Hearing Research 426 (2022) 108621

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Modeling temporal information encoding by the population of fibers in the healthy and synaptopathic auditory nerve

Peter T. Johannesen [a,b], Thibaud Leclère [a,b], Aswin Wijetillake [d], Manuel Segovia-Martínez [e], Enrique A. Lopez-Poveda [a,b,c,*]

[a] Instituto de Neurociencias de Castilla y León, Universidad de Salamanca, Salamanca, Spain
[b] Instituto de Investigación Biomédica de Salamanca, Universidad de Salamanca, Salamanca, Spain
[c] Departamento de Cirugía, Facultad de Medicina, Universidad de Salamanca, Spain
[d] Oticon Medical, Smørum, Denmark
[e] Oticon Medical, Vallauris, France

ARTICLE INFO

Article history:
Received 16 February 2022
Revised 16 September 2022
Accepted 20 September 2022
Available online 21 September 2022

Keywords:
Hearing
Deafness
Deafferentation
Temporal processing
Envelope
Amplitude modulation
Temporal fine structure

ABSTRACT

We report a theoretical study aimed at investigating the impact of cochlear synapse loss (synaptopathy) on the encoding of the envelope (ENV) and temporal fine structure (TFS) of sounds by the population of auditory nerve fibers. A computational model was used to simulate auditory-nerve spike trains evoked by sinusoidally amplitude-modulated (AM) tones at 10 Hz with various carrier frequencies and levels. The model included 16 cochlear channels with characteristic frequencies (CFs) from 250 Hz to 8 kHz. Each channel was innervated by 3, 4 and 10 fibers with low (LSR), medium (MSR), and high spontaneous rates (HSR), respectively. For each channel, spike trains were collapsed into three separate 'population' post-stimulus time histograms (PSTHs), one per fiber type. Information theory was applied to reconstruct the stimulus waveform, ENV, and TFS from one or more PSTHs in a mathematically optimal way. The quality of the reconstruction was regarded as an estimate of the information present in the used PSTHs. Various synaptopathy scenarios were simulated by removing fibers of specific types and/or cochlear regions before stimulus reconstruction. We found that the TFS was predominantly encoded by HSR fibers at all stimulus carrier frequencies and levels. The encoding of the ENV was more complex. At lower levels, the ENV was predominantly encoded by HSR fibers with CFs near the stimulus carrier frequency. At higher levels, the ENV was equally well or better encoded by HSR fibers with CFs different from the AM carrier frequency as by LSR fibers with CFs at the carrier frequency. Altogether, findings suggest that a healthy population of HSR fibers (i.e., including fibers with CFs around and remote from the AM carrier frequency) might be sufficient to encode the ENV and TFS over a wide range of stimulus levels. Findings are discussed regarding their relevance for diagnosing synaptopathy using non-invasive ENV- and TFS-based measures.

1. Introduction

The mammalian cochlea separates the frequency components of sounds such that low and high frequencies cause mechanical vibrations in apical and basal cochlear regions, respectively. For complex sounds, such as speech, the vibration of each cochlear site consists of slow amplitude fluctuations (referred to as the envelope or ENV) superimposed on faster amplitude fluctuations (termed the temporal fine structure or TFS) whose instantaneous frequency may vary over time around the characteristic frequency (CF) of the cochlear site in question. The ENV and TFS are responsible for several aspects of auditory perception, including speech recognition, pitch perception or sound source localization (e.g., Lorenzi et al., 2006; Rosen, 1992; Shannon et al., 1995; Smith et al., 2002; Van Tasell et al., 1987). Therefore, normal auditory perception requires both ENV and TFS to be accurately encoded in the auditory nerve (AN).

Envelope and TFS coding in the AN has been extensively investigated. Most animal studies have analyzed the response of individual AN fibers to sinusoidally amplitude-modulated (AM) pure-tones with carrier frequencies that match the CF of the fiber (e.g., Johnson, 1980; Joris and Yin, 1992; Kale and Heinz, 2012). The

* Corresponding author: Enrique A. Lopez-Poveda, Instituto de Neurociencias de Castilla y León, Universidad de Salamanca, Calle Pintor Fernando Gallego 1, 37007 Salamanca, Spain.
E-mail address: ealopezpoveda@usal.es (E.A. Lopez-Poveda).

https://doi.org/10.1016/j.heares.2022.108621

use of AM tones is convenient because they allow direct control over the ENV and TFS (i.e., the TFS is the AM carrier tone while the ENV is the modulating signal). However, limiting the examination to just the responses from individual fibers with the same CF as the AM carrier frequency may not provide the full picture of ENV and TFS coding. The AN contains thousands of fibers with different CFs and dynamic ranges and it is possible that encoding limitations of individual fibers are compensated for by the myriad fibers (e.g., Wever, 1949; Lopez-Poveda and Barrios, 2013; Lopez-Poveda, 2014). For example, fibers with high-spontaneous rates (HSRs) have lower thresholds and narrower dynamic ranges, while fibers with medium- (MSR) and low-spontaneous rates (LSRs) have higher thresholds and wider dynamic ranges (e.g., Liberman et al., 1978). It is possible that HSR fibers with CFs different from the carrier frequency and/or LSR fibers with CFs at the carrier frequency encode the ENV in their discharge rate at stimulus levels at which HSR fibers with CFs at the carrier frequency cannot because they discharge at saturation (Carney, 2018; Encina-Llamas et al., 2019; Kohlrausch et al., 2000). Because of this, rather than assessing the encoding capacity of individual fibers, it would be more appropriate and enlightening to investigate the encoding capacity of the whole AN.

Assessing ENV and TFS encoding by the population of AN fibers would be particularly important to understand the impact of fiber loss and/or synapse degradation (synaptopathy) on auditory perception. It has been suggested that synaptopathy affects predominantly LSR and MSR fibers (Furman et al., 2013; Schmiedt et al., 1996), although recent studies have questioned this notion (Suthakar and Liberman, 2021). In rodents, synaptopathy can occur with noise exposure (Kujawa and Liberman, 2009) and aging (Sergeyenko et al., 2013), even when audiometric thresholds remain clinically normal. In humans, synaptopathy occurs with aging (Makary et al., 2011; Viana et al., 2015; Wu et al., 2019) but the influence of noise exposure remains controversial (e.g., Bramhall et al., 2019; Johannesen et al., 2019). In theory, synaptopathy could degrade the neural encoding of the ENV and/or the TFS, which might contribute to the hearing difficulties experienced by older people with clinically normal-hearing thresholds (Lopez-Poveda and Barrios, 2013; Lopez-Poveda, 2014; Plack et al., 2014). Experimental support to these notions, however, remains elusive (e.g., Johannesen et al., 2019; Carcagno and Plack, 2021). Understanding to what extent synaptopathy degrades the encoding of the ENV and TFS could help elucidating these issues and might also guide the design of non-invasive methods to diagnose synaptopathy in humans.

Here, we investigate the effect of synaptopathy on ENV and TFS encoding by the population of AN fibers. Addressing this aim experimentally would be practically infeasible as it would require recording simultaneous responses (spike trains) from multiple AN fibers with different SRs and CFs to a single sound instance. For this reason, we instead use a computational model of the AN response. Furthermore, we apply information theory to reconstruct (i.e., 'decode') the ENV and TFS from the simulated spike trains in a mathematically optimal manner. The quality of the reconstruction is used as a metric of the ENV and TFS information encoded in the nerve. Various synaptopathy scenarios are simulated by removing fibers of specific types and/or from specific cochlear regions before reconstruction.

2. Materials and methods

2.1. Experimental design

Our approach is depicted in schematic form in Fig. 1. A computational model of the auditory periphery was used to simulate the timing at which spikes occur in the AN (spike trains) in response to sound (Fig. 1A). The model simulated the nonlinear vibration of the basilar membrane (BM) at multiple cochlear sites, or CFs. The simulated BM vibration at each site served as the input to a biophysical model of the inner hair cell (IHC). Each IHC was innervated by 3, 4 and 10 simulated AN fibers with LSR, MSR and HSR, respectively. For each cochlear site, the simulated spike trains were collapsed into three separate post-stimulus time histograms (PSTHs), one per fiber type (Fig. 1B). Information theory was applied to design a filter (Fig. 1C) that served to optimally reconstruct a reference signal (Fig. 1D) from one or more PSTHs. The reconstructed estimate was compared to the reference signal (Fig. 1E) and the quality of the reconstruction was regarded as a

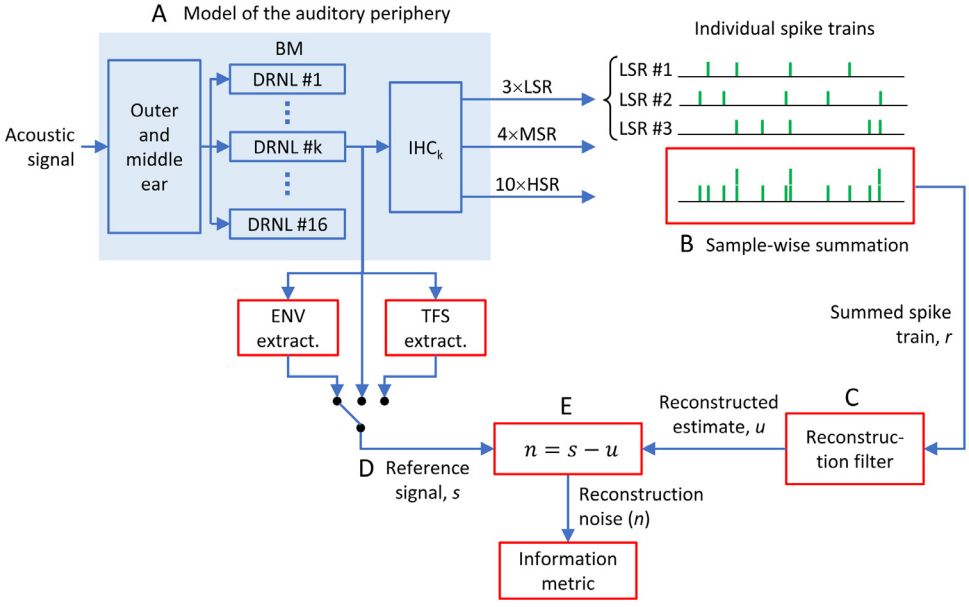

Fig. 1. Schematic representation of our approach. **A.** The auditory model with its various stages. **B.** Summation of spike trains. **C.** The reconstruction filter. **D.** The reference signal, i.e., the signal to be reconstructed from the spike train(s). In this study, three reference signals were considered: the waveform, the ENV and the TFS of the simulated BM response at the cochlear site whose CF matched the frequency of the AM tone carrier. **E.** The reconstruction noise used to calculate the information metric.

metric of the information about the reference signal present in the used PSTHs.

Stimuli were sinusoidally AM pure tones with various carrier frequencies (f_cs) and levels. A single AM tone was presented at a time. The modulation frequency was always 10 Hz. This value was chosen because the frequency side bands would not be resolved and all of the information about the ENV would be in the temporal fluctuations of AN discharges rather than in the rate profile (e.g., Moore and Glasberg, 2001). The modulation depth was always 100%. To minimize transient effects that may affect the information metric, stimuli included silence before and after the AM tone for a duration equal to that of the reconstruction filter plus 5 ms.

The approach was applied to quantify the information about the modulating signal (ENV), the carrier tone (TFS), and the full waveform (i.e., combined ENV and TFS) present in the AN. Reconstruction filters were calculated using all available PSTHs or a subset of them, depending on the question of interest (Fig. 1 illustrates a hypothetical example case where the ENV of the simulated BM response for channel #k was reconstructed using only the PSTH for LSR fibers innervating that channel). Synaptopathy was simulated by removing spike trains or entire PSTHs for specific fiber types and/or cochlear regions before reconstruction.

We addressed the following questions:

(1) Which fiber type(s) contributes the most to ENV and TFS encoding across a range of levels?
(2) To what extent is the ENV and TFS encoded by fibers with CFs equal to and different from the frequency of the AM tone carrier?
(3) Is there an appropriate stimulus to diagnose LSR fiber loss?

2.2. Model of the auditory periphery

There exist various computational models (and versions of them) of the AN response to arbitrary acoustic stimulation (reviewed by Lopez-Poveda, 2005; Meddis and Lopez-Poveda, 2010; Osses Vecchi et al., 2022). Here, we used the computational model of the guinea pig auditory periphery of Sumner et al. (2003a). This model accurately reproduces a wide range of phenomena and characteristics of experimental AN responses including rate-level functions to on- and off-CF stimuli for different fiber types (Figs. 3 and 4 in Sumner et al., 2002), phase locking and synchronization of AN discharges (Fig. 6 in Sumner et al., 2002), refractoriness, and different discharge-rate adaptation characteristics for different fiber types (Sumner et al., 2003b). It also reproduces the temporal representation of speech-like stimuli in the AN (Holmes et al., 2004). For a review of the merits of this model, see Table I in Lopez-Poveda (2005). Only a brief description of the model is given here because full details can be found in the original publications. The limitations of this model are discussed later.

The input to the model is an acoustic pressure waveform and its output is a collection of spike trains for several AN fibers with different CFs and SRs. The model consists of several stages where the output signal from each stage is used as the input to the next stage (Fig. 1A). The first stage models the velocity of stapes motion using a linear filter whose frequency response matches the combined frequency response of the outer and middle ear. The second stage consists of a bank of dual-resonance nonlinear (DRNL) filters (Lopez-Poveda and Meddis, 2001; Meddis et al., 2001). The output from each DRNL filter models the velocity of BM vibration at a cochlear site whose CF matches the filter CF. Here, we used 16 DRNL filters with CFs spanning the frequency range from 250 Hz to 8 kHz. The output from each DRNL filter is used as the input to a biophysical model of the IHC receptor potential (modified from Shamma et al., 1986). The receptor potential is used to drive a model of the synapse between the IHC and the AN (modified from Meddis, 1986). The synapse model has two elements: one to simulate the presynaptic calcium processes that lead to the release of neurotransmitter, and one to simulate the manufacture, release, loss, and reuptake of neurotransmitter vesicles at the synapse (Sumner et al., 2002). Two parameters control the calcium processes and can be adjusted to create synapses for LSR, MSR and HSR fibers. Depletion of neurotransmitter produces adaptation of the firing rate specific to each fiber type (Sumner et al., 2003b) and reproduces well the experimental data of Müller and Robertson (1991). The shapes of the rate-level functions (shown in Fig. 2) are 'saturating' for HSR fibers, with lower thresholds, steeper slopes, and narrower dynamic ranges; 'sloping saturation' for MSR fibers, with low-medium thresholds and typically wider dynamic ranges; and 'straight' for LSR fibers, with higher thresholds, shallower slopes, and wider dynamic ranges (Winter et al., 1990; Sumner et al., 2002, 2003a).

Table 1
Simulated synaptopathy profiles. CF: Characteristic frequency; HSR: high-spontaneous rate; MSR: medium-spontaneous rate; LSR: low-spontaneous rate.

