given probe level. For each combination of precursor and probe levels, firing rate is then computed over the 50 ms probe duration as the average firing rate of the 333 AN fibers of same BF and SR, thereby greatly improving computational efficiency. A similar approach is often used in psychophysical studies on the effects of efferent stimulation (e.g. Strickland, 2008). Here, the precursor duration was set long enough (400 ms, with 5 ms rise/fall times) that the modelled efferent reflexes fully stabilized. The 50-ms target probe was presented immediately after this precursor (with 2 ms rise/fall times). The precursor was the same type of sound as the probe (i.e., tones of the same frequency or noises of the same spectrum). The precursor levels were set to the same levels as the HPR paradigm mean levels, following which the probe level was selected between 0 and 80 dB SPL for tones or 20 and 100 dB for noise (in 4-dB steps). As in the baseline paradigm, each 450-ms (precursor + probe) combination was processed through the model independently.

The MAPsim decoder

The purpose of the decoder (right-hand section of Figure 1) in MAPsim is to invert the encoding process and reconstruct the original input signal as well as the encoding stage will allow. The role of the efferent reflexes in the efficient coding of sound can thus be studied psychophysically from the quality of the reconstructed acoustic signal. There are two steps in the decoding stage.

First, the decoder takes in the spike trains from the modelled AN fibers and feeds them through a bank of gammatone filters (fourth order) centered on corresponding BFs to generate wavelets (EQ.3):

$$O_n(t) = \Gamma_n * I_n(t) \qquad \text{(EQ.3)}$$

where n is the BF channel index (1 to 30), t is time, Γ_n is the gammatone filter centered on the channel n BF, $I_n(t)$ is the input AN spike train at time t in channel n, and $O_n(t)$ is the result of the convolution between the input spike train and the gammatone filter (i.e. the resulting gammatone wavelet train) at time t in channel n. Using this approach to decoding, the amplitude envelope of the output waveform is largely determined by the spike rate (and hence the number of wavelets at a given time), while the fine structure of the waveform is determined by the timing of the action potentials (the average wavelet phase).

Second, the wavelet trains are summed across BFs and SRs, as follows. Since the brain has access to efferent signals, we posit that it naturally incorporates them in its interpretation of input signal level. Efferent signals are thus used to re-expand the signal, i.e. to invert most of the compression the cochlear encoder had applied. To implement this re-expansion, the signal at each BF is multiplied by the inverted MOCR attenuation, before summing wavelet trains across BFs and finally multiplying the resulting signal by the inverted acoustic reflex attenuation. The channel-specific, MOCR attenuation, Att_n(t) and the broadband acoustic reflex attenuation, Att_b(t), both time-dependent, are extracted from the MAP model and expansion is implemented according to EQ. 4:

$$O(t) = \frac{\sum_{n=1}^{30}[O_n(t)/Att_n(t)]}{Att_b(t)} \qquad \text{(EQ.4)}$$

Finally, a spectral correction is applied to the reconstructed soundwave for its long-term spectrum to match that of the MAPsim input soundwave. The scripts for the MAP model and the decoder are available on request.

Psychophysical evaluation

If the efferent system is key to DRA, the absence of the system will result in widespread saturation of firing rates and drastically impair the ability to encode and recognize complex spectro-temporal patterns, such as those of speech. Additionally, previous simulations using automatic speech recognition have shown the potential improvement of speech intelligibility in noise under efferent reflexes (Clark et al., 2012). Here, speech recognition in noise with human subjects was used in a perceptual evaluation task to examine the role of efferent reflexes on efficient coding of intensity. The importance of efferent reflexes in MAPsim output quality were assessed through speech reception thresholds (SRTs) in noise. The experiment is designed to measure the beneficial effects of the two compressive efferent reflexes working together. Since these reflexes both act to compress the dynamic range, compensating expansions were explored in order to improve the quality of the output. Since the reconstructed signal from the simulator represents the brain's interpretation of the stimulus, and the brain has access to the reflex signals, it is presumed that it can take them into account. The SRTs were obtained with young normal-hearing adults presented with stimuli that underwent different processing conditions (Table 2).

	Efferent reflexes disabled	Efferent reflexes enabled		
	no eff.	no exp.	MOC exp.	MOC*AR exp.
MOC-based expansion	✗	✗	✓	✓
AR-based expansion	✗	✗	✗	✓

Table 2. Expansion applied under each experimental condition for the processed conditions.

To assess the importance of efferent-based expansion at the decoding stage, with efferent reflexes enabled at the encoding stage, three conditions applied different amounts of expansion. The first applied no expansion to the output of EQ. 3 (called 'no exp.'). The second applied only the EQ. 4 MOCR expansion (called 'MOCR exp.'). The third applied both (EQ.3 and EQ.4) acoustic reflex and MOCR expansions (called 'MOCR*AR exp.'). A control condition ('unproc.') employed the unprocessed, original stimuli. The condition applying the full expansion ('MOCR*AR exp.') was expected to yield SRTs closest to those obtained with unprocessed stimuli, which, if close enough, would constitute a validation of MAPsim. To demonstrate the importance of efferent reflexes, a final condition had both reflexes disabled at the encoding stage (called ‘no eff.’). Since efferent reflexes were disabled, no expansion was applied in this condition. SRTs for the ’no eff.’ condition were compared to those for the ‘MOC*AR exp.’ and ‘unproc.’conditions to measure the impact of knocking out efferent reflexes.

