Dynamic Reweighting of Auditory Modulation Filters

Eva R. M. Joosten[1], Shihab A. Shamma[2,3], Christian Lorenzi[2], Peter Neri[2]*

1 Laboratoire Psychologie de la Perception (CNRS UMR 8242) and Université Paris Descartes, Sorbonne Paris Cité, Paris, France, **2** Laboratoire des Systèmes Perceptifs (CNRS UMR 8248) and Département d'études cognitives, Ecole Normale Supérieure, PSL Research University, Paris, France, **3** Department of Electrical and Computer Engineering, Institute for Systems Research, University of Maryland, College Park, Maryland, United States of America

* neri.peter@gmail.com

Abstract

Sound waveforms convey information largely via amplitude modulations (AM). A large body of experimental evidence has provided support for a modulation (bandpass) filterbank. Details of this model have varied over time partly reflecting different experimental conditions and diverse datasets from distinct task strategies, contributing uncertainty to the bandwidth measurements and leaving important issues unresolved. We adopt here a solely data-driven measurement approach in which we first demonstrate how different models can be subsumed within a common 'cascade' framework, and then proceed to characterize the cascade via system identification analysis using a single stimulus/task specification and hence stable task rules largely unconstrained by any model or parameters. Observers were required to detect a brief change in level superimposed onto random level changes that served as AM noise; the relationship between trial-by-trial noisy fluctuations and corresponding human responses enables targeted identification of distinct cascade elements. The resulting measurements exhibit a dynamic complex picture in which human perception of auditory modulations appears adaptive in nature, evolving from an initial lowpass to bandpass modes (with broad tuning, $Q \sim 1$) following repeated stimulus exposure.

Author Summary

Amplitude modulations are considered the key carriers of intelligible information in auditory signals, and consequently it is of significant interest to discover how they are neurally analyzed and perceptually encoded. A dominant model has emerged from extensive experimental and theoretical studies of this phenomenon. This model posits that amplitude modulations are parsed into channels of different temporal rates via a bank of bandpass filters. Using exclusively data driven approaches with minimal assumptions about the structure of the model, the picture that emerges is of an adaptive process. Initially, human listeners in these tasks perceive modulations as if through a lowpass filter with very low cutoff frequency, which gradually evolves to become a broadly tuned bandpass process at higher modulation frequencies, reflecting the modulations of the target stimuli. This

surprising dynamic characteristic emphasizes the plastic nature of modulation analysis in sensory perception.

Introduction

Natural sounds carry salient amplitude modulations (AM) essential for successful interpretation of auditory landscapes and robust source identification [1–5]. The human auditory system is exquisitely sensitive to relatively slow AM cues, prompting extensive investigation of this ability over several decades [6, 7]. Recent success in modeling recognition of environmental sound textures [8], speech [9–11] and music material [12–14] has now prompted the need to consolidate the exact nature of modulation filtering [15–18].

Early measurements of sensitivity across AM rates seemed consistent with low-pass characteristics (running average over time) as envisaged by the 'linear envelope detector' (LED) model [6, 19, 20]. However, this class of experiments often posed interpretational challenges. For example, under some stimulus and experimental conditions only parts of the dataset conformed to low-pass characteristics [6, 15, 21–23] but not others [16, 24, 25]. Plausible accounts of these apparent inconsistencies involved shifts in task strategies and/or decision statistics [15, 21], whereby listeners would rely on different cues under conditions associated with different portions of the dataset and/or subtle stimulus alterations [26–28]. Strategy shifts of this kind are relatively common [29] (e.g. off-frequency listening [30]). Therefore, our first goal was to assess the extent to which *endogenous* adaptive strategies can influence the measurements of the modulation filter when no stimulus and/or task-related factors are concomitantly modified. Specifically, we sought to adopt a single stable measurement approach throughout (the *reverse correlation* method), reasoning that systematic changes in the subjects' detection strategies would be unveiled as a dynamic evolution of the filter estimates during the experimental session [31–33].