Synaptopathy profile	Fiber count per CF			
	HSR	MSR	LSR	Total
Normal hearing	10	4	3	17
Large loss of MSR/LSR fibers	10	2	1	13
Total loss of LSR fibers	10	1	0	11
Large loss of HSR fibers	4	4	3	11
Total loss of HSR fibers	0	4	3	7

Unless otherwise stated, each IHC was innervated by 17 AN fibers: 10 HSR + 4 MSR + 3 LSR. This number and the distribution of SRs reflect corresponding experimental values in cat (61% HSR, 23% MSR, and 16% LSR, respectively; Liberman, 1978). Synaptopathy was simulated by reducing the number of fibers per IHC (as shown in Table 1). Note that in this approach, fiber loss (deafferentation) is equivalent to synapse loss (synaptopathy).

2.3. Information metric

ENV and TFS encoding in the AN have been classically assessed using vector strength (e.g., Johnson, 1980; Joris and Yin, 1992) or spike correlograms (e.g., Louage et al., 2004; Joris et al., 2006). The former is only appropriate for periodic signals while the latter may be generalized to aperiodic signals. Here, we used an information theory approach with a view to generalize the present analyses in the future to investigate the effect of synaptopathy on the AN encoding of complex stimuli like speech.

Information theory provides tools for estimating the amount of information about an input signal transmitted through a communication channel to the output. In the present case, the input is the reference signal, the AN model is the communication channel and the multiple spike trains are the outputs. The transmitted information (or mutual information) between input and output (in units of bit/sec) may be interpreted as the reduction of uncertainty about what was the reference signal having observed one particular set of spike trains. For example, if 100 bits/sec were transmitted, the number of potential reference signals that evoked the spike train in question would be reduced by a factor 2^{100}. Therefore, a higher bitrate means less uncertainty about the reference signal.

Of the several approaches that exist to estimate the mutual information between the reference signal and the corresponding AN response, we chose one that consists of two steps (Bialek, 1991; see Section 3.2.3 in Rieke, 1997; Warland et al., 1997; Mesgarani et al., 2008; Pasley et al., 2012; Zai et al., 2015). Step #1 involved reconstructing an optimal estimate of the reference sig-

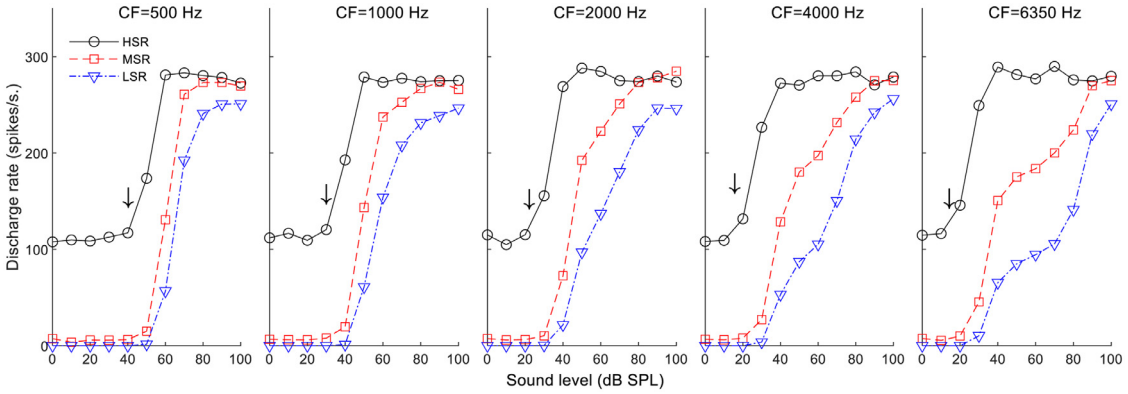

Fig. 2. Model rate-level functions for HSR, MSR and LSR fibers with different CFs (indicated at the top of each panel). Vertical arrows indicate the neural threshold for HSR fibers. Discharge rates were average values calculated from responses to pure tones at the fiber's CF.

nal from one or more spike trains. Step #2 involved assessing the quality of reconstruction. The latter was used as a proxy for the information encoded in the spike trains.

2.3.1. Step #1: Reconstruction of the reference signal

Reconstruction was done using the method of Warland et al. (1997). In this method, one or more spike trains are input to an optimal reconstruction filter that produces an estimate (u) of the reference signal (s). The optimal reconstruction filter is defined as the one that minimizes the root-mean-square (RMS) difference between the reference and reconstructed estimate signals, i.e., the filter that minimizes the reconstruction noise ($n = s - u$, Figs. 1E and 3D). Although the filter's coefficients can be found using an analytic mathematical expression (see the Appendix), we instead used MATLAB's `gmres` function to solve the associated matrix equation to find the optimal reconstruction filter. This procedure is justified because the matrices involved here are large but sparse and hence the iterative solution implemented in `gmres` is computationally convenient. In using `gmres`, the maximum number of inner iterations was set to 10, the maximum total iterations was set to 40, and the error tolerance was set to 10^{-4}. If `gmres` did not converge, a second try was done with the additional use of a preconditioner matrix (incomplete Cholesky factorization with threshold dropping). We set the reconstruction filter duration equal to 100 ms.

The reconstruction filter is designed to optimally reconstruct a reference signal from one or more spike trains. In theory, the quality of the reconstruction should improve by using a greater number of spike trains, a result that would be indicative that more spike trains convey more information about the reference signal than fewer spike trains. On the other hand, multiple spike trains evoked by a single instance of the stimulus are somewhat correlated with each other and the computational load increases with increasing the number of spike trains used for reconstruction. Here, to reduce the computational load as well as the risk for artefactual reconstruction (see the Results), we sample-wise summed the spike trains for all fibers of the same type innervating a given IHC (Figs. 1B and 3B). This approach resembles that of Encina-Llamas et al. (2019). As a result, reconstruction could be done using up to 48 summed spike trains (3 fiber types $\times$ 16 CF channels), although it was often done with only a subset of them. The summed spike trains may be thought of as PSTHs, although they are the result of summing spike trains from multiple fibers to a single instance of the stimulus, rather than summing spike trains from a single fiber evoked by multiple stimulus instances.

2.3.2. Step #2: Reconstruction quality and information metric

The reconstructed estimate (u, Fig. 1C) was compared with the reference signal (s, Fig. 1D) to assess the quality of the reconstruction and thus the amount of information from the reference signal present in the summed spike trains. A lower limit of the information encoded in the spike trains in units of bit/sec can be found by comparing the reference signal with the reconstruction noise (Bialek, 1991; see Section 3.2.3 in Rieke, 1997). An alternative metric, called the coding fraction (Wessel et al., 1996; Gabbiani and Metzner, 1999), however, is a simpler metric of the fidelity of the representation of the stimulus in the spike trains and is also a measure of (or a proxy for) the transmitted information. The coding fraction, γ, is defined as the proportion of the

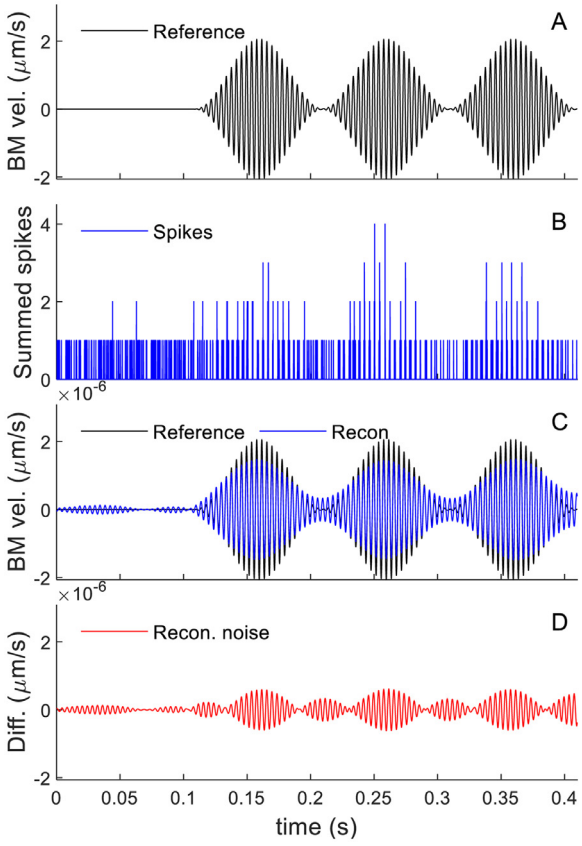

Fig. 3. An example of the reconstruction process. The stimulus was an AM tone with a carrier frequency of 1025 Hz. **A**. The reference signal, i.e., the output signal from the DRNL filter whose CF matched the carrier frequency. **B**. Summed spike train for the 10 HSR fibers innervating the relevant channel. **C**. Reconstructed estimate overlaid with the reference signal. **D**. Reconstruction noise, calculated as the difference between the reference signal and the reconstructed estimate.

standard deviation (STD) of the reference signal accounted for by the reconstructed estimate (Wessel et al., 1996; Gabbiani and Metzner, 1999):

$$\gamma = 1 - \frac{STD(Reconstruction\ noise)}{STD(Reference\ signal)} \quad (1)$$

If all information is transmitted, the reconstructed estimate is identical to the reference signal, the reconstruction noise is zero and the coding fraction equals one. Conversely, if there is zero information transmission, the reconstruction noise and the reference signal will have similar amplitude variance and the coding fraction will be zero.

We noted that the coding fraction is similar to the time-domain signal-to-noise ratio (SNR), where the only difference is the removal of the mean signal amplitude (i.e., the direct-current, DC, component) in the formulae for standard deviation. Because the SNR is simpler and more intuitive, we decided to remove any DC component from all signals, and use the time-domain reconstruction SNR (in decibels) as the metric for reconstruction quality and information. The reconstruction SNR is defined as follows:

$$Reconstruction\ SNR = 20 \cdot log_{10}\left(\frac{RMS[Reference\ signal - M(Reference\ signal)]}{RMS[Reconstruction\ noise - M(Reconstruction\ noise)]}\right) \quad (2)$$

where *RMS* and *M* are the root-mean-square and mean operators, respectively. Note that the higher the reconstruction SNR, the greater the quality of the reconstructed estimate and thus the greater the amount of information about the reference signal conveyed in the summed spike trains.

2.4. The reference signals

The choice of the reference signal is important and generally depends on the aim of the investigation. The tonotopic organization of the auditory system constitutes a 'place code' for sound frequency. For broadband stimuli, an accurate information metric should preserve this place information. The stimuli used here (pure tones modulated at 10 Hz) were narrowband. Nonetheless, to emphasize the appropriateness of preserving the place information in more general applications of our approach for complex stimuli, we chose to reconstruct the simulated BM responses rather than the acoustic stimulus (Fig. 1). Specifically, we chose to reconstruct the output signal from the DRNL filter whose CF matched the carrier frequency of the AM tone. Given the very low AM rate, this signal was like the acoustic stimulus except that it may have included nonlinear cochlear effects (e.g., compression, distortion, etc.) simulated by the DRNL filters (Meddis et al., 2001; Lopez-Poveda, 2003).

To separately quantify the encoded information about the modulating signal (ENV), the carrier tone (TFS), and the full waveform, we first extracted the ENV and TFS (Fig. 1) from the simulated BM response waveform and used either of them or the full waveform, as appropriate, as the reference signal (Fig. 1E). The ENV was calculated as the modulus of the Hilbert transform of the DRNL filter output waveform. The TFS was calculated as the cosine of the unwrapped phase of the Hilbert transform. The TFS was calculated only when the instantaneous amplitude of DRNL filter output waveform (i.e. the BM velocity in m/s) was above the neural threshold for each CF channel in the model; otherwise, the amplitude of the TFS was set to an arbitrary low value (10^{-20}). This prevented low-level signals from producing TFS without perceptual relevance. Neural threshold (indicated by arrows in Fig. 2) was defined as the pure tone level where HSR fibers discharged at a rate equal to the spontaneous rate plus 5% of their dynamic range, where the dynamic range was the difference between the maximum and the spontaneous discharge rate of the fiber.

Depending on the question of interest, the reconstruction of the ENV, TFS or full waveform was done using (1) summed spike trains from fibers that had the same CF as the DRNL filter (on-frequency reconstruction, as illustrated in Fig. 1); (2) summed spike trains from fibers whose CF was different from that of the DRNL filter (off-frequency reconstruction); or (3) summed spike trains from all fibers. In all three cases, summed spike trains were separately available for each fiber type and CF. Because the reconstruction filter was optimal (by calculation) for each combination of reference signal and set of summed of spike trains, the filter was different for each reconstruction.

2.5. Implementation

The model and analyses were implemented in MATLAB® (ver. 2019a). Processing was done with a sampling frequency of 20 kHz except for the IHC synapse model, which was run with a sampling frequency of 100 kHz (after Sumner et al., 2003a). To ensure the same sampling frequency for all signals, the output spike trains from the model were binned every 0.05 ms (1/20000) without discarding any spike.

3. Results

3.1. Reconstruction

The first step to quantify the encoded information was to optimally reconstruct a chosen reference signal from a set of summed spike trains. Fig. 3 illustrates an example of the process when the reference signal is the full waveform. The figure illustrates the reference signal, i.e., the output waveform from the DRNL filter-bank channel whose CF matched the AM carrier frequency (1025 Hz in this case) (Fig. 3A); the summed spike train obtained by summing, for each time instant, the spikes from the 10 HSR fibers innervating the channel's IHC (Fig. 3B); the reconstructed estimate overlaid with the reference signal (Fig. 3C); and the reconstruction noise calculated as difference time waveform between the reference signal and the reconstructed estimate (Fig. 3D). The figure shows only the first 400 ms of the signals involved in the reconstruction. The spikes before the onset of the AM tone (at 105 ms) are the spontaneous activity of the 10 HSR fibers.

3.2. Artefactual reconstruction

To avoid artefactual reconstruction, the method requires that summed spike trains are sufficiently longer in duration than the reconstruction filter and that those trains contain a minimum number of spikes. In practice, we found that spike trains should have a duration of at least six times that of the reconstruction filter and they should contain at least five spikes. Furthermore, we controlled for any artefactual reconstruction when it was more likely to occur, e.g., when it was done with all three fiber types and many CF channels. The control involved redistributing all spikes randomly in time ('shuffling') before reconstruction. Because shuffled spikes are unrelated to the reference signal, any reconstruction from shuffled spike trains would be artefactual. We found that, in these cases, the artefactual reconstruction SNR was always $<$1.0 dB.