Twelve young adults with self-assessed normal hearing (17-31 years old, mean 22 years old) were recruited from the Cardiff University undergraduate population to perform the SRT task. All participants were briefed in writing and verbally before signing a consent form. All testing and forms complied with the ethical rules of the Cardiff University School of Psychology Institutional Review Board.

SRT measurements employed a digit-triplet recognition task. Each stimulus comprised of a 400-ms precursor followed by three non-repeating, randomly selected digits from 0 to 9 (except disyllabic digit 7) uttered by a British female, each centered within a 700-ms audio file. The precursor was steady-state noise spectrally colored to the female voice, which set the stimulus context level and allowed the efferent reflexes of the MAP model to stabilize. The masker was the same speech-shaped noise as the precursor noise.

SRTs were measured using a one-down-one-up adaptive procedure. In each run, the signal-to-noise ratio (SNR) started with the digits being highly intelligible (at 0 dB SNR) and decreased by a step size of 4 dB as long as correct responses were given. After the first reversal, the step size was reduced to 2 dB. Correct recognition of two or three digits in the correct positions was scored a correct response. Recognition of one or zero digits was scored an incorrect response. The overall level of the speech and the noise mixed was maintained at 65 dB SPL, both at the input and the output of the simulator. Each run stopped when 10 reversals were reached, and the SNRs of all trials over the last 8 reversals were averaged to compute the SRT of that run. The SRT was taken as the average over 3 runs under each condition. Before testing, one practice run using unprocessed stimuli was given to the participants to familiarize them with the task. The practice run was also used to screen for unsuspected participant hearing impairment. The entire experiment took about 1 hour to complete. Participants received payment at the end of the experiment. Repeated-measure analysis of variance was conducted for the SRTs in the IBM SPSS software (version 26.0).

RESULTS

First, the model was used to simulate auditory nerve responses for two cases: baseline *vs.* HPR using tones, and baseline *vs.* HPR using broadband noise. These cases are compared with those from Wen et al. (2009) from their Figures 2 and 4, respectively, so we use simulated nerve fibers that are matched in best frequency and spontaneous rate with the fibers they observed. Second, the results of the precursor paradigm were compared to those of the HPR paradigm using tones to verify that the outcomes were similar. The precursor paradigm was also used to show the responses of the auditory nerves of different spontaneous rates presented with various types of stimuli. Third, the results of the speech-in-noise test were compared under deactivation *vs.* full activation of efferent reflexes and with varying amounts of expansion when efferent reflexes were activated. The validation outcome of the simulator is also reported in this section.

Dynamic range adaptation through the HPR paradigm

Figure 3 shows the average responses of a high SR fiber whose BF matched the probe tone frequency, comparing baseline and HPR-paradigm conditions. The rightmost column shows the physiological data of Wen et al. (2009) collected from a cat fiber responding to 550-Hz tones. The rest of the data were from a simulated high-SR human fiber responding to 580-Hz tones under ‘normal’ (leftmost column), ‘noMOC’ (second left column), ‘noAR’ (middle column), and ‘noEff’ (second right column) processing conditions.