The shortcomings of the low-pass AM filter model have since spurred development of alternative models which augmented it with, for example, an autocorrelation stage [34], or the current dominant framework of the modulation filterbank (MFB) [16, 35]. This latter model consists of a low-pass filter (up to 2.5 Hz) and a bank of band-pass filters spanning the ~ 5–100 Hz range. Thanks to its increased sophistication and flexibility, the MFB is able to account for a wide range of psychophysical [16, 24, 36–39] and physiological [35, 40–42] results. Nevertheless, there remain several important details of these hypothesized filters that have not been adequately constrained by available data. For instance, filter tuning (Q-value) has often been assumed to measure ~ 2 [35], yet other studies have indicated a significantly lower value (<1) [25]. Furthermore, another study set out to test the ability of the modulation filter bank model to account for dependence of AM envelope detection on the nature of the carrier. It reported that the model significantly underestimated the detrimental effects of carrier periodicity on modulation detection, a finding that has cast a shadow of uncertainty on certain aspects of the filterbank model [34]. Another unexplained finding concerns the lack of persistent low-level adaptive effects in these postulated filters [18], suggesting that they are more dynamic and likely more susceptible to cognitive control (e.g. by the details of the measurement task [43]). With specific relation to psychoacoustical literature using decision weights [44] (a technique related to the reverse correlation approach used here), the bandpass signatures expected of the MFB model are not directly evident through previous filter estimates [45–48] (we return to this issue in Discussion). Consequently, the second goal of the present study is to measure these bandpass filters in an unconstrained manner to allow for data-driven conclusions that are

largely independent of model specifics. To do so, we relied on a combination of powerful system identification tools and AM-tailored stimulus perturbations that enabled us to describe a more dynamic picture of the underlying process encompassing both LED and MFB modes of operation.

Our findings largely confirm the three critical questions we sought to address: (1) the measured bandpass channels undergo characteristic changes to reflect listeners' strategy shifts from using a combination of loudness and spectral-profile cues (both existing in the presented stimuli) in the early part of the experiment, to relying primarily on the spectral-profile cue during later phases; (2) the Q-values associated with these channels are ~1, although this characteristic is itself subject to recalibration as assessed by our measurement task and analysis; and (3) the modulation filter bank (augmented by a low-pass characteristic with a low cut-off [35]) remains the most parsimonious model for auditory modulation processing.

Materials and Methods

Ethics statement

Ethics approval was obtained from the College Ethics Review Board (CERB) at Aberdeen University (http://www.abdn.ac.uk/clsm/working-here/cerb.php). All listeners gave written informed consent.