3.3. On-frequency information encoding across CFs

We first assessed the information encoded by on-frequency fibers across channels. Stimuli were AM tones with carrier frequencies at the CF and were presented 20 dB above the threshold for HSR fibers at each CF (defined above). One AM tone was presented at a time. Only spike trains from fibers with CFs at the carrier

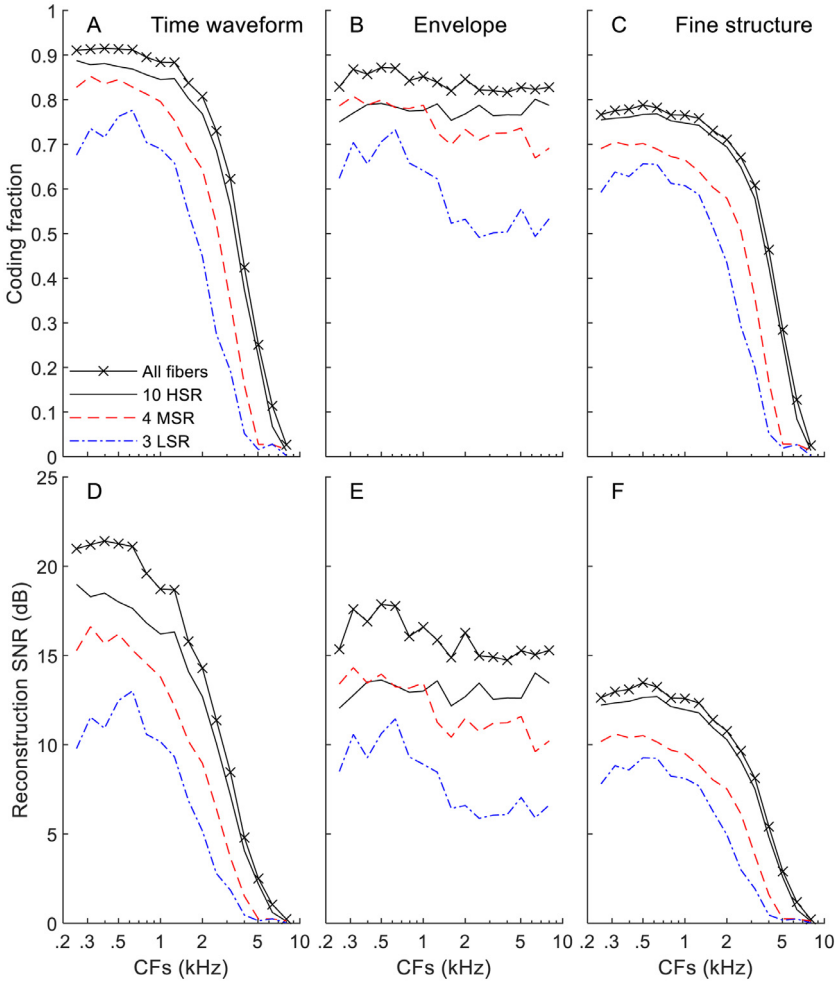

Fig. 4. On-frequency encoding as a function of CF. Each data point is the information encoded by fibers innervating the cochlear channel whose CF matched the frequency of the AM tone carrier. Two information metrics are shown: the coding fraction (**top**) and the reconstruction SNR (**bottom**). Each metric is shown as a function of CF and for the waveform (**A, D**), the ENV (**B, E**), and the TFS (**C, F**). Stimuli were 20 dB above the neural threshold for each individual CF (see the main text for details). Each panel shows information encoded separately by each fiber type (HSR, MSR and LSR), as well as jointly by all fiber types.

frequency were used for reconstruction (on-frequency reconstruction). To verify that the reconstruction SNR (in dB) provides similar results as the coding fraction, which is a more conventional metric (Wessel et al., 1996; Gabbiani and Metzner, 1999), the quality of reconstruction was assessed here using the two metrics.

Fig. 4 shows the values for the two metrics as a function of CF, for each fiber type (HSR, MSR and LSR) and separately for the waveform, ENV and TFS. A visual comparison of the results in the top and bottom panels shows that the two metrics yield similar results. This corroborates that they are roughly equivalent measures of reconstruction quality. For this reason and because of its conceptual simplicity, we will use the reconstruction SNR as the information metric in the rest of this paper.

For the waveform (Fig. 4D) and TFS (Fig. 4F), the reconstruction SNR is highest for HSR fibers, followed by MSR and LSR fibers. This indicates that 10 HSR fibers per CF encode more waveform and TFS information than 4 MSR or 3 LSR fibers. For all fiber types, the waveform and TFS information decreases with increasing CF. This is because neural phase-locking decreases with increasing frequency (Johnson, 1980; Palmer and Russell, 1986) and, here, the CF matched the AM carrier frequency. Phase-locking rolls off above around 1.5 kHz primarily because of the low-pass filter characteristics of the IHC (Palmer and Russell, 1986) and secondarily because of the refractory period (0.75 ms) (Sumner et al., 2002). Unlike TFS or waveform information, ENV information is approximately constant across CFs (Fig. 4E). At lower CFs ($<$ 1 kHz), HSR and MSR fibers carry approximately the same amount of ENV information, while at higher CFs HSR fibers encode more ENV information than MSR or LSR fibers.

Fig. 4 also shows that sometimes the amount of ENV and TFS information encoded by all fibers together (10 HSR + 4 MSR + 3LSR; crosses) is close to the information encoded by a single fiber group type alone. This typically occurs when the fiber type in question encodes $>$3 dB more information than other fiber types and indicates that the fiber type in question 'dominates' information encoding. For example, HSR fibers dominate TFS information encoding throughout the frequency range (i.e., the continuous line nearly overlaps the crosses in Fig. 4F). When two fiber types encode information equally well (e.g., HSR and MSR fibers at CF=0.5 kHz in Fig. 4E), the total information is typically 3 dB higher than the information encoded by each individual fiber type.

The results in Fig. 4 are for a fixed stimulus level of 20 dB above the discharge-rate threshold of HSR fibers. This may have favored the encoding of the stimuli by a particular fiber type if the chosen level excites a particular fiber type within its 'best' dynamic range (i.e., within the steepest part of the rate-level function). Fig. 2 shows that the chosen level (20 dB above the arrows) was so low that (1) LSR fibers responded at or just above their

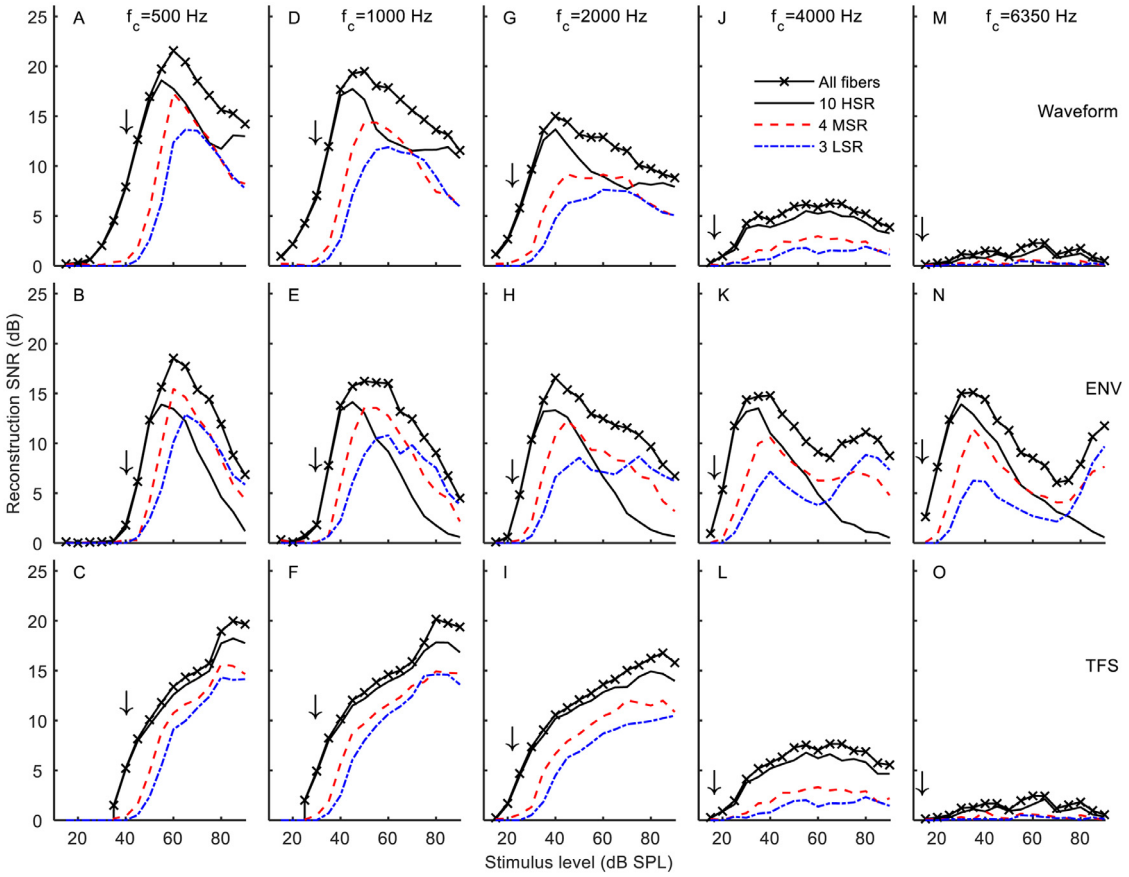

Fig. 5. On-frequency encoding as a function of level. Each data point is the information encoded by fibers innervating the cochlear channel whose CF matched the frequency of the AM tone carrier (indicated at the top of each column). Information is shown for the time waveform (**top** panels), ENV (**mid** panels) and TFS (**bottom** panels). Each panel shows information encoded separately by each fiber type (HSR, MSR and LSR), as well as jointly by all fiber types.

threshold, and hence were hardly able to encode the AM tone; (2) MSR fibers could encode the AM tone with moderate fidelity as the peaks of the AM tone (27.2 dB above the arrows) were clearly above the thresholds for MSR fibers; and (3) HSR fibers were excited in their best dynamic range except that the AM tone peaks may have produced momentary saturation. In other words, the different encoding fidelity of HSR/MSR/LSR fibers in Fig. 4 is reasonably well explained by the rate-level functions.

3.4. On-frequency information encoding across sound levels

We investigated the extent that different fiber types encode information at different levels. Fig. 5 shows the reconstruction SNR separately for each fiber type and for all three types combined, for five AM tones with different carrier frequencies (indicated at the top) and levels (abscissae). Different rows illustrate the reconstruction SNR for the full waveform (top), ENV (middle) and TFS (bottom). Reconstruction was done using only spike trains from fibers with CFs equal to the carrier frequency (on-frequency reconstruction).

The ENV (middle row in Fig. 5) was better encoded by HSR fibers at levels below the saturation threshold of HSR fibers (~40-50 dB SPL, depending on the CF, Fig. 2). At higher levels, MSR and LSR fibers encoded the ENV better than did HSR fibers. In other words, HSR fibers encoded a very large proportion of the total ENV information (compare continuous lines with crosses) at lower levels, while MSR and LSR fibers jointly encoded a very large proportion of the total ENV information at higher levels. The total ENV information sometimes varied nonmonotonically with increasing level. For example, at CF=6.3 kHz (Fig. 5N) the total information increased with increasing level from 20 to 40 dB SPL, decreased with further increases in level from 40 to 60-70 dB SPL, and increased again above 70 dB SPL. Similar nonmonotonic trends have been reported in the synchrony of spikes to the AM modulation frequency of single AN fibers (e.g., Fig.5C in Joris and Yin, 1992).

The TFS (bottom row in Fig. 5) was better encoded by HSR fibers than by MSR or LSR fibers throughout the range of levels, even at levels where HSR fibers discharged at saturation and MSR/LSR fibers were operating within their best dynamic range (~60-70 dB SPL, Fig. 2). Indeed, HSR fibers encoded a very large proportion of the total TFS information over virtually the whole range of levels. The good encoding of TFS at high levels, even when fibers discharge at saturation, has already been reported by experimental studies of single-fiber responses (e.g., Joris and Yin, 1992; HSR fibers: their Fig. 4A; LSR fibers: their Fig. 8A).

For all input levels, TFS and waveform information decreased gradually with increasing carrier frequency above 1 kHz, i.e., information was overall less in the right-most panels than in the left-most panels of the top and bottom rows in Fig. 5. This is due to phase-locking decreasing gradually with increasing frequency above 1.5 kHz (Johnson, 1980; Palmer and Russell, 1986).

3.5. Effect of fiber count on on-frequency information

In the previous sections, we have analyzed on-frequency ENV and TFS information encoded by different fiber types. Information, however, was based on summed spike trains for 'normal' fiber counts, i.e., 10 HSR, 4 MSR, and 3 LSR fibers per CF. Because more fibers can theoretically encode more information, differences in information across fiber types may reflect differences in fiber counts.

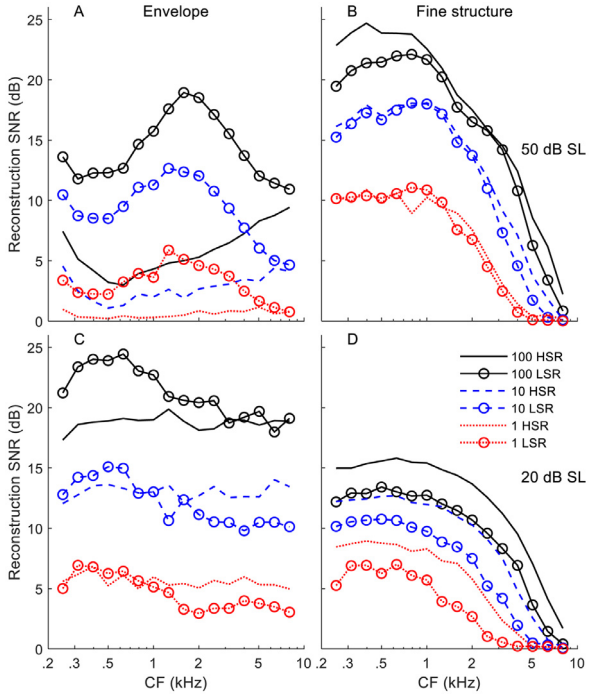

Fig. 6. Effect of fiber count on ENV (left) and TFS (right) encoding by on-frequency HSR and LSR fibers. Results are shown for AM tones at 50 dB (top) and 20 dB (bottom) above the threshold for HSR fibers at each CF.

In this section, we investigate ENV and TFS information encoded by equal counts of fibers per CF. Only spike trains from HSR or LSR fibers with CFs at the carrier frequency were used for reconstruction (on-frequency reconstruction). Results are shown in Fig. 6 for AM tones at 20 and 50 dB above the discharge rate of HSR fibers at each CF, denoted as 20 and 50 dB sensation level (SL), respectively.

Regarding ENV information, at 20 dB SL (Fig. 6C), a single HSR fiber tends to encode more ENV information than a single LSR fiber because at this level the HSR fiber is in its best dynamic range while the LSR fiber is only just above threshold. At 50 dB SL (Fig. 6A), a single LSR fiber can encode the ENV while a single HSR fiber cannot because it is in saturation. At either of the two levels, however, 100 fibers of either type encode much more ENV information than one fiber of either type alone.

Regarding TFS encoding (Fig. 6B, D), at the two levels tested, a single fiber of either type encodes approximately constant TFS information up to ~1 kHz; information decreases at higher frequencies because of decreasing phase locking. However, 100 fibers encode more information and over a wider range of frequencies than one fiber alone. Indeed, 100 fibers encode approximately the same amount of TFS information at 4 kHz as one fiber does at 500 Hz. Furthermore, for an equal number of fibers, there is more TFS information at higher than at lower levels, e.g., for one HSR fiber at CF=1 kHz the reconstruction SNR is 10 dB at 50 dB SL compared to 5 dB at 20 dB SL.

In summary, increasing the number of fibers increases the ENV and TFS information encoded by on-frequency fibers, as well as the range of frequencies over which the TFS is encoded. Increasing level also increases TFS information encoded by on-frequency fibers.

3.6. Off-frequency encoding at various sound levels

So far, we have focused on information encoding by fibers whose CF matched the AM carrier frequency (on-frequency encoding). The information encoded by fibers with CFs different from the carrier frequency (off-frequency fibers) was ignored. However, as sound level increases, cochlear excitation spreads and off-frequency fibers can contribute to encoding. In other words, on-frequency information (as shown in Fig. 5) is likely an underestimate of the total information encoded in the AN, particularly at high levels. In this section, we assess information encoding by on- and off-frequency fibers for increasing levels of the AM tone.