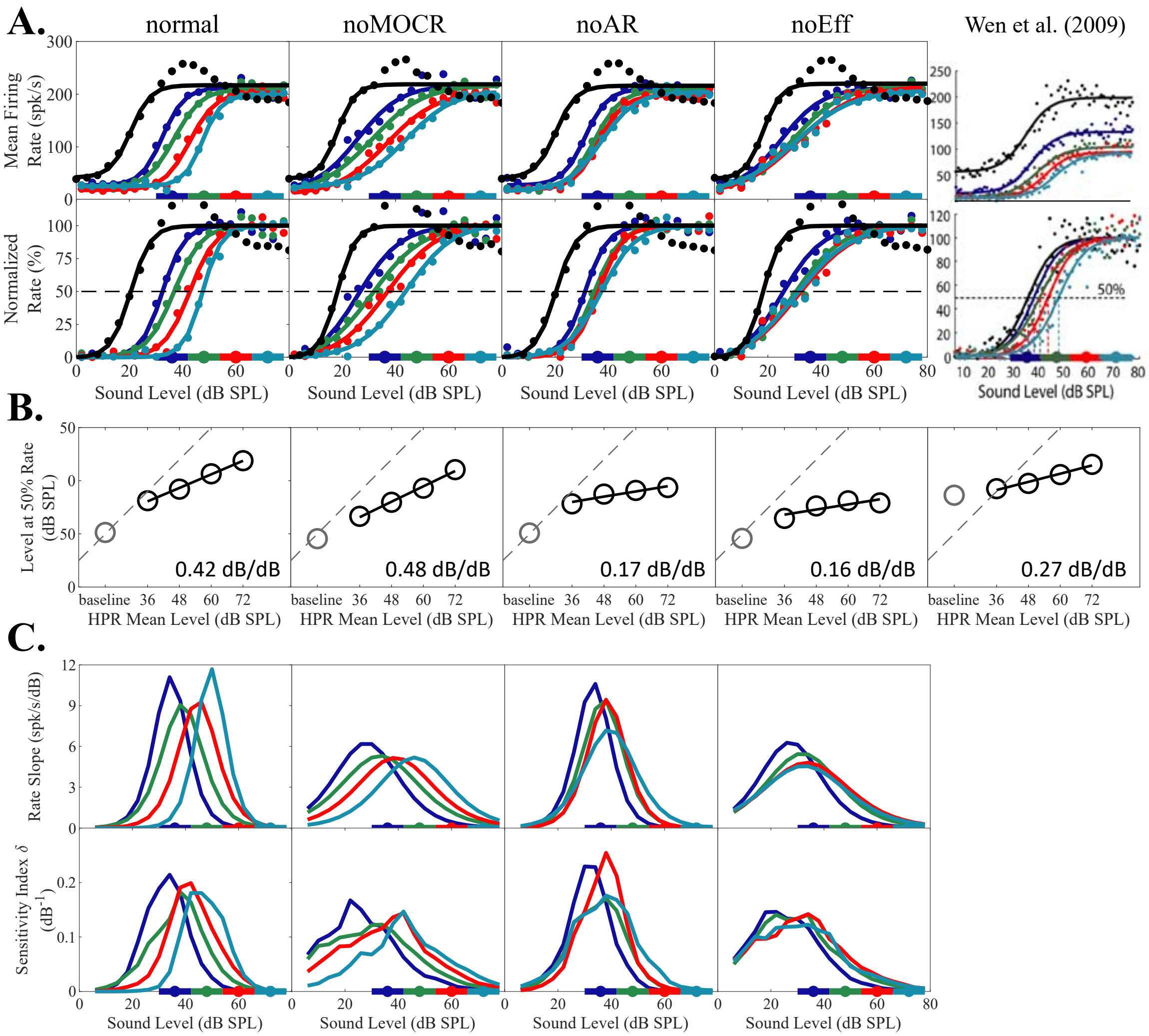

Figure 3. Response of a high SR fiber (BF = 580 Hz) to 580-Hz tones. From left to right panels: modeled human data for normal, noMOCR, noAR and noEff conditions, and Cat electrophysiological data (BF = 550 Hz, tone at 550 Hz, from Figure 2 of Wen et al., 2009, adapted with permission, Copyright © 2009 Society for Neuroscience). Different colored symbols and lines are data points and fitted curves for different HPR levels, indicated by colored

segments on the x-axis, while black is the baseline condition (with no DRA). A: RLFs (top) and normalized RLFs (bottom). B: level at 50% rate. C: rate slope (top) and sensitivity index δ' (bottom).

Under 'normal' condition, the RLFs shift toward the right with increasing HPR levels (Figure 3A), a DRA that was observed in the physiological data. Classical firing rate adaptation, the decrease of the maximum firing rate with increasing HPR level, is minimal in the simulation, but when the RLFs are normalized for maximum firing rate (second row of panels), there is greater DRA than in the physiological data. The 50% point shifts by 0.42 dB per dB change in HPR level for the modeled data, and only by 0.16 dB/dB for the physiological data. As HPR level increases, there is also a clear rightward shift in the peak rate slope and the peak sensitive index *δ'* in the 'normal' condition.

DRA is present under 'noMOCR' condition and reaches 0.48 dB/dB. However, the rate slope and sensitive index *δ'* of 'noMOCR' are shallower compared to 'normal', suggesting a reduction of sensitivity in encoding intensity change. On the other hand, 'noAR' shows a drastically reduced DRA with HPR levels compared to 'normal' or 'noMOCR' conditions, reaching only 0.17 dB/dB. The absence of acoustic reflex does not affect the sensitivity of intensity change coding as its sensitivity indices are comparable to those of 'normal'. Finally, the absence of both acoustic reflexes (i.e., 'noEff') shows combined effects of severe reduction, but not an eradication, of DRA.

Figure 4 shows modeling of the second fiber type measured by Wen et a. (2009): the average response of a medium-SR fiber to broadband noise in the baseline and the HPR paradigm conditions. The BF of the modeled human fiber was selected at 1280 Hz to best match the 1300 Hz BF of the Cat fiber. Overall, the results using noise and a medium-SR fiber are similar to those observed for tones with a high SR fiber, but with two small differences. First, the maximum firing rate decreases more markedly with increasing HPR level. Second, the amount of DRA is larger for noise than for tones, which is 0.55 and 0.52 dB/dB for ‘normal’ and ‘noMOCR’ under noise stimulation, respectively.