Stimulus and task

The auditory stimuli (delivered binaurally via Sennheiser HD202 headphones) were specifically designed to encode task-relevant AM modulations into the envelope of the acoustic signal in a manner suitable for experimental characterization using psychophysical reverse correlation [49], and their parameters were specified to lie in the perceptually salient range for AM processing (<30 Hz). The 300-ms carrier was a segment of white noise (5 kHz bandwidth, 10 kHz sampling rate) that was generated once at the beginning of each block and identically replicated throughout the entire block; on any given block, no variability/perturbation was therefore introduced by the fine temporal structure of the stimulus (see further below and S6 Fig for additional analyses demonstrating that carrier-induced AM deviations had no impact on our results). Stimulus perturbation was instead applied to the overall amplitude of the carrier in a stepwise fashion: the stimulus waveform was subdivided into 9 temporal segments (each lasting ~30-ms) and the amplitude of each segment was controlled independently [46]. In the absence of an applied increment/decrement (signal), a given segment was assigned a fixed baseline level (indicated by leading and trailing ends of red line in Fig 1A) of ~62 dB SPL. The 'increment' signal was generated by increasing the amplitude of the central segment within the stimulus waveform (peak of red line in Fig 1A) to ~68 dB SPL (we use ~ because the exact values were tailored to each listener to target individual threshold performance and therefore differed across listeners); the temporal scale of this modulation was chosen to match ecologically relevant AM cues [1, 50], and its AM pulse-like specification was chosen to ease application of analytical/theoretical results dependent on signal shape [51]. The target signal was then added to a noise waveform generated by applying a random Gaussian-distributed increment/decrement to the overall amplitude of each segment, independently for different segments (Fig 1C); the resulting signal+noise trace (Fig 1E) was presented together with a noise-only trace on every trial, and listeners were asked to detect the former in a two-interval-forced-choice (2IFC) protocol. The jitter introduced by the noisy modulation was approximately ±2.2 dB standard deviation around baseline level. In a separate series of experiments we asked the same listeners to detect a signal defined by a decrement (rather than an
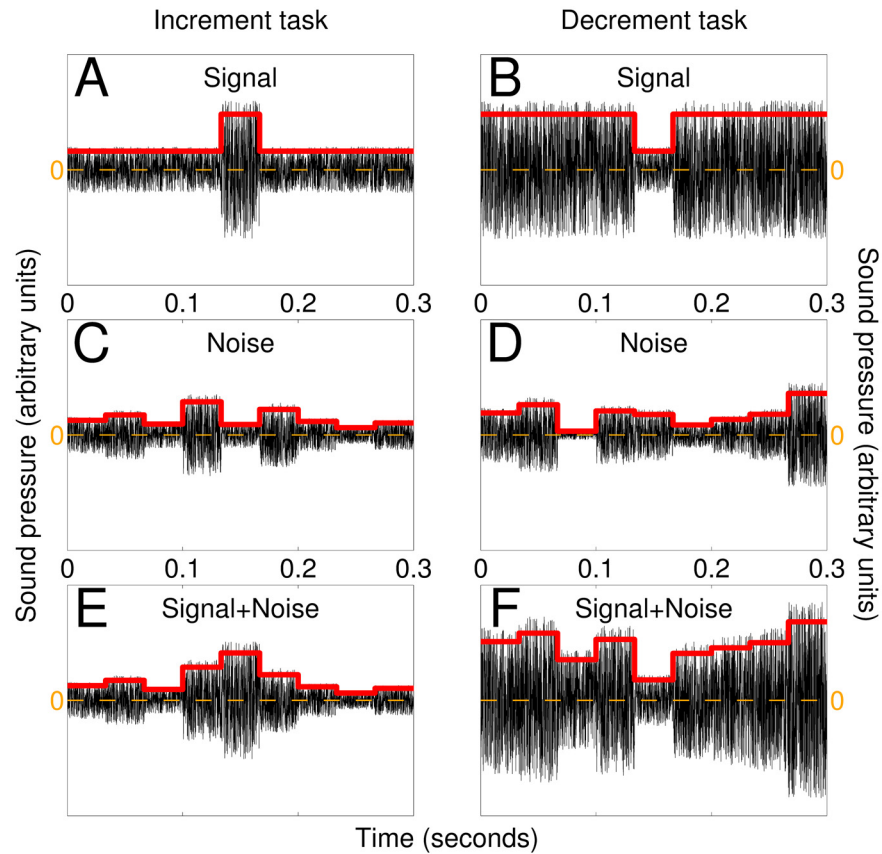
Fig 1. AM modulations as targets and interferers. The signal to be detected consisted of a centered 33-ms square-pulse change in AM envelope that was either larger (A) or smaller (B) than baseline modulation depending on whether listeners were engaged in the increment or decrement task respectively. (C,D) Noise was applied in the form of Gaussian modulations around baseline value every 33-ms to generate signal +noise stimuli (E-F). Black traces show sound waves, red traces show amplitude modulations. See Materials and Methods for details.