Results for the ENV and TFS are shown in Fig. 7 and Fig. 8, respectively. In each figure, each column corresponds to a different carrier frequency (indicated at the top) and each row corresponds to a different level (indicated on the right).

Fig. 7 shows that the range of CFs that encodes the ENV widens as level increases, i.e., when going from the bottom to the top row in the figure. The widening is comparatively greater for lower carrier frequencies (compare Fig. 7E with 7A) than for higher carrier frequencies (compare Fig. 7T with **7**P). This suggests that virtually all neurons (all fiber types at all CFs) can encode the ENV for a low-frequency carrier at high levels. This is probably due to the basalward spread of cochlear excitation at higher levels. For CFs close to the carrier frequency, the discharge of HSR fibers starts to saturate at levels <60 dB SPL (Fig. 2). Saturation causes a reduction of ENV information encoded by on-frequency HSR fibers together with an increase in the information encoded by off-frequency HSR fibers, resulting in an inverted "W" pattern (Fig. 7A, B, or C). Remarkably, sometimes the ENV is better encoded by off-frequency HSR fibers than by on-frequency LSR fibers. For example, for a 90-dB SPL, 500-Hz AM tone (Fig. 7A), HSR fibers with CF=4 kHz encode the ENV just as well as do MSR/LSR fibers with CF=2 kHz, and the two groups encode the ENV better than do on-frequency LSR fibers (CF=500Hz).

Fig. 8 shows TFS encoding as a function of CF. As shown, HSR fibers encode more TFS information than LSR or MSR fibers across all CFs and levels, even at the highest levels tested where HSR fibers discharge at saturation (top row in Fig. 8).

3.7. ENV and TFS encoding by saturated HSR fibers

The greater number of HSR than LSR fibers may explain why HSR fibers encoded the TFS better (e.g., Fig. 8E) than do the other fiber types at low stimulus levels (< 40 dB SPL, Fig. 2) but would not explain why they also encoded the TFS better at higher levels (Fig. 8B), where HSR fibers discharge at saturation and LSR do not. It is also unclear why saturated, on-frequency HSR fibers encode well the TFS (Fig. 8) but not the ENV (Fig. 7). Here, we investigate the reason behind these paradoxical findings.

First, we investigated why HSR fibers can encode the TFS even when saturated. For this purpose, Fig. 9 illustrates a 15-ms-long excerpt of the summed spike trains for 10 and 100 on-frequency HSR fibers in response to an AM tone with 2-kHz carrier frequency at 70 dB SPL (as in Figs. 7M and 8M), a level where HSR fibers were clearly saturated (Fig. 2). Fig. 9 also illustrates corresponding excerpts for the reference signal TFS (Fig. 9A, D), the reconstructed signals in each case (red lines) and the reconstruction SNRs (insets) computed over the whole stimulus duration. For comparison, the figure also illustrates corresponding plots for 3 and 100 LSR fibers (Fig. 9E, F).

Fig. 9 demonstrates that the HSR fibers produce more discharges than LSR fibers at levels at which HSR fibers are saturated and LSR are not (compare Fig. 9B, C with Fig. 9E, F). The reconstruction SNRs (insets in Fig. 9) show that HSR fibers also encoded the TFS slightly better than LSR fibers. The reason is that because of phase locking, spikes are more likely to occur in synchrony with the carrier (even in saturation) but they occur stochastically, thus at different times for different fibers. Therefore, the TFS is better encoded in the summed spike train for a larger group of fibers than in the spike trains for a small group of fibers. By the same

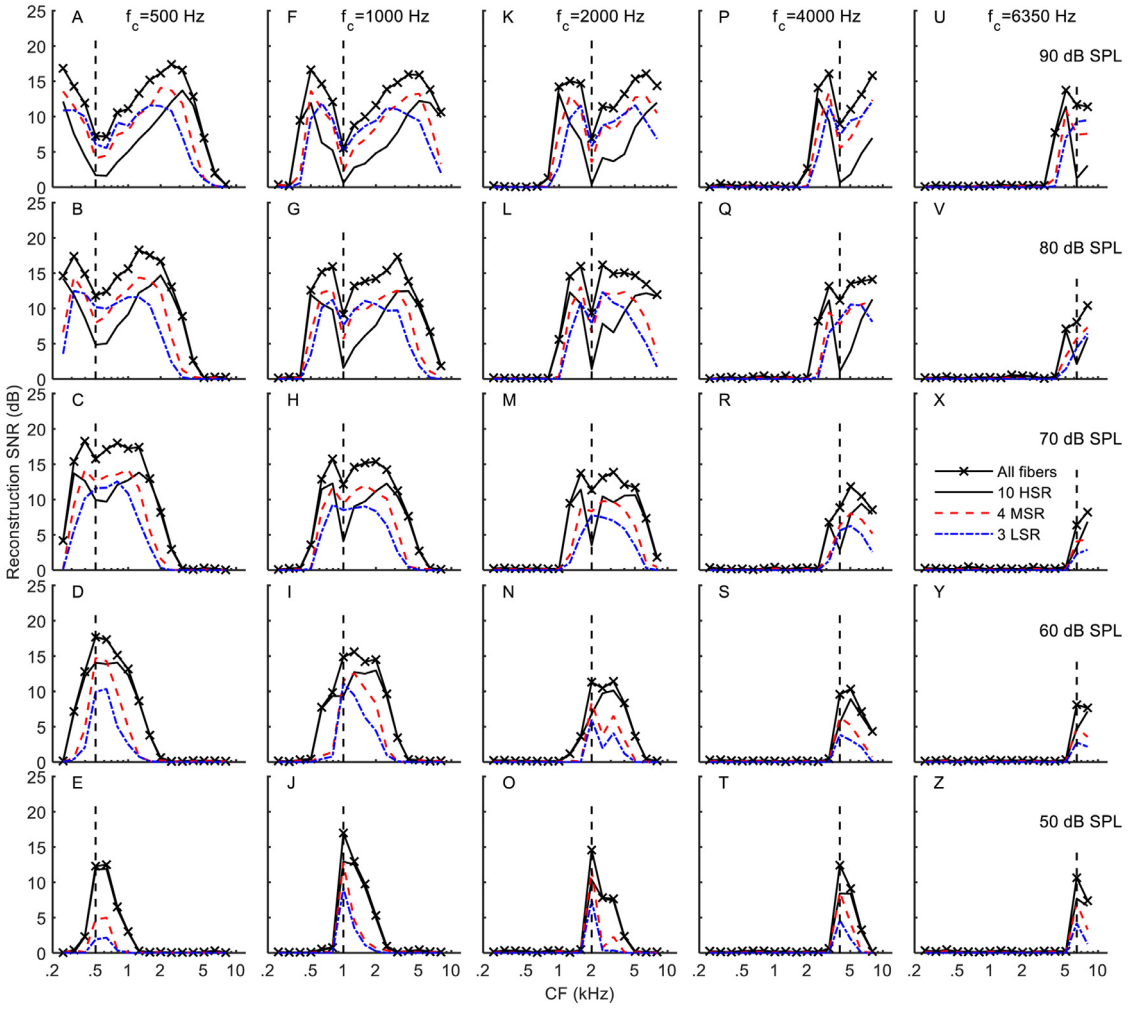

Fig. 7. ENV encoding by fibers with CFs equal to and different from the frequency of the AM tone carrier. Each column is for a different AM carrier tone frequency, as indicated at the top. Each row is for a different level of the AM tone as indicated in the right-most panels. Vertical dashed lines indicate CF=carrier frequency.

principle, the TFS is better encoded by fibers with higher than with lower discharge rates.

Paradoxically, the same saturated on-frequency HSR fibers that encoded the TFS well (Fig. 8M), encoded the ENV poorly (Fig. 7M). To investigate the reason for this, Fig. 10 shows a 250-ms-long excerpt of the summed spike trains produced by 10 and 100 on-frequency HSR fibers (Fig. 10B, C) in response to the same AM tone as in Fig. 9**.** Also shown are the corresponding reference ENV signals (Fig. 10A, D), the reconstructed reference signal ENVs (red lines) and reconstruction SNRs (insets) computed over the whole stimulus duration. For comparison, Fig. 10 also shows plots for 3 and 100 on-frequency LSR fibers (Fig. 10E, F).

Consistent with the results shown in Fig. 6, ENV encoding improves with increasing the number of fibers for both HSR (Fig. 10B, C) and LSR fibers (Fig. 10E, F). Interestingly, the ENV is better encoded (i.e., the reconstruction SNR is higher) by LSR than HSR fibers, no matter if we compare for large and equal fiber counts (Fig. 10C versus Fig. 10F) or for the nominal fiber count (10 HSR, Fig. 10B versus 3 LSR, Fig. 10E). One feature that differentiates the summed spike trains of HSR and LSR fibers is the higher spontaneous rate of HSR fibers, which is visible in the soft parts of the ENV (at times of around 1.305, 1.405 and 1.505 secs). Another differentiating feature is that the summed spike trains of HSR fibers are more square-like while those of LSR fibers are more sinusoid-like (compare the blue traces in Figs. 10C and **10**F). The more square-like response of HSR fibers at high levels possibly reflects the stronger onset response of HSR fibers (Müller and Robertson, 1991) and is consistent with experimental responses for high-CF fibers. For example, the cycle histograms in Fig. 4B of Joris and Yin (1992) are more square-like at higher than at lower levels, although they show smaller onset responses than the present simulations. Together with the spontaneous activity, the greater onset responses of HSR fibers, even when they discharge at saturation, possibly contributes to distorting the ENV representation in the discharge rate of saturated HSR fibers.

We also showed earlier that off-frequency HSR fibers contribute substantially to ENV encoding and this raises the question of why they are not also affected by these distortions. A likely explanation is that because onset responses are level-dependent (e.g., Müller and Robertson, 1991), the lower excitation levels present at off-frequency regions produce smaller or no onset responses in each ENV cycle (bottom panels in Fig. 10).

In summary, the TFS is better encoded by HSR than LSR fibers probably because HSR fibers are more numerous and have greater discharge rates. The stronger onset response of HSR fibers in each ENV cycle probably distorts ENV encoding when HSR fibers discharge at saturation. On-frequency LSR fibers and off-frequency HSR fibers have smaller (or absent) onset responses at levels where

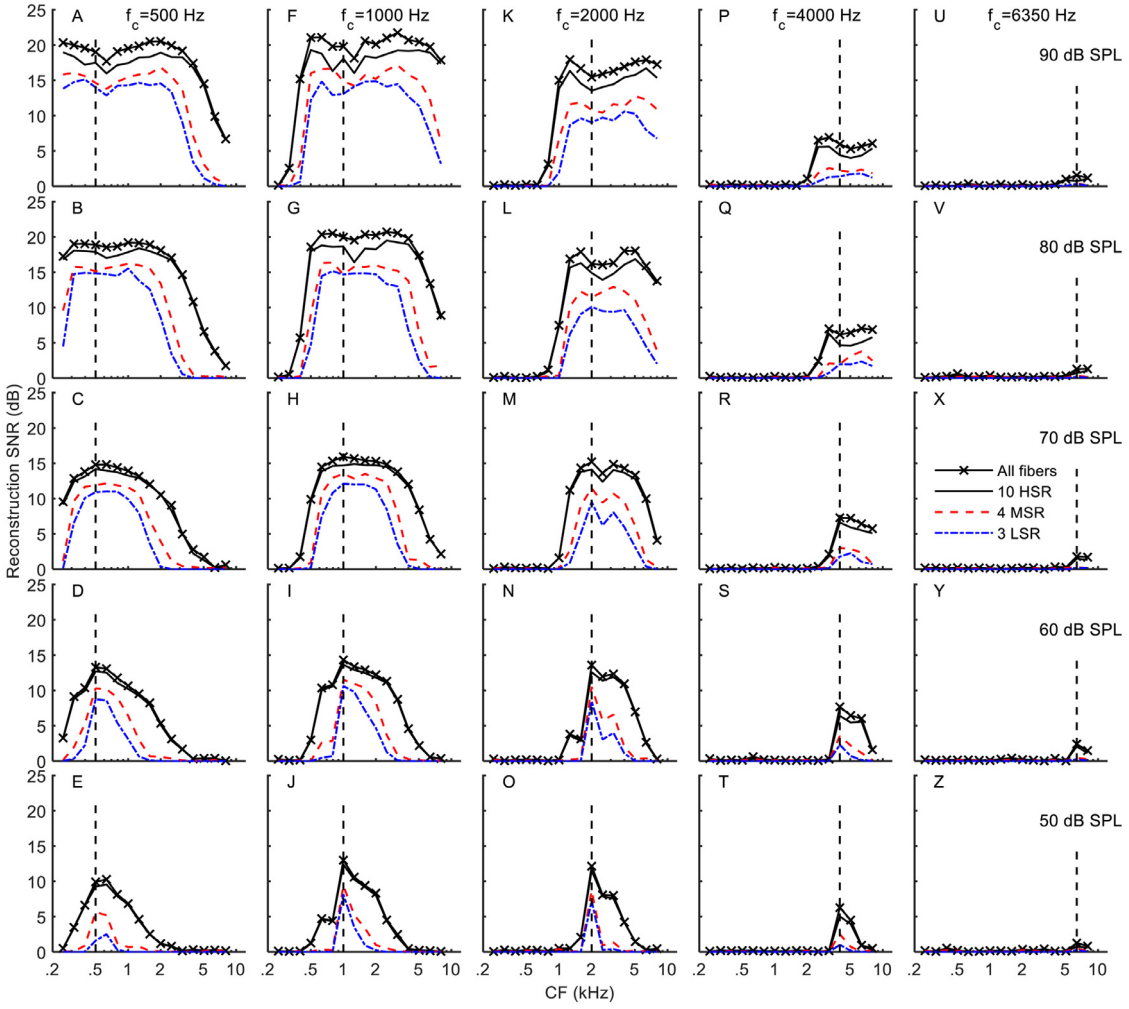

Fig. 8. As Fig. 7 but for TFS.