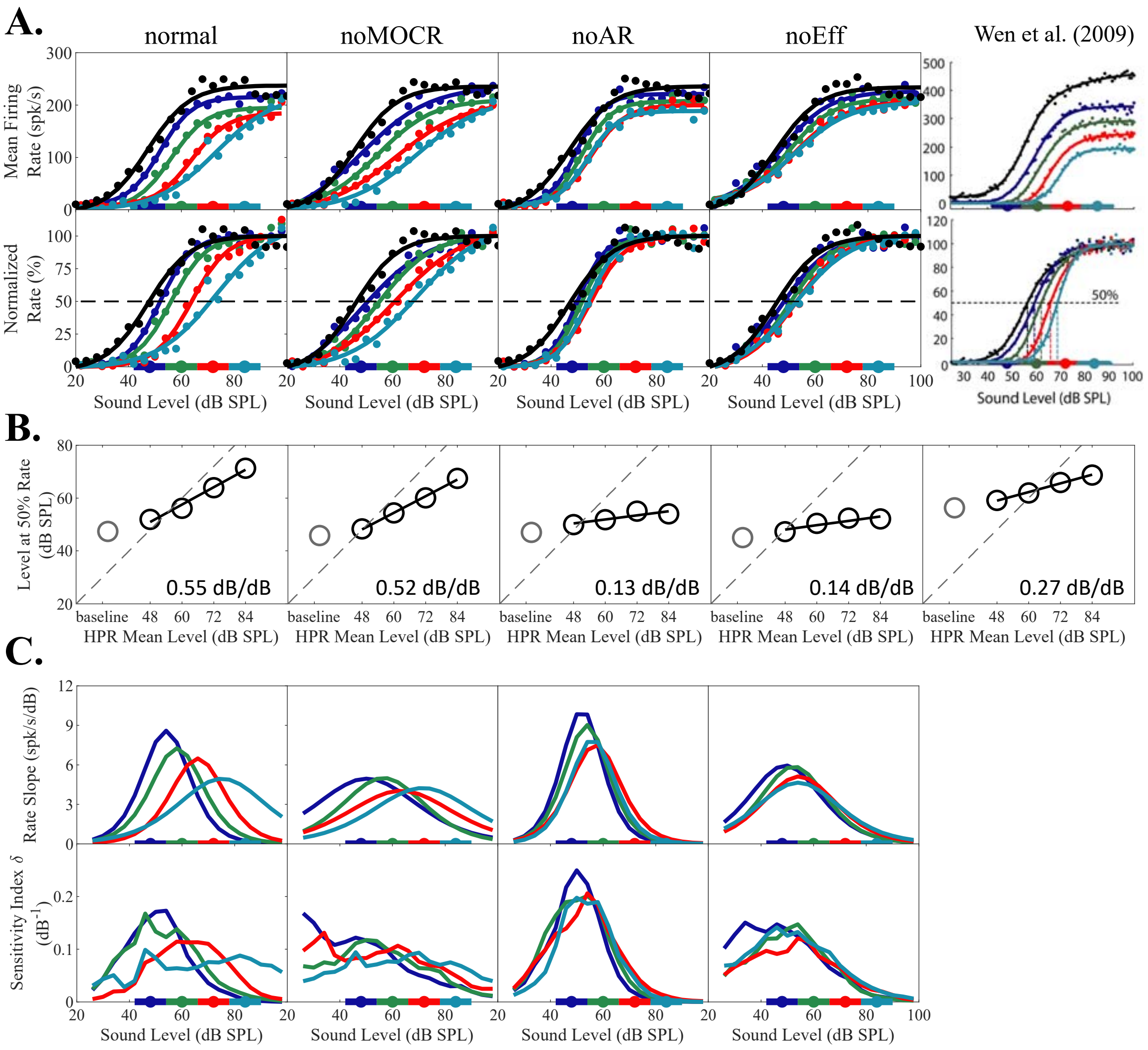

Figure 4. As figure 3, but for a medium SR fiber (BF = 1280 Hz) responding to broadband noise.

Precursor paradigm shown as a more efficient alternative to HPR paradigm

The results of the precursor paradigm are similar to those of the HPR paradigm (Figure 5). With the precursors, the normal RLFs show rightward shift with increasing precursor level, and the amount of such DRA is slightly larger than that of the HPR level, yielding 0.57 dB/dB shift for HSR fibers with 580-Hz BF responding to tones at the BF. The deactivation of efferent reflexes reduces DRA to 0.15 dB/dB.

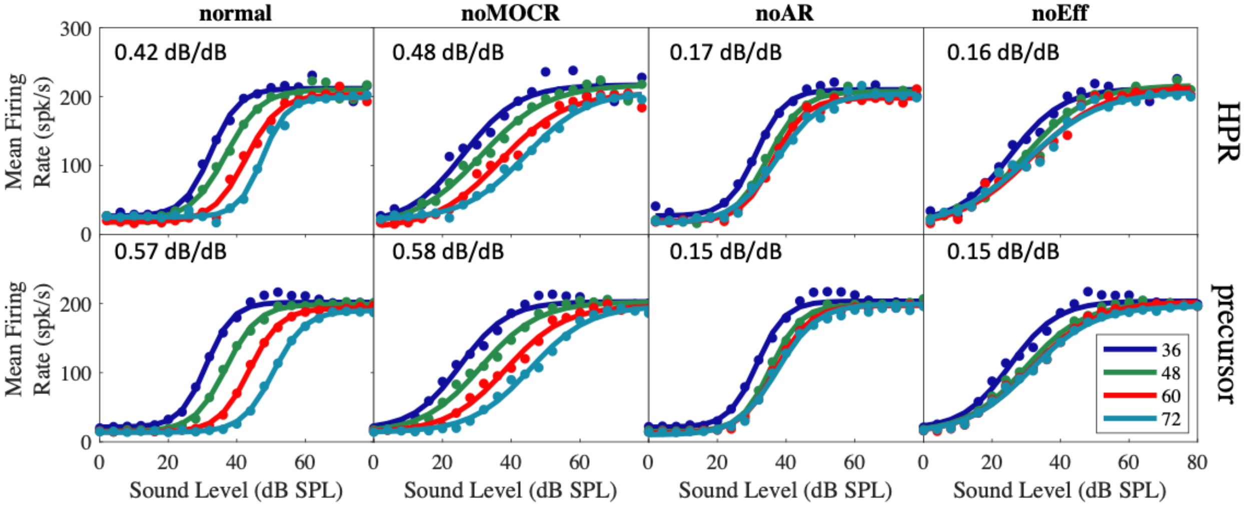

Figure 5. RLFs of high-SR fibers (BF 580 Hz) to 580-Hz tones under different paradigms, hearing conditions and context levels. From left to right panels: modeled human data for normal, noMOCR, noAR and noEff conditions. Top panels: HPR paradigm. Bottom panels: precursor paradigm. Dotted lines: RLFs with context levels in the 36-72 dB range. Solid lines: logistic fits of predicted RLFs. Top left of each panel: DRA (dB/dB).