increment) in the amplitude of the central segment (Fig 1B) from a baseline level of ∼68 dB to ∼62 dB. Some previous studies on AM minimized the role of loudness cues by roving the absolute levels of individual stimuli [44, 52]; we deliberately avoided this manipulation because it amplifies the role of gain control [53], a nonlinear effect that may have compromised interpretation of the nonlinear kernels [51, 54]: roving stimuli span a wide level range for the purpose of rendering overall absolute level ineffective as a cue for performing discrimination, however they also require listeners to *factor out* overall level via gain control (or analogous mechanisms) to compute relative filter outputs; this dynamic nonlinear operation is difficult to incorporate into the cascade models that form the theoretical underpinnings of the present study [51, 58], potentially complicating kernel analysis beyond the level of interpretability afforded by analyses like those favoured here that do not involve parameter fitting. Following their response (via button press), auditory feedback (correct/incorrect) was provided immediately; the next trial automatically initiated after a 1-second delay. At the end of each block, listeners received an auditory summary (via an automated system based on segments of human voice) detailing the total number of collected trials and the percentage of correct responses on the last block as well as averaged across all blocks.

Targeting optimal performance regimes

We tested 10 listeners, all naive except author EJ (indicated by square symbol in all Figures), with ages 28±9 years (mean±SD across listeners). Listeners were initially familiarized with the task during preliminary training sessions consisting of ∼ 20 trials without noise. They were presented with noiseless versions of both signal+noise and noise-only stimuli, preceded by explicit verbal indication of whether they would hear the former or the latter. They were then asked to indicate the signal+noise interval and provided with trial-by-trial feedback (correct/incorrect). After this preliminary phase, we adjusted noise levels individually to target optimal threshold regimes for the deployment of reverse correlation [55] (S1 Fig). Percent correct was 74%±6% (increment detection, 5.5k±2.8k trials per listener) and 76%±7% (decrement detection, 3.6k±3.3k trials per listener). We successfully minimized response bias across listeners (S1B Fig); this is particularly important when attempting nonlinear system identification to avoid bias-induced modulations within nonlinear kernels: in the presence of bias, 2nd order kernels (see below for detailed description/definition) may contain modulations that do not reflect the structure of the perceptual machinery preceding the binary behavioural decision, but are instead produced by the nonlinear nature of the conversion from decision variable to binary output which, in general, is not relevant for characterizing the perceptual machinery itself as it is protocol-specific [51] (i.e. it depends on the arbitrary way in which listeners are asked to express their percept, whether via binary choice ('yes'/'no', 'present'/'absent') or rating scale, for example). Internal noise was within the expected range for human psychophysics [56] (S1C Fig), indicating that listeners adopted a robust task strategy, as also evidenced by the relatively high absolute efficiency (within the normal range for detection [57] and much higher than observed for other auditory tasks [58], see S1D Fig). All aspects of performance metric analysis indicate that 1) listeners performed the above-detailed tasks in a stable and efficient manner, with no discernible difference in overall performance metrics between increment and decrement experiments (data points fall around diagonal unity line in S1 Fig); 2) our protocols successfully established optimal conditions for the application of behavioural reverse correlation [49, 51, 56].