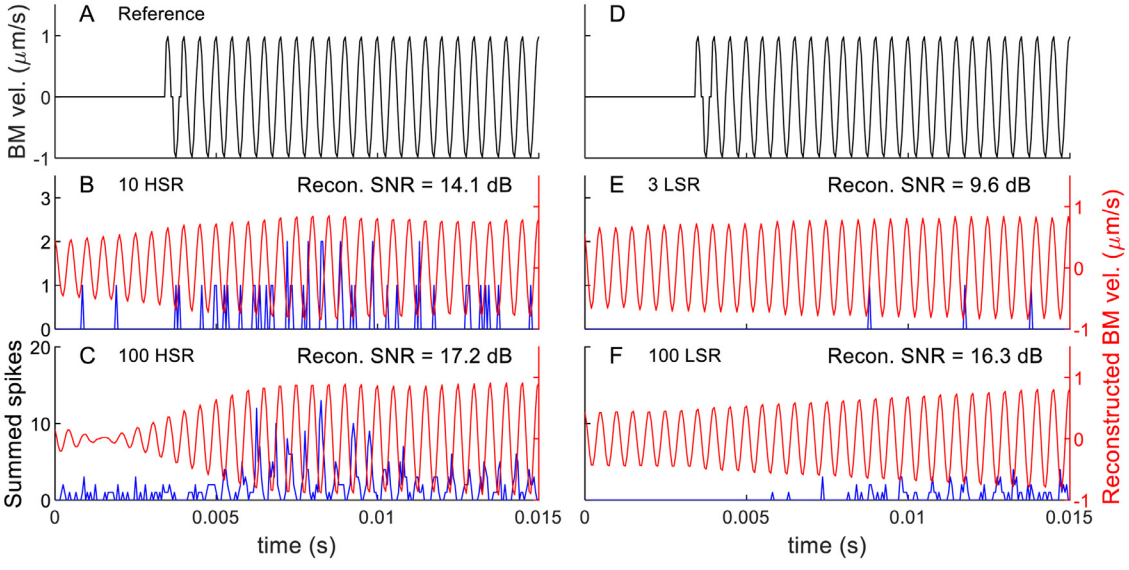

Fig. 9. A comparison of TFS encoding in the summed spike trains with different number and types of fibers. The stimulus was an AM tone at 70 dB SPL, with a carrier frequency of 2 kHz. The top panels (**A, D**) illustrate the reference signal, which in this case was the TFS of the simulated BM response for CF=2 kHz. The left and right columns illustrate results for HSR and LSR fibers, respectively. Each row illustrates summed spike trains for the nominal number of fibers (10 HSR and 3 LSR) and for 100 HSR and LSR fibers, hence the different ordinate scales. Red lines depict the reconstructed BM velocity (right ordinate) and the insets show the corresponding reconstruction SNRs (in dB SNR). Note that visually, the TFS is better encoded with more fibers and more spikes, hence it is better encoded by 100 HSR (**C**) than by 100 LSR fibers (**F**), even though HSR fibers were in saturation and LSR fibers were not.

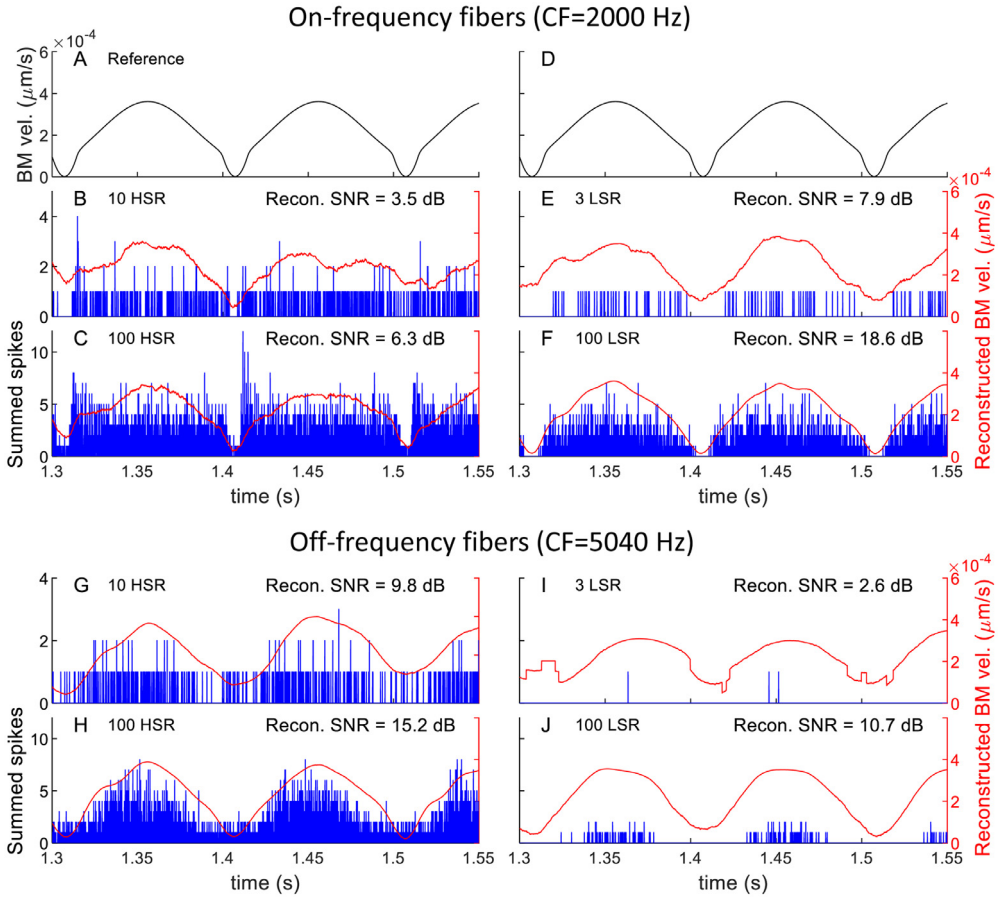

Fig. 10. A comparison of ENV encoding in the summed spike trains with different number and types of fibers from on- and off-frequency regions. The stimulus was an AM tone at 70 dB SPL, with a carrier frequency of 2 kHz. The top panels (A, D) illustrate the reference signal, which in this case was the ENV of the simulated BM response for CF=2 kHz. Panels B, C, E and **F** illustrate summed spike trains for on-frequency fibers (CF=2 kHz), while panels **G**, H, I, and J illustrate summed spike trains for off-frequency fibers (CF=5040 Hz). The left and right columns illustrate results for HSR and LSR fibers, respectively. Each row illustrates summed spike trains for the nominal number of fibers (10 HSR and 3 LSR) and for 100 HSR and LSR fibers, hence the different ordinate scales. Reds lines depict the reconstructed BM velocity (right ordinate) and the insets show the corresponding reconstruction SNRs (in dB SNR).

HSR fibers are saturated, which makes them better suited to encode the ENV when on-frequency HSR fibers discharge at saturation (Fig. 7).

3.8. ENV encoding by the population of fibers

So far, we have analyzed ENV information encoding in independent cochlear channels with CFs equal to or different from the AM carrier frequency. Information, however, can be *jointly* encoded by multiple channels and we have shown (Fig. 7) that off-frequency fibers can encode the ENV as well or better than on-frequency fibers, particularly at high levels. In this section, we investigate ENV encoding by all fiber types jointly and by an increasingly greater number of channels with CFs around the AM carrier frequency.

Stimuli were AM tones at 80 dB SPL. Because up to 48 summed spike trains (3 fiber types × 16 CF channels) could be used in the reconstruction, the duration of the AM tone was longer here (30 seconds) than in the preceding simulations to avoid artefactual reconstruction. Due to computational limitations, the reconstruction was done at a sampling frequency of 5 kHz; that is, the model was run at the regular sampling frequency of 20 kHz, but the reference ENV was downsampled to 5 kHz and spikes were binned every 0.2 ms (1/5000) without discarding any spike. The reconstruction was done for a gradually increasing number of channels, starting from one channel (the on-frequency channel with CF=f_c), followed by three channels (the on-frequency channel plus two adjacent channels on each side of the CF: CF-1, CF=f_c, CF+1), followed by five channels (CF-2, CF-1, CF=f_c, CF+1, CF+2), and so on. The range of CFs used in each case is shown in Table 2.

Results show that for normal fiber counts (circles in Fig. 11A-E), ENV information increases asymptotically as spike trains from an increasingly number of channels (or equivalently, an increasingly wider BM region around the CF=f_c) are used in the reconstruction. Information 'saturates' (i.e., reaches its maximum value) only when a sufficient number of channels around the AM tone carrier frequency are included. The specific number varies depending on the carrier frequency, but it is overall larger for lower than for higher carrier frequencies.

3.9. Effect of fiber loss on total ENV information

So far, we have investigated information encoding in the 'healthy' AN. In this section, we investigate the impact of fiber loss on ENV information encoding. Several synaptopathy scenarios are considered, as shown in Table 1. These include partial and total loss of MSR and LSR fibers with normal counts of HSR fibers, as well as partial and total loss of HSR fibers with normal counts of LSR and MSR fibers. In all cases, fiber counts (and losses) were identical across channels. Stimuli and simulations were as described in the preceding section and results are illustrated in color in Fig. 11A-E for direct comparison with results for the 'healthy' ear (circles in Fig. 11A-E).

Table 2
Range of CFs (in Hz) of the channels used in the reconstructions shown in Fig. 11. The used number of channels (left column) ranges from one (the on-frequency channel with CF=f_c) to 16 (all channels). For example, when 3 channels were used to reconstruct the ENV of an AM tone with a carrier frequency of 500 Hz, the used channels had CFs from 400 Hz to 630 Hz. Note that the number of channels used for the reconstruction of AM tones with f_cs of 500 Hz and 6350 Hz are not always exactly equal to the number of channels indicated in the left column because the lowest and highest CFs were 250 Hz and 8000 Hz, respectively, and when increasing the number of channels, the channels were included symmetrically around the channel with CF=f_c.

	AM carrier frequency (f_c)				
Number of channels	500	1000	2000	4000	6350
1 (on-frequency)	500	1000	2000	4000	6350
3	400–630	794–1260	1587–2520	3175–5040	5040–8000
5	315–794	630–1587	1260–3175	2520–6350	4000–8000
7	250–1000	500–2000	1000–4000	2000–8000	3175–8000
11	250–1587	315–3175	630–6350	1260–8000	2000–8000
16 (all)	250–8000	250–8000	250–8000	250–8000	250–8000

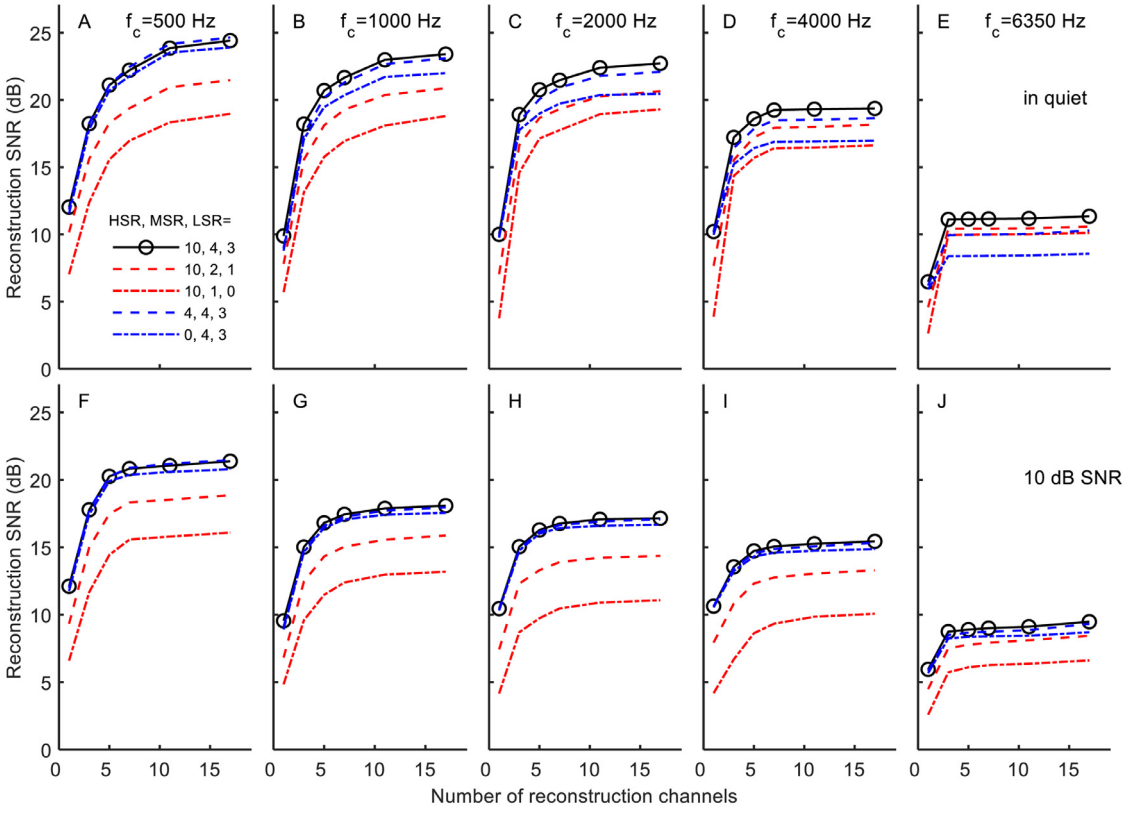

Fig. 11. Reconstruction SNR for the ENV of an AM tone as a function of the number of channels used in the reconstruction (abscissa). Reconstruction was done using spike trains from fibers with CFs at and around the AM tone carrier frequency. The AM tone was presented at 80 dB SPL either in quiet (upper row) or embedded in a notch noise at 10 dB SNR (lower row). The spike trains from all fiber types were used jointly in the reconstruction. Different traces illustrate reconstruction SNR values for different synaptopathy scenarios, as indicated by the inset where number triads refer to the number of HSR, MSR, and LSR respectively. In other words, circles illustrate reconstruction SNRs for the healthy ear with 10 HSR, 4 MSR and 3 LSR fibers. Red lines indicate moderate or almost total loss of LSR and MSR fibers, and blue lines illustrate moderate or total loss of HSR fibers.

For lower carrier frequencies, ENV information was reduced more by LSR fiber loss (red lines) than by HSR fiber loss (blue lines). For example, for the 500-Hz carrier (Fig. 11A), total loss of LSR fibers with normal counts of HSR fibers reduced the total information from about 25 dB SNR to 18 dB SNR, while total exclusive HSR fiber loss barely reduced the information. This is probably because LSR fibers are overall better suited than HSR fibers to encode ENV information at the chosen level of 80 dB SPL (Figs. 5, 6). The relative importance of LSR versus HSR fiber loss, however, changes gradually with increasing carrier frequency. Indeed, the pattern reverses for the 6350-Hz carrier, where total, exclusive loss of HSR fibers causes a greater reduction of ENV information than total, exclusive loss of LSR fibers (Fig. 11E).

Importantly, even for low carrier frequencies, the total ENV information (in the asymptote) remains high with total, exclusive loss of LSR fibers. This is probably because off-frequency HSR fibers encode the ENV at 80 dB SPL, when on-frequency HSR fibers do not (Fig. 7).