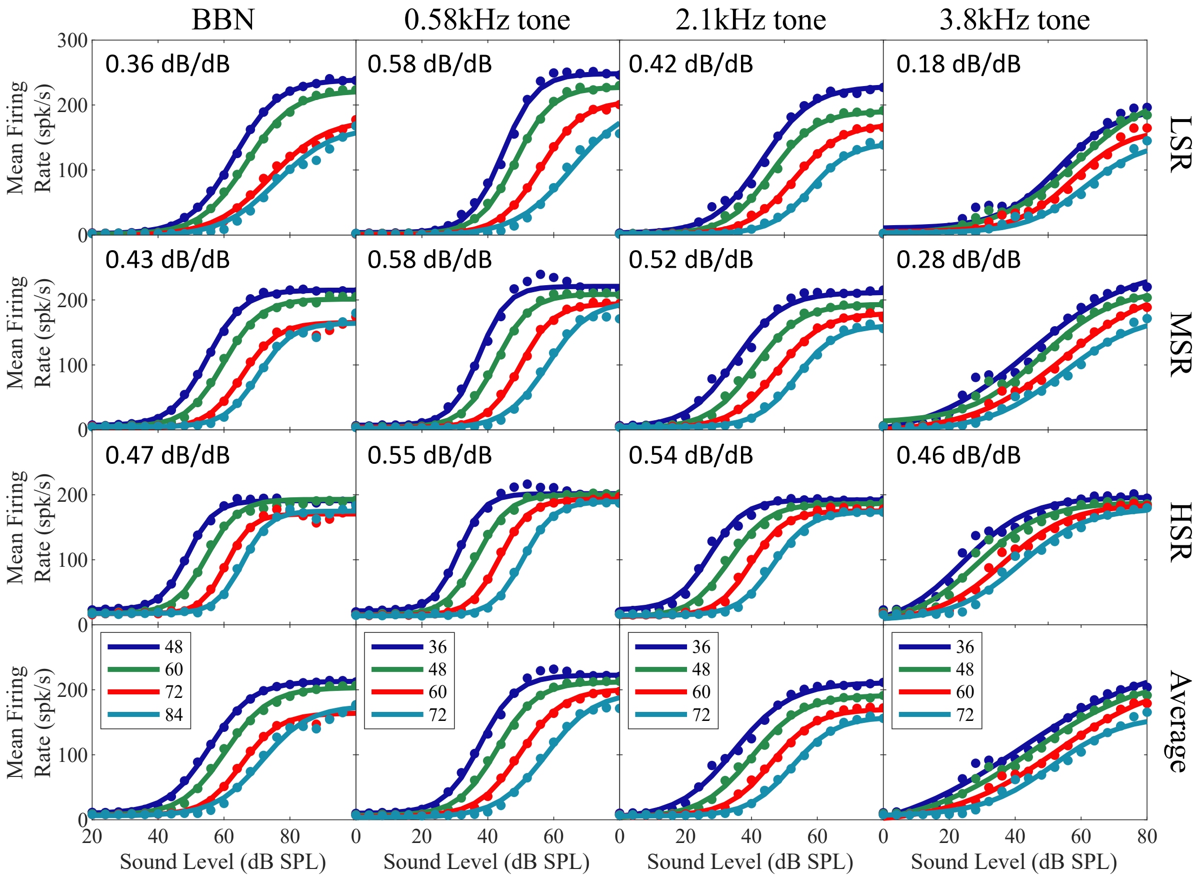

Figure 6. RLFs for model AN fibers of different SRs for broadband noise (BBN) or tones. From left to right panels: BF = 1 kHz, noise stimuli; BF of 580 Hz, 2.1 kHz and 3.8-kHz with matching-frequency tone stimuli. Top to bottom panels: low, medium- and high-SR fibers, and average responses of fibers of the three spontaneous rates.

Analysis of normal fibers responding to various types of stimuli (noise and tones of different frequencies) was performed for each spontaneous rate class using the precursor paradigm. The results (Figure 6) show that (1) RLFs tend to saturate at lower probe levels for fibers with high SR than for fibers with low SR regardless of the stimulus frequencies, but robust DRA occurs for fibers of all three SRs, (2) the amount of DRA decreases with increasing tone frequency, especially for low-SR fibers, and (3) the amount of DRA increases with fibers' SRs for high-frequency tones, but the effect is not obvious for low-frequency tones.

Efferent reflexes in the efficient encoding of speech

Figure 7 shows the SRTs (signal-to-noise ratio for 50% digits correctly reported) achieved by listeners attending to the MAPsim output. Intelligible speech was thus heard using each simulation, but SRTs were improved by including certain features in the simulation.