LED/MFB models recast as LNL cascades

The modulation filtering models are referred to as LED/MFB models. They are normally implemented as illustrated in Fig 2. For the purpose of examining how these models relate to our protocols, we treat the incoming stimulus as defined in AM space (i.e. as a 9-element vector where individual entries indicate the amplitude of each segment) because this is the stimulus subspace within which input noise was applied (see Fig 1); in other words, cochlear filtering (top in Fig 2) is reduced by our stimulus generation protocols, because noise is applied not to the fine structure of the stimulus waveform but to its AM profile. The LED involves application of a low-pass filter [6] to the (modulation) envelope, while the MFB applies primarily bandpass filtering [16]. This front-end stage is illustrated in Fig 2A and 2B for LED and MFB respectively. With reference to the 9-element input vector where each entry refers to a different time point, this filtering stage consists of convolution between this input vector and a temporal impulse response $\mathbf{L}_1$. We represent the filter frequency characteristics in Fig 2 (blue) because they are easier to interpret as lowpass/bandpass, but our $\mathbf{L}_1$ estimates are initially recovered as temporal impulse responses (Fig 3B) because they are obtained via direct reverse-correlation of the input stimulus (defined across time). The output from the $\mathbf{L}_1$ layer (which is itself a function of time) is then passed onto a decision stage that generates a psychophysical response (red rectangle in Fig 2). The details of how this stage operates are still unclear [15, 21]. We can describe the sequence of operations carried out by LED/MFB models using the same general
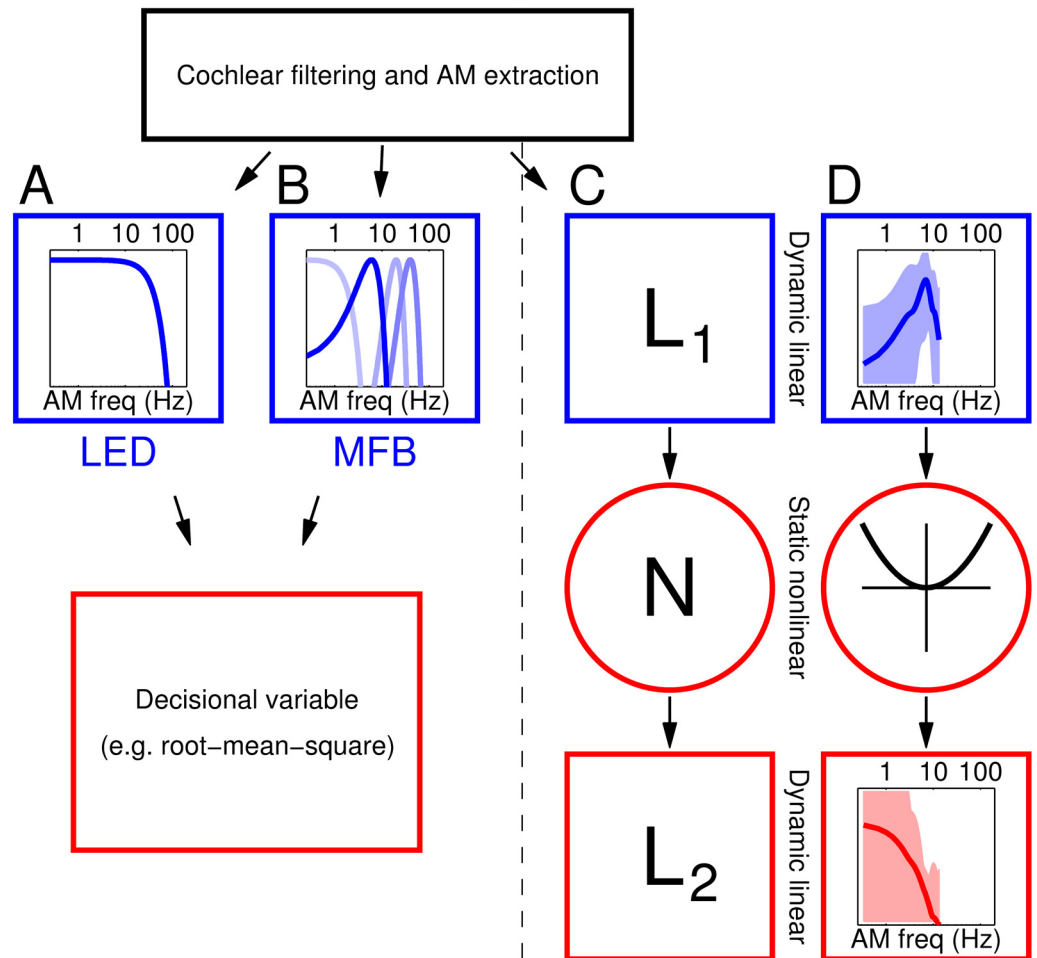
Fig 2. Models of AM processing in the auditory system. All such models consist of a cochlear filtering and AM envelope extraction stage (top), followed by one of two competing formulations: (A) a linear lowpass filter that smooths the AM envelope, better known as the linear envelope detector (LED) model; or (B) a modulation frequency bank (MFB) model that bandpass filters the AM rates into different channels. Both models feed into a decisional stage (red). (C) They can be reformulated as a cascade of linear-nonlinear-linear (L_1NL_2) stages, with the additional combined nonlinear-linear (NL_2) operation accommodating different decisional rules, e.g. squaring followed by sum in the case of energy-based rules. When formulated as such, the two standard models in (A-B) are primarily distinguished by the nature of their L_1 modules (blue): lowpass in one and bank of bandpass filters in the other. (D) Our cumulative empirical estimates of L_1 and L_2 support MFB characteristics (blue trace in D is compatible with blue traces in B, not A).