3.10. Diagnosing LSR fiber loss

Cochlear synaptopathy is thought to predominantly affect LSR synapses and fibers (Furman et al., 2013; Schmiedt et al., 1996), although this notion has been recently challenged (Suthakar and Liberman, 2021). The results shown earlier suggest that the population of HSR fibers dominates TFS encoding (Fig. 8), and can

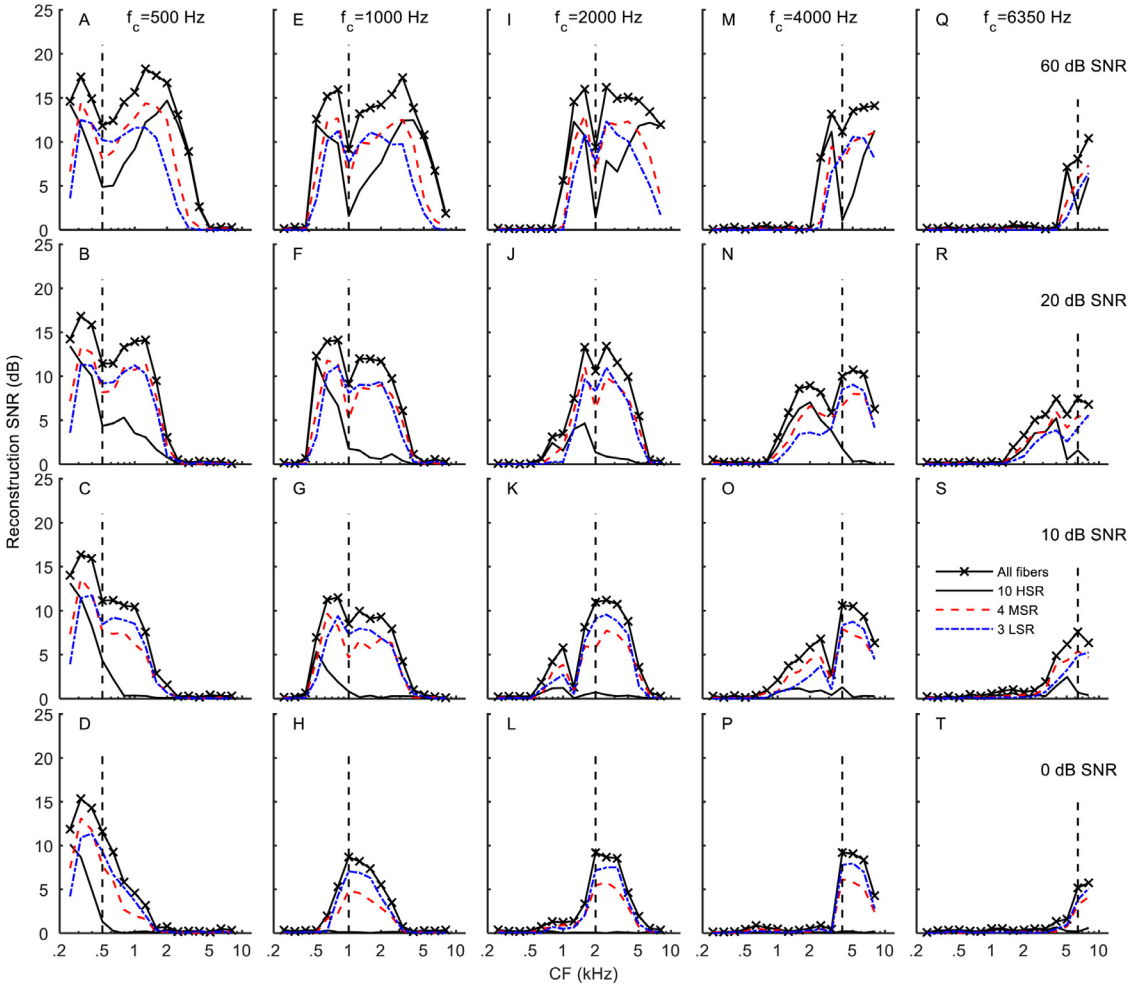

Fig. 12. ENV encoding by fibers with CFs equal to and different from the frequency of the AM tone carrier (indicated at the top) when the AM tone is embedded in a notch noise. Reconstruction was done using spikes from fibers of different types (see the inset) with CFs as indicated in the abscissa. The level of the AM tone was 80 dB SPL. Each column is for a different AM carrier tone frequency, as indicated at the top. Each row is for a different noise level, as indicated by the probe-to-noise ratio (in dB SNR) in the right-most column.

encode large amounts of ENV information (Fig. 11A-E). These findings suggest that LSR fibers may not be essential for ENV or TFS encoding in quiet, which could make it challenging to diagnose LSR fiber loss using TFS- or ENV-based measures.

It might nevertheless be possible, however, to diagnose LSR fiber loss by using (behavioral or electrophysiological) tests that emphasize the use of ENV information encoded by on-frequency LSR fibers. One potential way to do it is by presenting an AM probe embedded in noise with a spectral notch around the AM carrier frequency. The idea is that the noise will mask ENV encoding by off-frequency HSR fibers with minimal or no effect on the information encoded by on-frequency LSR fibers. If synaptopathy predominantly causes a loss of LSR fibers, ENV information would be more sensitive to LSR fiber loss at and around the cochlear region tuned to the carrier frequency with than without the notch noise.

To explore this idea, we considered the ENV information for an AM tone at 80 dB SPL as the best starting point before adding a notch noise. As shown in Fig. 7 (second row), at this level, LSR and MSR fibers contribute the most to the total information at or near on-frequency channels and hence a test at this intensity would likely provide the best sensitivity to LSR and/or MSR fiber loss at cochlear regions around the carrier frequency. Although the relative contribution from LSR and MSR fibers compared to HSR fibers was even larger at 90 dB SPL (top row in Fig. 7), we deemed 90 dB SPL as too loud for listener comfort in an eventual diagnostic test. We then searched for the optimal level for an added notch noise. The noise had a flat spectrum except for the notch, which was spectrally centered at the carrier frequency (f_c) and spanned the frequency range [$0.8\times f_c$, $1.2\times f_c$]. The overall level of the noise relative to the overall probe level (i.e., the signal-to-noise ratio, SNR) varied from 0 to 20 dB in steps of 10 dB with an additional reference condition without noise (60 dB SNR). Note that the results for this reference condition are those in the second row of Fig. 7.

The results (Fig. 12) suggested that 10 dB SNR was a good probe-to-noise ratio, because off-frequency HSR-fiber contributions to ENV encoding were highly reduced (with respect to the no-notch condition) for lower carrier frequencies (Fig. 12C, G) and close to zero for higher frequencies (Fig. 12K, O, S) with only a moderate reduction in the information encoded by LSR fibers (compare the blue lines in the no-notch condition in the first row of Fig. 12 with those of the 10 dB SNR notch-noise condition in the third row of Fig. 12). A large reduction of total information (as would occur if noise was presented at 0 dB SNR, bottom row of Fig. 12) might be inconvenient because it would make an eventual behavioral test harder for the subject and/or might increase the variability in AM scores when there is no clear envelope cue to track.

Fig. 11F-J illustrates the total ENV information and the effect of fiber loss when 80-dB SPL AM tones are presented together with the selected notch noise at 10 dB SNR. As expected, the

addition of the notch noise removed ENV-encoding contributions from remote cochlear sites, and hence information asymptoted to a smaller value and saturated with fewer channels (compare the circles in the top and bottom rows of Fig. 11).

As noted earlier, without the noise, total exclusive HSR fiber loss typically reduced the maximum information by 2-3 dB (compare the black circles with the blue dashed-dotted line in Fig. 11A-E). With the added noise, total HSR fiber loss had almost no effect on information (i.e., the blue lines overlap with the black circles in Fig. 11F-J). This indicates that adding a notch noise was a successful way of eliminating contributions from off-frequency HSR fibers.

Another effect of adding notch noise was that the information loss due to LSR loss was overall a larger proportion of the total information available without fiber loss. For a 500 Hz carrier without notch noise (Fig. 11A), the reconstruction SNR decreased from 24.6 dB to 19.0 dB when all LSR fibers were lost (i.e. a reduction of 22.7%); with the added notch noise (Fig. 11F), the reconstruction SNR decreased from 21.5 dB to 16.1 dB (a reduction of 25.1%). The reduction was larger for higher carrier frequencies. For example, for a 2 kHz-carrier without noise (Fig. 11C), the reconstruction SNR decreased from about 23 dB to 19 dB (i.e. a reduction of 17.3%) when all LSR fibers were lost. With the added notch noise, the reconstruction SNR decreased from about 17 dB to 11 dB (i.e. a reduction of 35.3%), again when all LSR fibers were lost (Fig. 11H).

In summary, adding a carefully chosen notch noise to the AM tone minimized ENV encoding by off-frequency HSR fibers, and caused LSR fiber losses to yield greater proportional losses of total ENV information. Therefore, ENV-based measures that employ high-level probes embedded in a notch noise seem a promising way to diagnose LSR fiber loss, at least when fiber loss occurs throughout the cochlea, as was the case here.

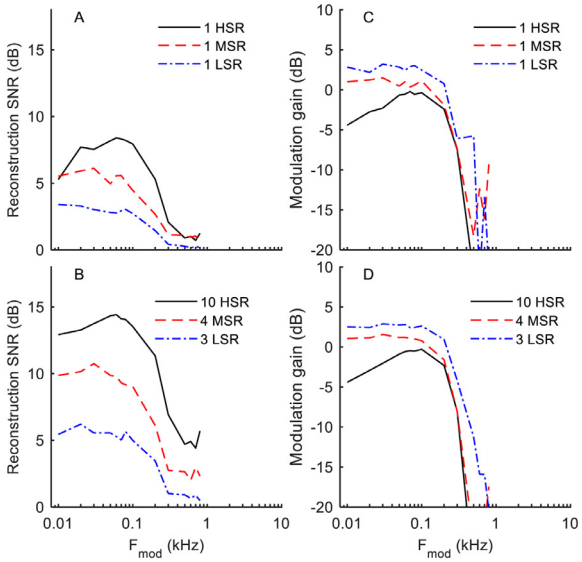

Fig. 13. A comparison of reconstruction SNR (left column) and modulation gain (right column) for different modulation frequencies (modulation transfer function). The top row depicts simulations for a single fiber of each type, while the bottom row shows simulations for the nominal fiber counts, as indicated by the inset. The AM tone carrier frequency was 2 kHz and the stimulus level was 40 dB SPL. Only spike trains generated by on-frequency fibers were used in the calculations. After Kale and Heinz (2012), modulation gain was calculated as $20\times\log_{10}(2R/m)$, where R is the vector strength and m the modulation depth, which was equal to 1.0.

4. Discussion

We have applied information theory to investigate temporal information encoding in the (simulated) AN. Our approach involves the optimal reconstruction of the stimulus from the neural response followed by an assessment of the reconstruction quality. The latter is used as the metric for information encoding. With this approach, we have assessed (1) the relative importance of HSR, MSR and LSR fibers for encoding ENV and TFS information; (2) the relative importance of fibers with CFs equal to and different from the stimulus frequency, i.e., the relative importance of on- and off-frequency encoding; and (3) the impact of fiber loss, particularly LSR fiber loss (synaptopathy), on ENV encoding. Last, we have applied the proposed approach to explore ways of increasing the sensitivity of ENV-based measures to LSR fiber loss.

Our main findings were: (1) TFS is predominantly encoded by the population of HSR fibers across all stimulus levels, even when HSR fibers discharge at saturation; (2) ENV is poorly encoded by saturated, on-frequency HSR fibers possibly because their larger onset responses distort the neural representation of the ENV; (3) Off-frequency fibers of all types contribute significantly to ENV encoding. At high levels, where individual on-frequency LSR fibers are better suited for ENV encoding, off-frequency HSR fibers encode as much ENV information as do on-frequency LSR fibers; (4) ENV-based measures can increase their sensitivity to LSR fiber loss by using high-level probes embedded in a notch noise that masks ENV information encoded by off-frequency HSR fibers.

4.1. Model performance and limitations

As explained in the Methods section, the chosen AN model (Sumner et al., 2003a) accurately reproduces a wide range of AN experimental data and phenomena (reviewed by Lopez-Poveda, 2005; Meddis and Lopez-Poveda, 2010). Fig. 13C,D show that the model also simulates modulation transfer functions realistically (compare Fig. 13C, D with the experimental data in Fig. 3A of Kale and Heinz, 2012) as well as the fact that the synchronization of spike times with the AM modulating signal tends to be greater for LSR/MSR than for HSR fibers (Fig. 13C, D) (Johnson, 1980; Joris and Yin, 1992).

The model, however, does not simulate the middle ear reflex or the olivocochlear reflexes. Either or both these reflexes may be activated by the AM probe alone or in combination with background noise (reviewed by Lopez-Poveda, 2018) and their activation could shift the dynamic range of AN fibers to higher levels (e.g., Winslow and Sachs, 1987) extending the range of levels over which on-frequency fibers can encode the ENV (Marrufo-Perez et al., 2018, 2019; reviewed by Marrufo-Perez and Lopez-Poveda, 2022). Future investigations may explore the effect of middle ear and/or olivocochlear reflex activation on ENV and TFS encoding.

The present model implementation used 16 filterbank channels with CFs logarithmically distributed from 0.25 to 8 kHz. This number of channels is sufficient to demonstrate the importance of off-frequency contributions to TFS and ENV encoding (e.g., the inverted W-shape in Fig. 7) with a reasonable computational load. However, a greater number of channels with sufficient overlap between them might be necessary to investigate spectral (rather than temporal) information encoding for broadband stimuli.

The present model implementation involved 17 fibers per frequency channel (10 HSR, 4 MSR, and 3 LSR fibers). The total number of fibers was 255 (=16 channels ×17 fiber/channel), thus much smaller than the number of fibers in a healthy cochlea (~30,000). Three arguments, however, suggest that this is not an important limitation for the present study. First, Fig. 11 shows that ENV encoding reaches a plateau after combining spike trains from about 10 frequency channels around the on-frequency channel (which corresponds to 170 fibers). Second, Fig. 6 shows that increasing the fiber count from 1 to 10 improves ENV encoding more than increasing the fiber count from 10 to 100 fibers, which suggests

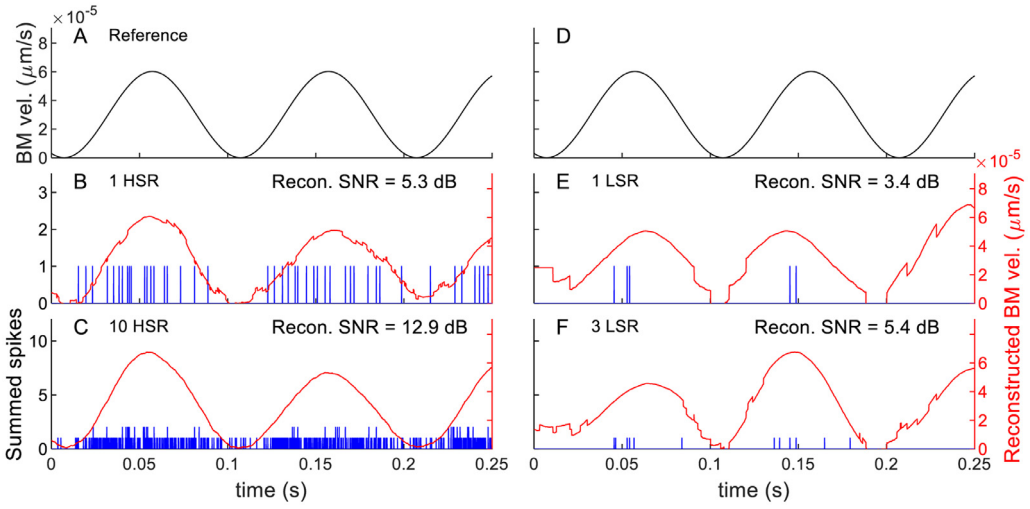

Fig. 14. A comparison of ENV encoding in the summed spike trains with different number and types of on-frequency fibers. The stimulus was an AM tone at 40 dB SPL with a carrier frequency of 2 kHz. Otherwise as in Fig. 10.

that ENV information probably saturates for fiber counts above about 200-500 fibers at least for the present AM tones in quiet. Third, it is conceivable that spike times become correlated with each other when the fiber count grows beyond a certain limit and hence adding more fibers does not increase information. However, a more realistic number of fibers may be necessary to investigate the encoding of more complex stimuli over a wide range of levels (see, e.g., Lopez-Poveda and Barrios, 2013).