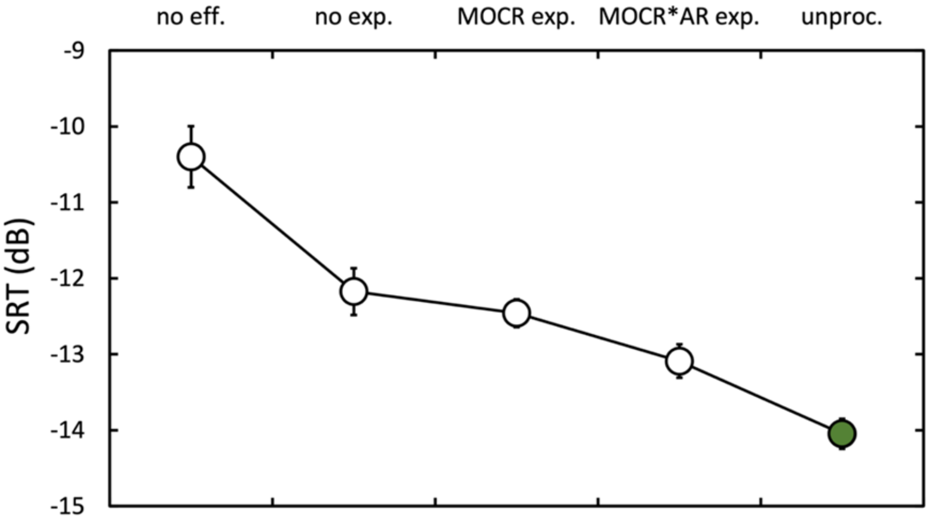

*Figure 7. Digit-triplet SRTs obtained from listeners attending to the original signal (unproc.) and MAPsim outputs with efferent reflexes disabled (no eff.), with them enabled but without expansion (no exp.) and with expansion based on inverted efferent signals (MOCR exp. and MOCR*AR exp.). Error bars are standard errors of means.*

		Mean Difference	SE	*p*
no eff.	no exp.	1.773*	0.410	0.012
	MOCR exp.	2.059**	0.452	0.008
	MOCR*AR exp.	2.689**	0.299	0.001
	unproc.	3.647**	0.509	0.001
no exp.	MOCR exp.	0.286	0.265	1.000
	MOCR*AR exp.	0.916	0.356	0.260
	unproc.	1.873**	0.376	0.004
MOCR exp.	MOCR*AR exp.	0.630	0.308	0.657
	unproc.	1.587**	0.222	0.001
MOCR*AR exp.	unproc.	0.958	0.343	0.176

*Table 3. Post hoc pairwise comparisons between MAPsim processing conditions. Adjustment for multiple comparisons: Bonferroni. *: $p < 0.05$. **: $p < 0.01$.*

The importance of compensating for the peripheral compression introduced by the MAP model was evaluated. The SRTs of ‘no exp.’ (neither expansion applied), ‘MOCR exp.’ (MOCR expansion only), ‘MOCR*AR exp.’ (both expansions applied), and ‘unproc.’ (original, unprocessed stimuli) were compared (Table 3). The mean thresholds were progressively reduced by adding compensation for the MOCR and then the MOCR and AR, with deficit compared to the unprocessed case reaching less than one dB. However, they did not improve significantly over the ‘no exp.’ case.

The role of efferent reflexes in coding speech in noise was examined by comparing 'no eff.' to 'MOCR*AR exp.' and 'unproc.' conditions. Under 'no eff.', efferent reflexes were deactivated in the MAP model, hence no expansion was applied. The results show that when efferent reflexes are absent, the SRT increases significantly, elevating nearly 2.7 dB from that of 'MOCR*AR exp.', $p < 0.001$, and 3.6 dB from that of 'unproc.', $p < 0.001$.

DISCUSSION

The modeling based on the MAP model (Panda et al., 2014), shows how DRA may occur at sound levels up to at least 72 dB, such that the system can remain mostly saturation-free and efficiently transmit to the brain information about temporal modulations of speech uttered at normal levels. Specifically, DRA is brought about by two efferent feedback loops: the acoustic reflex shifts RLFs with context level, by attenuating transmission through the middle ear; the MOCR works in parallel with the acoustic reflex by modulating the electromotility of the outer hair cells, fine-tuning the slope of the RLFs to ensure optimal and precise encoding of sound intensity. Compared to the Wen et al. data, the MAP model predicts a greater effect of DRA but much smaller classical adaptation effects. Greater DRA results from the inclusion of the two efferent processes, which were suppressed by anesthesia in the physiological work. Reduced classical adaptation may come from the use of much shorter HPR stimuli (8 seconds, compared to 5 minutes) in our study, combined with a model that, in any case, only simulates short-term adaptation.

After decoding the firing patterns predicted by MAP back into an acoustic signal, speech recognition in noise through MAPsim significantly improves with activated efferent reflexes, illustrating the role of efferent reflexes in efficient coding of speech, which is a signal highly modulated in spectral and temporal domains (Drullman et al., 1994).