doi:10.1371/journal.pcbi.1005019.g002

cascade, as illustrated in Fig 2C. In this formulation, the LED/MFB filters correspond to different characteristics for the first filtering stage L_1. Subsequent decision stages are approximated by a combined nonlinear-linear operation (red outlines in Fig 2C). For example, if read-out involves energy extraction from the temporal output returned by L_1, N corresponds to squaring and L_2 to sum over time. Similar approximations can be adopted for root-mean-square and MAX rules [51, 58]. The two filters L_1 and L_2 are referred to as the *cascade filters*. Our goal is to estimate their structure via the *psychophysical kernels* we can measure from data (see below); this is achieved by exploiting a set of analytical tools that establish important connections between cascade filters and psychophysical kernels [59].
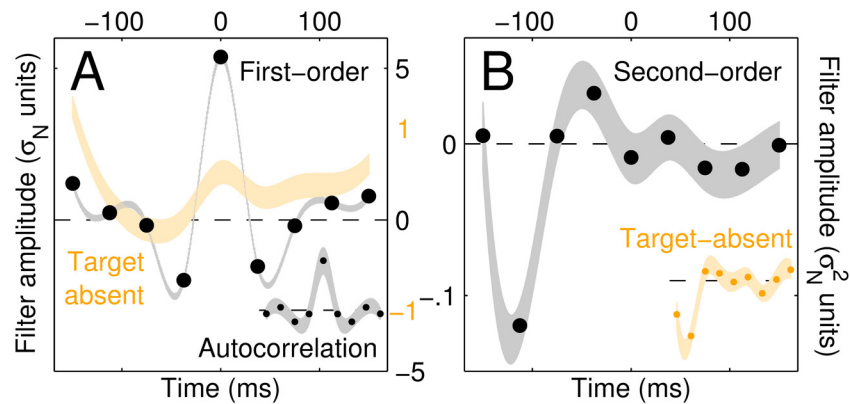
Fig 3. Measured kernels and associated system identification estimates of $\mathtt{L}_1$. Target-present/target-absent (black/orange) first-order kernels for increment detection are plotted in A (shaded regions show ±1 SEM). B plots first row from second-order kernel (or equivalently column, see S2C–S2E Fig); inset (orange) shows same from target-absent data. The black trace in A is an estimate of the autocorrelation of $\mathtt{L}_1$ [51], while the trace in B is an (independent) estimate of $\mathtt{L}_1$ itself [59, 64]. To confirm this, we plot the autocorrelation of B in the inset of A: it matches the black trace shown in the main panel of A (see Materials and Methods for details).