4.2. Information encoding versus periodicity encoding

The reconstruction SNR (the metric used here to assess ENV and TFS encoding) differs significantly from vector strength and modulation gain (the metrics used in many previous studies of AM encoding) (e.g., Goldberg and Brown, 1969; Johnson, 1980; Joris and Yin, 1992; Kale and Heinz, 2012). Vector strength and modulation gain quantify the synchronization of spike times with the cycle of the stimulus. The reconstruction SNR, by contrast, quantifies how well a signal can be reconstructed on a time-by-time basis from the spike train(s). In other words, for a given spike train, vector strength depends on the stimulus frequency and hence it would be equal regardless of whether the evoking stimulus is a sinusoidal, triangular or a square wave with the same frequency (discussed by Joris et al., 2004), whereas the reconstruction SNR would be different in each case because it reflects how well the waveform can be reconstructed from the spike train.

Fig. 13 allows a direct comparison of the two metrics when used to assess ENV encoding by on-frequency fibers and for different AM modulation frequencies, i.e., when used to produce modulation transfer functions. The level was 40 dB SPL and the carrier frequency was a 2-kHz. The figure shows that increasing the number of fibers improves the reconstruction SNR but not the modulation gain (compare Fig. 13A with Fig. 13B and Fig. 13C with Fig. 13D). More importantly, a single HSR fiber shows less modulation gain (poorer synchrony) than a single LSR fiber (Fig. 13C) and 10 HSR fibers do not supersede the modulation gain of 3 LSR fibers (Fig. 13D). The opposite occurs for the reconstruction SNR, i.e., the ENV is reconstructed better from the spike train produced by a single HSR than a single LSR fiber (Fig. 13A) and the reconstruction SNR improves with increasing the number of fibers (compare Fig. 13B with 13A). Fig. 14 confirms that at the low stimulus level chosen for this example (40 dB SPL), the fewer spikes produced by LSR fibers (Fig. 14E, F) are more synchronous with the ENV (Fig. 14D) than the spikes produced by HSR fibers (Fig. 14B, C), hence the higher modulation gain for LSR fibers, and yet the reconstructed ENV is more accurate (the reconstruction SNR is higher) for HSR than LSR fibers.

In summary, the reconstruction SNR increases (within limits) with increasing the number of spikes and/or the number of fibers while vector strength does not. This raises an important question: which of the two metrics (vector strength or reconstruction SNR) is more representative of perception? In other words, does perception rely on the periodicity at which spike occurs or does it rely on an optimal reconstruction of the signal? Further research is necessary to investigate this issue.

4.3. TFS encoding

The present results show that HSR fibers convey more TFS information than LSR fibers. This occurs despite the fact that in the model, like in the physiology (e.g., Johnson, 1980; Joris and Yin, 1992; Louage et al., 2004; see also Fig. 13C, D), the synchronization of spikes with the carrier is greater for LSR than HSR fibers. The reason for the discrepancy is likely due to the different properties of the synchronization index and the present reconstruction SNR (see previous section).

Altogether, the larger number of HSR than LSR fibers together with the overall greater discharge rates of HSR fibers and the encoding contributions from off-frequency HSR fibers suggest that the population of HSR fibers is more important, and possibly sufficient, for TFS encoding.

4.4. Off-frequency contributions to ENV encoding

It is sometimes assumed that AN fibers with CFs matching the peak spectral region of a sound are the most important for ENV encoding (e.g., Prendergast et al., 2017a, 2017b; Marrufo-Pérez et al., 2018). To accommodate the wide range of sounds levels within the limited dynamic range of fiber responses, HSR, MSR and LSR fibers are thought to each encode low, medium, and high sound levels, respectively, so that each fiber type encodes changes in level (the ENV) within its best dynamic range (Joris and Yin, 1992). In contrast with this view, our results show that off-frequency fibers contribute to ENV information encoding as much as, and possibly more than, on-frequency fibers. They further show that at high levels, off-frequency HSR and MSR fibers contribute substantially to ENV encoding. These findings are in line with the

theoretical considerations of Carney (2018). They are also consistent with the analysis and modeling of human envelope following responses (Encina-Llamas et al., 2019), as well as with evidence that off-frequency information is important for AM detection (Kohlrausch et al., 2000). Off-frequency ENV encoding probably explains why behavioral AM detection thresholds improve at high levels (e.g., Kohlrausch et al., 2000; Moore and Glasberg, 2001) despite the fact that on-frequency ENV information is largest at ~40-60 dB SPL (when all fiber types contribute to ENV encoding) and less at higher levels (Fig. 5).

The present findings, however, apply to ENV encoding in quiet. In noise, or, more generally, in the presence of other simultaneous sounds, the relative contributions of on- and off-frequency fibers to ENV encoding may depend on the spectral content of the noise. For example, Fig. 11F-J shows that ENV encoding by off-frequency fibers may be reduced ('masked') to some extent for AM probes in notch noise. In addition, noise can activate the middle-ear and/or the olivocochlear reflexes, which can adapt the dynamic range of AN fibers and facilitate the encoding of ENV and TFS by on-frequency fibers over a wider range of levels (reviewed by Marrufo-Pérez and Lopez-Poveda, 2022).

4.5. Information loss versus fiber loss

We have shown that large fiber losses are required to cause substantial reductions of ENV information. For example, a complete loss of LSR fibers together with 75% loss of MSR fibers across the length of the (simulated) cochlea was required to reduce the ENV information from about 23 to 19 dB SNR for a 2-kHz carrier (red, dashed-dotted lines in Fig. 11C). For an arguably more realistic (but still extreme) synaptopathy scenario, where 50% of MSR and 66% of LSR fibers are lost (Furman et al., 2013), ENV information was only reduced from about 23 dB SNR to 21 dB SNR (red, dashed line in Fig. 11C). Overall, this is consistent with the theoretical considerations of Oxenham (2016), who suggested that large fiber losses would be required before any substantial decrease of detection-relevant information occurs, at least for long-duration sound probes. Fiber loss, however, could have larger effects on the detection and/or discrimination of brief sounds (Marmel et al., 2015).

4.6. Diagnosing synaptopathy using temporal-processing measures

The present approach assessed encoded information in an optimal manner, and findings suggest that the population of HSR fibers may be sufficient to encode the ENV (Fig. 7) and TFS (Fig. 8). If synaptopathy affects predominantly LSR fibers and/or synapses, and ENV and TFS information can be reasonably encoded without LSR fibers, this helps understanding the mixed findings when synaptopathy has been investigated in humans using non-invasive TFS-based measures, such as inter-aural phase detection, or ENV-based measures like AM detection or envelope-following responses (e.g., Grose et al., 2017; Prendergast et al., 2017a, 2017b; Guest et al., 2018; reviewed by Bramhall et al., 2019). The contribution from off-frequency HSR fibers might have also confounded the relationship between LSR fiber loss and the amplitude of tone-burst ABR wave I, another popular proxy for synaptopathy (e.g., Bramhall et al., 2017; Fulbright et al., 2017; reviewed by Bramhall et al., 2019).

We have shown that it might be possible to increase the sensitivity of ENV-based measures (and perhaps also of tone-burst ABRs) to LSR fiber loss by using high-level probe sounds embedded in notch noise (Fig. 11). This would minimize ENV encoding by both on- and off-frequency HSR fibers because on-frequency HSR fibers will be saturated due to the high probe level and off-frequency information will be masked by the notch noise.

Notch noise has long been used to reduce off-frequency encoding, thus prevent off-frequency listening, in many hearing experiments (e.g., Patterson, 1976; Johnson-Davies and Patterson, 1979), including AM detection (e.g., Almishaal et al., 2017). It has also been used to refine behavioral and electrophysiological techniques aimed at diagnosing synaptopathy (e.g., Bharadwaj et al., 2015; Carcagno and Plack, 2021). The present findings corroborate that the use of notch noise can be effective. We note, however, that our recommended levels for the AM probe (80 dB SPL) and noise (10 dB SNR) may be appropriate for our purpose but may not generalize to other stimuli or tasks. Furthermore, we could find those levels only because we knew the baseline conditions where HSR fibers contributed the most to ENV encoding for the present stimuli, and this information may not be always available. Also, it remains uncertain that the decrease in total ENV information caused by total exclusive loss of LSR fibers (Fig. 11F-J) be detectable by behavioral or electrophysiological tests, even with the proposed stimulus.

4.7. Other potential uses of the proposed information metric

The proposed method to quantify information transmission from sound to spikes may be helpful in other applications, such as, for example, the design of sound coding strategies for cochlear implants (CI). It is conceivable that the benefits of novel strategies designed to encode TFS information be obscured if those strategies simultaneously degrade ENV encoding. The present approach may be used to control that ENV information is preserved while improving TFS encoding. In this example application, information could be assessed in the pattern of electrical current pulses generated by the CI (the 'electrodogram') or in the pattern of neural responses evoked by the CI (the 'neurogram'). Depending on the choice, the setup would involve reconstructing the stimulus ENV and TFS from the electrodogram or from a neurogram simulated with a computational model of the AN response to electrical stimulation (see Leclère et al., submitted).

5. Conclusions

1 The population of high-spontaneous rate fibers dominates the encoding of temporal fine structure across all sound levels both because they are more abundant and because they have higher discharge rates than medium- or low-spontaneous rate fibers.
2 Envelope is poorly encoded by high-spontaneous rate fibers at high levels probably because their larger onset response in each modulation cycle distorts the neural representation of the envelope, rather than because of saturation *per se*.
3 Measures based on temporal fine structure encoding are likely insensitive to a loss of low-spontaneous rate fibers.
4 Off-frequency high-spontaneous rate fibers contribute significantly to envelope encoding when on-frequency fibers saturate at and above moderate sound levels.
5 Envelope-based measures aimed at diagnosing a loss of low-spontaneous rate fibers can improve their sensitivity by using a high-level probe embedded in a notch noise to mask off-frequency envelope encoding, particularly by fibers with high-spontaneous rates.

Author statement

Conceptualization: EALP; Methodology: EALP, PTJ; Software: PTJ; Formal analysis: PTJ; Investigation: PTJ; Writing-Original Draft: PTJ, EALP; Writing-Review & Editing: All authors. Visualization: PTJ, EALP. Supervision: AW, MSM, EALP. Project administration: EALP. Funding acquisition: MSM, EALP.

Conflicts of interests

The authors declare no competing financial interests. AW and MSM are employees of Oticon Medical.

Acknowledgements

We thank Christian J. Sumner and Almudena Eustaquio-Martín for sharing their code for the auditory model, and Attila Fráter for useful discussions. Work supported by Oticon Medical, the William Demant Foundation, Junta de Castilla y León (grant SA0252P20), European Regional Development Fund, and Spanish Ministry of Science and Innovation (grant PID2019-108985GB-I00) to EALP.

Appendix

The reconstruction procedure was almost identical to that of Warland et al. (1997) and is reproduced below for convenience. The procedure consists of filtering a spike train (r) through a reconstruction filter (f) to obtain an estimate (u) as close as possible to a reference signal (s) by optimizing the reconstruction filter. Note that, here, the term spike train refers to the sum of spike trains from all fibers of the same type innervating an IHC. In other words, r is a set of spike counts with integer values at each given sample. Also note that reconstruction may be done using one summed spike train and a single reconstruction filter, or using several summed spike trains each with their corresponding reconstruction filter. In other words, when multiple summed spike trains are used for reconstruction, each of them is convolved with its corresponding reconstruction filter.

An estimate, u, of the reference signal is calculated as a constant term, a, plus the convolution of each summed spike train, r^v, with its corresponding linear reconstruction filter, f^v, and added together across the used spike trains (Eq. 3 of Warland et al., 1997):

$$u_i = a + \sum_v \sum_{j=1}^{N} r^v_{i-j} f^v_j \tag{A1}$$

where u_i is the reconstructed estimate for the i-th sample number, v indicates summed spike-train number (going from 1 up to a maximum of 48 = 16 CFs × 3 fiber types per CF), f^v_j is the filter coefficient j samples before i, and N is the length of the reconstruction filters (in number of samples). In other words, each summed spike train produces an additive component to the reconstructed stimulus.

To facilitate the calculation of the reconstruction filter coefficients, we introduce a matrix notation (bold font) and define the stimulus vector **s,** the reconstruction **u** and the response matrix **R** as follows (Eq. 4 of Warland et al., 1997):

$$\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ \cdots \\ s_M \end{bmatrix}, \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \cdots \\ u_M \end{bmatrix} \tag{A2}$$

$$\mathbf{R} = \begin{bmatrix} 1 & r^1_1 & r^1_2 & \cdots r^1_N & \cdots & r^v_1 & r^v_2 \cdots r^v_N \\ 1 & r^1_2 & r^1_3 & \cdots r^1_{N+1} & \cdots & r^v_2 & r^v_3 \cdots r^v_{N+1} \\ 1 & r^1_3 & r^1_4 & \cdots r^1_{N+2} & \cdots & r^v_3 & r^v_4 \cdots r^v_{N+2} \\ 1 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & r^1_M & r^1_{M+1} & \cdots r^1_{N+M} & \cdots & r^v_M & r^v_{M+1} \cdots r^v_{N+M} \end{bmatrix}$$

The first row of **R** consists of a one followed by the first spike train ($v = 1$) for sample indices $i = 1$ to N (r^1_1 to r^1_N), followed by the second spike train ($v = 2$) for sample indices $i = 1$ to N (r^2_1 to r^2_N), and so on until the v-th spike train. The second row in **R** contains the same as row one but for one time instant later ($i = 2$ to $N + 1$). Following rows are for successively later time instants until time instant M, where M is the number of samples in the reference signal. The required spike train duration is $N + M - 1$ samples.

In matrix notation, the reconstructed estimate of the reference signal is calculated as (Eq. 6 of Warland et al., 1997):

$$\mathbf{u} = \mathbf{Rf} \tag{A3}$$

where the transposed reconstruction filter **f** is defined as (Eq. 7 of Warland et al., 1997):

$$\mathbf{f^T} = \begin{bmatrix} a & f^1_1 & f^1_2 & \cdots f^1_N & \cdots & f^v_1 & f^v_2 & \cdots f^v_N \end{bmatrix} \tag{A4}$$

Note that **f** is a concatenation of the individual reconstruction filters for each summed spike train, as indicated by the superscript v. The optimal filter **f** is defined as the one that minimizes the squared difference $\mathbf{e}^2$ between the reference signal, **s**, and the reconstructed estimate, **u**:

$$\mathbf{e}^2 = (\mathbf{s} - \mathbf{u})^{\mathrm{T}} (\mathbf{s} - \mathbf{u}) \tag{A5}$$

The optimal reconstruction filter that minimizes $\mathbf{e}^2$ is obtained from the analytical expression (Eq. 8 of Warland et al., 1997):

$$\mathbf{f} = \left(\mathbf{R}^{\mathrm{T}}\mathbf{R}\right)^{-1} \left(\mathbf{R}^{\mathrm{T}}\mathbf{s}\right) \tag{A6}$$

The term $(\mathbf{R}^{\mathrm{T}}\mathbf{R})$ corresponds to correlation functions among the spike trains and the term $(\mathbf{R}^{\mathrm{T}}\mathbf{s})$ corresponds to the reverse correlation between the stimulus and the spike trains and hence the optimal filter is determined by the statistics of both the stimulus and the spikes.