Mechanisms of dynamic range adaptation in AN fibers

The shifting of RLFs toward higher levels as context level increases was first shown in animal studies at the auditory nerve (Wen et al., 2009) and the inferior colliculus (Dean et al., 2005, 2008) levels. Many adaptive properties of the AN are associated with the synapses between inner hair cells and fibers (Moser and Beutner, 2000; Goutman and Glowatzki, 2007), inspiring auditory modeling scientists to simulate DRA through changing the dynamics of inner-hair-cell-auditory-nerve synapses. Zilany & Carney (2010) have successfully simulated DRA by implementing power-law dynamics at the junction between inner hair cells and fibers in their auditory model. However, it is unclear whether these power-law dynamics are physiologically plausible. The current study suggests that DRA at the AN could originate from the efferent reflexes, especially the AR, which would not be evident from studies with anesthetized small mammals. Interestingly, anesthetized animals still show DRA at higher centers (Dean et al. 2005, 2008), suggesting that other mechanisms are also at work at these levels of the nervous system.

The MAP model predicts that the MOCR and the acoustic reflex take on different roles in DRA. The modeled MOCR receives contributions from the AN fibers of all three spontaneous rates. When disabling the MOCR, the slope of the RLF decreases, suggesting that the auditory system becomes less sensitive to sound intensity change. In other words, a slight change in sound intensity does not induce as much difference in the firing rates in the absence of MOCR as in the normal condition. On the other hand, the acoustic reflex is activated only at high intensities to attenuate the stapes displacement, and the amount of attenuation solely depends on the output from the stream that involves the low-SR fibers. When the acoustic reflex is disabled, the firing rates at high probe levels are no longer suppressed, causing the RLFs of the higher context levels to shift leftward and overlap with the RLFs of the lower context levels. Therefore, the absence of acoustic reflex impacts the sensitivity and accuracy of intensity coding at higher context levels and the auditory system's ability to perform DRA efficiently.

The efferent reflexes have been suggested as a source of DRA but their role could not be examined in small mammals because anesthesia in physiological studies at least partially suppresses the efferent system. Note that, in the current study, when efferent reflexes are activated, the amount of DRA far exceeds what has been found in physiological studies, suggesting that the contribution of efferent reflexes to DRA was obscured under anesthesia but can be revealed using computational modeling.

Some DRA remained in both modeled and empirical data, even with both efferent reflexes disabled, suggesting an additional source of adaptation in the peripheral auditory system. The most plausible explanation for this remaining adaptation resides in the dynamics of neurotransmitter vesicle release into the cleft, replenishment within the inner hair cell and reuptake by the hair cell from the cleft, as emulated by the 3-store model (Meddis, 1986, in its probabilistic implementation, and Sumner et al., 2002, in the quantized implementation used in this study). While such depletion accounts for some firing-rate adaptation, the presence of DRA with deactivated efferent reflexes shows that non-efferent-related DRA is an emergent property of the 3-store hair cell model.

Classical adaptation in AN fibers

Figure 6 shows some evidence of classical adaptation, but mainly in the low SR fibers and much less than seen throughout the Wen et al. data. Firing-rate adaptation occurs on different timescales. Short timescales (a few milliseconds or tens of milliseconds) are expressed in the 3-store model via fast available-store depletion but long-term firing-rate adaptation (Kiang et al., 1965) is not. Long-term adaptation may stem from a gradual decrease, under steady stimulation, of the ion flux (Strimbu et al., 2019) required by inner hair cells to drive neurotransmitter release into the cleft. It is not captured in the MAP model and therefore not in our predictions. The HPR paradigm used in Wen et al. (2009) may capture such adaptation in high SR fibers because the stimulus is minutes in duration (see Figures 3 and 4).

The model predicts differences in short-term adaptation as a function of spontaneous rate (Figure 6) due to differences in the time constant τ_{Ca}, which reflects the dwell time of presynaptic calcium in the vicinity of the synapse and therefore determines the release characteristics of the synapse. At saturation, despite high depletion of the available store, the probability of release of vesicles is much higher in high-SR than in low SR-fibers, such that short-term firing-rate adaptation of high-SR fibers is limited in the HPR or precursor paradigms.

The precursor paradigm

The precursor paradigm significantly improves the efficiency of setting up the context level compared to the HPR paradigm. The precursor precedes the probe with an identical signal that is 400-ms long and sets the context level. The precursor allows sufficient time to activate the efferent system to produce a given level of DRA. The precursor paradigm performed equivalently to the HPR paradigm in revealing DRA. Importantly, signals could be processed much more efficiently under the precursor paradigm so that the roles of efferent reflexes could be studied with perceptual measures using MAPsim. The equivalence of the HPR and precursor paradigms is reassuring given that psychophysical studies generally employ the latter when attempting to activate the efferent system.

Future use of the MAPsim simulator

MAPsim provides a new simulation framework for efficiently exploring peripheral auditory physiology, its pathologies and the corresponding perceptual impacts. Since all hearing depends upon the signal encoded on the AN, the decoded sound will reflect any loss of information occurring within the model of peripheral transduction and thus the effects of modeled pathologies. MAPsim proved successful in that SRTs at the simulator validation stage differed from those obtained with unprocessed stimuli by just 1 dB, suggesting very limited information loss when simulating normal hearing.

MAPsim could serve as a powerful tool to simulate perceptual effects of specific hearing pathologies, such as loss of inner vs. outer hair cells, loss of endocochlear potential and synaptopathy. The present simulations enable us to see via psychophysical measures that a deficient caudal efferent system could cause unrecoverable information loss and severely impair the ability to recognize speech in steady-state noise. Previously, the role of the efferent system, especially the MOCR, on speech recognition in noise was only studied through coupling the MAP model with an artificial observer, such as an automatic speech recognition system (Clark et al., 2012; Yasin et al., 2020), or through correlational studies where speech performance was examined under different levels of efferent activation (Mertes et al., 2018). Here, the simulator indicates specific effects of both MOCR and AR on human speech reception thresholds.