doi:10.1371/journal.pcbi.1005019.g003

Derivation of psychophysical kernels

Psychophysical kernels are used as compact descriptors of stimulus properties that impact listeners' decisions in simple detection/discrimination tasks [51, 60]. The most effective approach to an intuitive understanding of the 1st order kernel is to think of the underlying perceptual process as a matched template that assigns a set of weights to different elements of the stimulus, sums across all elements, and finally converts this weighted sum into a binary decision of the kind 'I saw the target' or 'I did not see it' [49, 61]. For this model, the psychophysical 1st order kernel (computed as described below) is an image of the template [49, 60]: it details the perceptual impact associated with different portions of the auditory waveform. It also does not matter whether the kernel is computed from noise modulations associated with the target or not: in both cases, it will reflect the template associated with the model outlined above, if this model provides an adequate account of the perceptual process [49, 60]. There are many conditions, however, when the 1st order kernel does not retain the intuitively transparent interpretation proposed above, e.g. in the presence of a nonlinear transformation between the stimulus and the response such as a dependence on the power or correlational structure of the stimulus [62, 63]. In order to extract useful information about the underlying process, it then becomes necessary to study higher-order descriptors such as the 2nd or 3rd order kernels [64], requiring more data and elaborate models. For example, the 2nd order kernel is useful if we suspect that the perceptual process assigns a set of weights to all possible pairwise *interactions* between different elements of the stimulus, e.g. between the amplitudes of the first and the second segments of the auditory waveform, or between the amplitudes of the first and third segments, and so on. These interaction terms would provide additional information about the stimulus properties that affect listeners' choices beyond the description afforded by the 1st order kernels [51, 59, 64]. To compute 1st and 2nd order kernels, we denote the AM noise modulation applied on the target-present ($q = 1$) or target-absent ($q = 0$) interval of a trial to which listeners responded correctly ($r = 1$) or incorrectly ($r = 0$) by the 9-element vector $\mathbf{n}^{[q,r]}$. The first-order target-present psychophysical kernels (i.e. those obtained *only* from noise trials containing the target) were computed as $\mathbf{p}_1^{[1]} = \mathtt{avg}(\mathbf{n}^{[1,1]}) - \mathtt{avg}(\mathbf{n}^{[1,0]})$ where $\mathtt{avg}(.)$ is used to indicate average across trials of the specified type [49]; the target-absent

kernels were $\mathbf{p}_1^{[0]} = \mathrm{avg}(\mathbf{n}^{[0,0]}) - \mathrm{avg}(\mathbf{n}^{[0,1]})$. The second-order psychophysical kernels were similarly computed as $\mathbf{p}_2 = \mathrm{cov}(\mathbf{n}^{[1,1]}) + \mathrm{cov}(\mathbf{n}^{[0,0]}) - \mathrm{cov}(\mathbf{n}^{[1,0]}) - \mathrm{cov}(\mathbf{n}^{[0,1]})$ where $\mathrm{cov}(.)$ indicates covariance across trials. Please see [51, 60] for further details of these methods.

$\mathtt{L_1}/\mathtt{L_2}$ estimates via cascade characterization

Cascade filters and psychophysical kernels are different classes of objects. Cascade filters are filtering components of a hypothesized cascade model; for the $\mathbf{L_1NL_2}$ cascade described previously, they cannot be estimated directly from data via simple rules. They can, however, be estimated *indirectly* via the psychophysical kernels. Psychophysical kernels are data descriptors computed directly from the raw data using simple rules (see above); in this sense, they are not dissimilar from simply computing a summary statistic (e.g. mean or median) from a dataset. Their estimation is robust and does not depend on any assumed underlying model. If a model is assumed, the kernels can then be used to characterize specific components of the model. For example, the $\mathbf{LN}$ model is widely adopted for this type of application [49]; in its psychophysical variant, this model reduces to template matching for the $\mathbf{L}$ stage [51], i.e. inner product between the input stimulus and the $\mathbf{L}$ template [65–67]. Under this model, the first-order psychophysical kernel returns a scaled image of the template $\mathbf{L}$ [60], allowing for direct transparent estimation of the linear filtering stage.

For the purpose of our study the $\mathbf{LN}$ model is inapplicable as it predicts [49, 51, 58, 68, 69] that perceptual kernels derived from noise modulations associated with signal+noise stimuli ('target-present') must match those derived from noise-only stimuli ('target-absent'). This property is a consequence of the linear nature of the $\mathbf{L}$ stage, combined with the classic reverse-correlation result that the static nonlinear $\mathbf{N}$ stage is bypassed by the kernel estimation procedure [60, 70, 71]. The response of $\mathbf{L}$ to signal+noise is simply the sum of its response to signal plus its response to noise, effectively decoupling the noise-driven response from that associated with the signal. The filter perturbation associated with the noise element is therefore statistically analogous between signal+noise and noise-only stimuli, leading to equivalent kernel estimates [68, 69]. This prediction is not born out by the results in S2A and S2B Fig which show that target-absent and target-present first-order kernels respectively are markedly different (compare not only the shape of the traces, but also the scaling of ordinate between S2A and S2B Fig; see also [72] and S7A Fig). This finding is consistent with the expectation that the underlying mechanism would more likely conform to cascade models including additional filters, e.g. the $\mathbf{L_1NL_2}$ cascade in Fig 2C (see more below); because this cascade does not belong to the $\mathbf{LN}$ family of models, it does not predict that target-present and target-absent 1st order kernels should match but rather that (in general) they should differ [51], as we observe.