There are two minor differences between the procedure described by Warland et al. (1997) and the present procedure. First, here the reference signal was *not* normalized to zero mean and unit variance before the reconstruction procedure because, unlike Warland et al. (1997), we also reconstructed envelope information which contains a constant component (DC) and hence should not have the DC removed. Second, except when specifically indicated, we did not downsample the spike trains that were input to the reconstruction algorithm. This has no consequence as the mathematics governing the procedure are equally valid and independent of sampling rate.

References

Almishaal, A., Bidelman, G.M., Jennings, S.G., 2017. Notched-noise precursors improve detection of low-frequency amplitude modulations. J. Acoust. Soc. Am. 141, 324–333.

Bharadwaj, H.M., Masud, S., Mehraei, G., Verhulst, S., Shinn-Cunningham, B.G., 2015. Individual differences reveal correlates of hidden hearing deficits. J. Neurosci. 35, 2161–2172.

Bialek, W., Rieke, F., de Ruyter van Steveninck, R.R., Warland, D., 1991. Reading a neural code. Science 252, 1854–1857.

Bramhall, N., Beach, E.F., Epp, B., Le Prell, C.G., Lopez-Poveda, E.A., Plack, C.J., Schaette, R., Verhulst, S., Canlon, B, 2019. The search for noise-induced cochlear synaptopathy in humans: mission impossible? Hear. Res. 377, 88–103. http://doi.org/10.1016/j.heares.2019.02.016.

Bramhall, N.F., Konrad-Martin, D., McMillan, G.P., Griest, S.E., 2017. Auditory brainstem response altered in humans with noise exposure despite normal outer hair cell function. Ear Hear. 38, e1–e12.

Carcagno, S., Plack, C.J., 2021. Effects of age on psychophysical measures of auditory temporal processing and speech reception at low and high levels. Hear. Res. 400, 108117. http://doi.org/10.1016/j.heares.2020.108117.

Carney, L.H., 2018. Supra-threshold hearing and fluctuation profiles: implications for sensorineural and hidden hearing loss. J. Assoc. Res. Otolaryngol. 19, 331–352. http://doi.org/10.1007/s10162-018-0669-5.

Encina-Llamas, G., Harte, J.M., Dau, T., Shinn-Cunningham, B., Epp, B., 2019. Investigating the effect of cochlear synaptopathy on envelope following responses using a model of the auditory nerve. J. Assoc. Res. Otolaryngol. 20, 363–382.

Fulbright, A.N.C., Le Prell, C.G., Griffiths, S.K., Lobarinas, E, 2017. Effects of recreational noise on threshold and suprathreshold measures of auditory function. Semin. Hear. 38, 298–318.

Furman, A.C., Kujawa, S.G., Liberman, M.C., 2013. Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates. J. Neurophysiol. 110, 577–586.

Gabbiani, F., Metzner, W., 1999. Encoding and processing of sensory information in neural spike trains. J. Exp. Biol. 202, 1267–1279.

Goldberg, J.M., Brown, P.B., 1969. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. J. Neurophysiol. 22, 613–636.

Grose, J.H., Buss, E., Hall III, J.W, 2017. Loud music exposure and cochlear synaptopathy in young adults: isolated auditory brainstem response effects but no perceptual consequences. Trends Hear. 21, 1–18.
Guest, H., Munro, K.J., Prendergast, G., Millman, R.E., Plack, C.J., 2018. Impaired speech perception in noise with a normal audiogram: no evidence for cochlear synaptopathy and no relation to lifetime noise exposure. Hear. Res. 364, 142–151. http://doi.org/10.1016/j.heares.2018.03.008.
Holmes, S.D., Sumner, C.J., O'Mard, L.P., Meddis, R, 2004. The temporal representation of speech in a nonlinear model of the guinea pig cochlea. J. Acoust. Soc. Am. 116, 3534–3545.
Johannesen, P.T., Buzo, B.C., Lopez-Poveda, E.A., 2019. Evidence for age-related cochlear synaptopathy in humans unconnected to speech-in-noise intelligibility deficits. Hear. Res. 374, 35–48. http://doi.org/10.1016/j.heares.2019.01.017.
Johnson, D.H., 1980. The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. J. Acoust. Soc. Am. 68, 1115–1122.
Johnson-Davies, D., Patterson, R.D, 1979. Psychophysical tuning curves: restricting the listening band to the signal region. J. Acoust. Soc. Am. 65, 765–770.
Joris, P.X., Louage, D.H., Cardoen, L., van der Heijden, M., 2006. Correlation index: a new metric to quantify temporal coding. Hear. Res. 216-217, 19–30.
Joris, P.X., Yin, T.C.T., 1992. Responses to amplitude-modulated tones in the auditory nerve of the cat. J. Acoust. Soc. Am. 91, 215–232.
Joris, P.X., Schreiner, C.E., Rees, A, 2004. Neural processing of amplitude-modulated sounds. Physiol. Rev. 84, 541–577.
Kale, S., Heinz, M.G., 2012. Temporal modulation transfer functions measured from auditory-nerve responses following sensorineural hearing loss. Hear. Res. 286, 64–75. http://doi.org/10.1016/j.heares.2012.02.004.
Kohlrausch, A., Fassel, R., Dau, T., 2000. The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. J. Acoust. Soc. Am. 108, 723–734. http://doi.org/10.1121/1.429605.
Kujawa, S.G., Liberman, M.C., 2009. Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. J. Neurosci. 29, 14077–14085.
Liberman, M.C., 1978. Auditory-nerve response from cats raised in a low-noise chamber. J. Acoust. Soc. Am. 63, 442–455. doi:10.1121/1.381736.
Lopez-Poveda, E.A., 2005. Spectral processing by the peripheral auditory system: facts and models. Int. Rev. Neurobiol. 70, 7–48. doi:10.1016/S0074-7742(05)70001-5.
Lopez-Poveda, E.A., Barrios, P, 2013. Perception of stochastically undersampled sound waveforms: a model of auditory deafferentation. Front. Neurosci. 7, 124.
Lopez-Poveda, E.A., Meddis, R, 2001. A human nonlinear cochlear filterbank. J. Acoust. Soc. Am. 110, 3107–3118. http://doi.org/10.1121/1.1416197.
Lopez-Poveda, E.A., 2014. Why do I hear but not understand? Stochastic undersampling as a model of degraded neural encoding of speech. Front. Neurosci. 8, 348.
Lopez-Poveda, E.A., 2018. Olivocochlear efferents in animals and humans: from anatomy to clinical relevance. Front. Neurol. 9, 197. doi:10.3389/fneur.2018.00197.
Lopez-Poveda, E.A., 2003. An approximate transfer function for the dual-resonance nonlinear filter model of auditory frequency selectivity. J. Acoust. Soc. Am. 114, 2112–2117.
Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., Moore, B.C.J., 2006. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proc. Nat. Acad. Sci. 103, 18866–18869.
Louage, D.G., van der Heijden, M., Joris, P.X., 2004. Temporal properties of responses to broadband noise in the auditory nerve. J. Neurophysiol. 91, 2051–2065. http://doi.org/10.1152/jn.00816.2003.
Makary, C.A., Shin, J., Kujawa, S.G., Liberman, M.C., Merchant, S.N., 2011. Age-related primary cochlear neuronal degeneration in human temporal bones. J. Assoc. Res. Otolaryngol. 12, 711–717.
Marmel, F., Rodríguez-Mendoza, M.A., Lopez-Poveda, E.A., 2015. Stochastic undersampling steepens auditory threshold/duration functions: implications for understanding auditory deafferentation and aging. Front. Aging. Neurosci. 7, 63. http://doi.org/10.3389/fnagi.2015.00063.
Marrufo-Pérez, M.I., Eustaquio-Martín, A., Fumero, M.J., Gorospe, J.M., Polo, R., Gutiérrez Revilla, A., Lopez-Poveda, E.A, 2019. Adaptation to noise in amplitude modulation detection for cochlear implant users. Hear. Res. 377, 133–141. http://doi.org/10.1016/j.heares.2019.03.017.
Marrufo-Pérez, M.I., Eustaquio-Martín, A., López-Bascuas, L.E., Lopez-Poveda, E.A., 2018. Temporal effects on monaural amplitude-modulation sensitivity in ipsilateral, contralateral and bilateral noise. J. Assoc. Res. Otolaryngol. 19, 147–161. http://doi.org/10.1007/s10162-018-0656-x.
Marrufo-Pérez, M.I., Lopez-Poveda, E.A., 2022. Adaptation to noise in normal and impaired hearing. J. Acoust. Soc. Am. 151, 1741–1753. doi:10.1121/10.0009802.
Meddis, R., Lopez-Poveda, E.A., 2010. Auditory periphery: from pinna to auditory nerve. In: Meddis, R, Lopez-Poveda, EA, Popper, AN, Fay, RR (Eds.), Computational Models of the Auditory System. Springer, New York, pp. 7–38 Chapter 2.
Meddis, R., O'Mard, L.P.O., Lopez-Poveda, E.A., 2001. A computational algorithm for computing non-linear auditory frequency selectivity. J. Acoust. Soc. Am. 109, 2852–2861. http://doi.org/10.1121/1.1370357.
Meddis, R., 1986. Simulation of mechanical to neural transduction in the auditory receptor. J. Acoust. Soc. Am. 79, 702–711.
Mesgarani, N., David, S.V., Fritz, J.B, Shamma, S.A, 2008. Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. J. Neurophysiol. 102, 3329–3339.
Moore, B.C.J., Glasberg, B.R., 2001. Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners. J. Acoust. Soc. Am. 110, 1067–1073. http://doi.org/10.1121/1.1385177.
Müller, M., Robertson, D., 1991. Relationship between tone burst discharge pattern and spontaneous firing rate of auditory nerve fibers in the guinea-pig. Hear. Res. 57, 63–70.
Osses Vecchi, A., Varnet, L., Carney, L.H., Dau, T., Bruce, I.C, 2022. A comparative study of eight human auditory models of monaural processing. Acta Acust. 6, 17.
Oxenham, A.J., 2016. Predicting the perceptual consequences of hidden hearing loss. Trends Hear. 20, 1–6.
Palmer, A.R., Russell, I.J., 1986. Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. Hear. Res. 24, 1–15. http://doi.org/10.1016/0378-5955(86)90002-x.
Pasley, B.N., David, S.V., Mesgarani, N., Flinker, A., Shamma, S.A., Crone, N.E., Knight, R.T., Chang, E.F., 2012. Reconstructing speech from human auditory cortex. PLoS Biol. 10 (1), e1001251. http://doi.org/10.1371/journal.pbio.1001251.
Patterson, R.D., 1976. Auditory filter shapes derived with noise stimuli. J. Acoust. Soc. Am. 59, 640–654.
Plack, C.J., Barker, D., Prendergast, G., 2014. Perceptual consequences of "hidden" hearing loss. Trends Hear. 18, 1–11.
Prendergast, G., Guest, H., Munro, K.J., Kluk, K., Léger, A., Hall, D.A., Heinz, M.G., Plack, C.J., 2017a. Effects of noise exposure on young adults with normal audiograms I: electrophysiology. Hear Res 344, 68–81.
Prendergast, G., Millman, R.E., Guest, H., Munro, K.J., Kluk, K., Dewey, R.S., Hall, D.A., Heinz, M.G., Plack, C.J., 2017b. Effects of noise exposure on young adults with normal audiograms II: behavioral measures. Hear Res 356, 74–86.
Rieke, F., Warland, D., de Ruyter van Steveninck, R.R., Bialek, W., 1997. Spikes: Exploring the Neural Code. MIT Press, Cambridge, MA.
Rosen, S., 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. Philos. Trans. R. Soc. Lond. B, Biol. Sci. 336, 367–373.
Schmiedt, R.A., Mills, J.H., Boettcher, F.A., 1996. Age-related loss of activity of auditory-nerve fibers. J. Neurophysiol. 76, 2799–2803. doi:10.1152/jn.1996.76.4.2799.
Sergeyenko, Y., Lall, K., Liberman, M.C., Kujawa, S.G., 2013. Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline. J. Neurosci. 33, 13686–13694.
Shamma, S.A., Chadwick, R.S., Wilbur, W.J., Morrish, K.A., Rinzel, J., 1986. A biophysical model of cochlear processing: intensity dependence of pure tone responses. J. Acoust. Soc. Am. 80, 133–145. http://doi.org/10.1121/1.394173.
Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304. http://doi.org/10.1126/science.270.5234.303. PMID: 7569981.
Smith, ZM, Delgutte, B, Oxenham, AJ, 2002. Chimaeric sounds reveal dichotomies in auditory perception. Nature 416, 87–90. doi:10.1038/416087a.
Sumner, C., Lopez-Poveda, E.A., O'Mard, L.P.O., Meddis, R, 2002. A revised model of the inner hair cell and auditory nerve complex. J. Acoust. Soc. Am. 111, 2178–2188. http://doi.org/10.1121/1.1453451.
Sumner, C., Lopez-Poveda, E.A., O'Mard, L.P.O., Meddis, R, 2003b. Adaptation in a revised model of the inner-hair cell. J Acoust Soc Am 113, 893–901. http://doi.org/10.1121/1.1515777.
Sumner, C., O'Mard, L.P.O., Lopez-Poveda, E.A., Meddis, R., 2003a. A non-linear filterbank model of the guinea-pig cochlear nerve: rate responses. J Acoust Soc Am 113, 3264–3274. http://doi.org/10.1121/1.1568946.
Suthakar, K., Liberman, M.C., 2021. Auditory-nerve responses in mice with noise-induced cochlear synaptopathy. J. Neurophysiol. 126, 2027–2038. http://doi.org/10.1152/jn.00342.2021.
Van Tassel, D.J., Soli, S.D., Kirby, V.M., Widin, G.P., 1987. Speech waveform envelope cues for consonant recognition. J. Acoust. Soc. Am. 82, 1152–1161.
Viana, L.M., O'Malley, J.T., Burgess, B.J., Jones, D.D., Oliveira, C.A., Santos, F., Merchant, S.N., Liberman, L.D., Liberman, M.C., 2015. Cochlear neuropathy in human presbycusis: confocal analysis of hidden hearing loss in post-mortem tissue. Hear. Res. 327, 78–88.
Warland, D.K., Reinagel, P., Meister, M., 1997. Decoding visual information from a population of retinal ganglion cells. J. Neurophysiol. 78, 2336–2350.
Wessel, R., Koch, C., Gabbiani, F., 1996. Coding of time-varying electric field amplitude modulations in a wave-type electric fish. J. Neurophys. 75, 2280–2293.
Wever, E.G., 1949. Theory of Hearing. Wiley, New York, NY.
Winslow, R.L., Sachs, M.B., 1987. Effect of electrical stimulation of the crossed olivocochlear bundle on auditory nerve response to tones in noise. J. Neurophysiol. 57, 1002–1021.
Winter, I.M., Robertson, D., Yates, G.K., 1990. Diversity of characteristic frequency rate-intensity functions in guinea pig auditory nerve fibers. Hear. Res. 45, 191–202.
Wu, P.Z., Liberman, L.D., Bennett, K., de, Gruttola V, O'Malley, J.T.O., Liberman, M.C, 2019. Primary neural degeneration in the human cochlea: evidence for hidden hearing loss in the aging ear. Neuroscience 407, 8–20. doi:10.1016/j.neuroscience.2018.07.053.
Zai, A.T., Bhargava, S., Mesgarani, N., Liu, S.-.C., 2015. Reconstruction of audio waveforms from spike trains of artificial cochlea models. Front Neurosci. 9, 347. http://doi.org/10.3389/fnins.2015.00347.