Conclusion

Our findings confirm the potential of efferent reflexes to maintain DRA and enable efficient coding of speech at the auditory nerve level. The MAP model predicts that the acoustic reflex shifts the dynamic range of auditory-nerve fibers towards contextual levels and the MOC reflex increases fiber sensitivity around that level. Our MAPsim simulator was validated for normal hearing of speech stimuli. Being based on MAP, MAPsim can be used to simulate specific sensorineural pathologies, opening the door to establishing their psychophysical signatures, such that they may be differentially diagnosed.

REFERENCES

Chimento TC, Schreiner CE (1991) Adaptation and recovery from adaptation in single fiber responses of the cat auditory nerve. J Acoust Soc Am 90:263–273.

Clark NR, Brown GJ, Juergens T, Meddis R (2012) A frequency-selective feedback model of auditory efferent suppression and its implications for the recognition of speech in noise. J Acoust Soc Am 132:1535–1541.

Colburn HS, Carney LH, Heinz MG (2003) Quantifying the information in auditory-nerve responses for level discrimination. JARO - J Assoc Res Otolaryngol 4:294–311.

Dean I, Harper NS, McAlpine D (2005) Neural population coding of sound level adapts to stimulus statistics. Nat Neurosci 8:1684–1689.

Dean I, Robinson BL, Harper NS, McAlpine D (2008) Rapid neural adaptation to sound level statistics. J Neurosci 28:6430–6438.

Drullman R, Festen J, Plomp R (1994) Effect of temporal envelope smearing on speech reception. J Acoust Soc Am 95:1053–1064.

Goutman JD, Glowatzki E (2007) Time course and calcium dependence of transmitter release at a single ribbon synapse. Proc Natl Acad Sci U S A 104:16341–16346.

Guinan JJ, Gifford ML (1988) Effects of electrical stimulation of efferent olivocochlear neurons on cat auditory-nerve fibers. I. Rate-level functions. Hear Res 33:97–113.

Harris DM, Dallos P (1979) Forward masking of auditory nerve fiber responses. J Neurophysiol 42:1083–1107.

Hung IJ, Dallos P (1972) Study of the acoustic reflex in human beings. I. Dynamic characteristics. J Acoust Soc Am 52:1168–1180.

Kiang NY, Watanabe T, Thomas EC, Clark LF (1965) Discharge patterns of single fibers in the cat’s auditory nerve. MIT Press.

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

Lopez-Poveda EA, Meddis R (2001) A human nonlinear cochlear filterbank. J Acoust Soc Am 110:3107–3118.

MacGregor RJ (1987) Neural and Brain Modeling. Academic Press.

Meddis R (1986) Simulation of mechanical to neural transduction in the auditory receptor. J Acoust Soc Am 79:702–711.

Meddis R, Lecluyse W, Clark NR, Jürgens T, Tan CM, Panda MR, Brown GJ (2013) A computer model of the auditory periphery and its application to the study of hearing. In: Advances in Experimental Medicine and Biology, pp 11–20.

Mertes IB, Wilbanks EC, Leek MR (2018) Olivocochlear Efferent Activity is Associated With the Slope of the Psychometric Function of Speech Recognition in Noise. Ear Hear 39:583–593.

Moser T, Beutner D (2000) Kinetics of exocytosis and endocytosis at the cochlear inner hair cell afferent synapse of the mouse. PNAS 97:883–888.

Panda MR, Lecluyse W, Tan CM, Jürgens T, Meddis R (2014) Hearing dummies: Individualized computer models of hearing impairment. Int J Audiol 53:699–709.

Smith RL (1979) Adaptation, saturation, and physiological masking in single auditory-nerve fibers. J Acoust Soc Am 65:166–178.

Smith RL, Zwislocki JJ (1975) Short-term adaptation and incremental responses of single auditory-nerve fibers. Biol Cybern 17:169–182.

Strickland EA (2008) The relationship between precursor level and the temporal effect. J Acoust

Soc Am 123:946–954.

Strimbu CE, Prasad S, Hakizimana P, Fridberger A (2019) Control of hearing sensitivity by tectorial membrane calcium. Proc Natl Acad Sci U S A 116:5756–5764.

Sumner CJ, Lopez-Poveda EA, O'Mard LP, Meddis R (2002) A revised model of the inner-hair cell and auditory-nerve complex. J Acoust Soc Am 111:2178.

Watkins P V., Barbour DL (2008) Specialized neuronal adaptation for preserving input sensitivity. Nat Neurosci 11:1259–1261.

Wen B, Wang GI, Dean I, Delgutte B (2009) Dynamic range adaptation to sound level statistics in the auditory nerve. J Neurosci 29:13797–13808.

Winter IM, Robertson D, Yates GK (1990) Diversity of characteristic frequency rate-intensity functions in guinea pig auditory nerve fibres. Hear Res 45:191–202.

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.

Young ED, Sachs MB (1979) Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. J Acoust Soc Am 66:1381–1403.

Zilany MSA, Carney LH (2010) Power-Law Dynamics in an Auditory-Nerve Model Can Account for Neural Adaptation to Sound-Level Statistics. J Neurosci 30:10380–10390.