Under the $\mathbf{L_1NL_2}$ cascade model, the connection between filter components ($\mathbf{L_1}/\mathbf{L_2}$) and psychophysical kernels ($\mathbf{p}_1^{[1]}/\mathbf{p}_1^{[0]}/\mathbf{p}_2$) is provided by the following three theoretical results (#1-2 pertaining to $\mathbf{L_1}$, #3 pertaining primarily to $\mathbf{L_2}$). Result #1: target-present first-order kernels (Fig 3A) return an approximate image of $\mathbf{L_1}$ autocorrelation [51]; we can exploit this result to study the characteristics of $\mathbf{L_1}$, with the cautionary note that the relationship between $\mathbf{p}_1^{[1]}$ and $\mathbf{L_1}$ involves other terms besides $\mathbf{L_1}$ autocorrelation and that the relative contribution of these terms depends on stimulus SNR [51]. Result #2: $\mathbf{L_1}$ can also be estimated from *second-order* kernels by exploiting the established result that the first row (or column) of the second-order kernel (black and red rectangles in S2C and S2D Fig, replotted as traces in Fig 3B) returns an approximate image of $\mathbf{L_1}$ [51, 59, 64]. This is the approach typically adopted for solving $\mathbf{L_1NL_2}$ cascades [59, 64]; we conform to this practice by relying primarily on these $\mathbf{L_1}$ estimates here. We can cross-check the consistency of $\mathbf{L_1}$ estimates returned by these two approaches: if we take the autocorrelation of the trace in Fig 3B and plot it in the inset to Fig 3A, it should

resemble the trace in Fig 3A. This prediction is well realized by data (see also S2B and S2E Fig), thus lending further support to the applicability of the associated analytical tools to the present context.

We can also estimate L_2 by relying on the additional result (#3) that target-absent first-order psychophysical kernels (S2A Fig) return the cross-correlation between L_1 and L_2, as detailed in [51, 73]. We can then deconvolve the L_1 estimates (obtained as described above) out of target-first order kernels to obtain estimates for L_2. The additional deconvolution step involved in deriving L_2 partly justifies the noisiness associated with the aggregate estimate in Fig 2D (shading shows ±1 SEM; this is compounded by computing the power spectrum from time-based filter estimates before combining them across listeners, who naturally displayed a significant degree of individual variability). Notice that all our conclusions are based on quantitative analysis of individual listener data (Figs 4, 5B–5D and 5F); aggregate estimates are shown for visualization purposes only.

In practice, L_1/L_2 estimates will be distorted images of the kernels associated with theoretical accounts of L_1NL_2 cascades due to analytical approximations [51, 73] and the highly non-linear properties of AM extraction [16], but the qualitative nature of their filtering characteristics (whether low-, high- or band-pass) is preserved in the presence of these distortions (as we have verified via Monte Carlo simulations of full-scale models explicitly encompassing all stages from cochlear filtering to binary choice; see S5 Fig for a diagrammatic representation of one such simulation).

Scalar metrics for quantitative evaluation of L_1/L_2 estimates

We computed the power spectrum $\mathbf{w}$ of each L_1 and L_2 estimate and gauged its band-pass characteristics using four parameter-free metrics; because target-present kernels (Fig 3A) approximate L_1 autocorrelation, the corresponding power spectrum was obtained via Fourier transform [74]. The spectral centroid (Fig 4A) was $\mathbf{f} \bullet \hat{\mathbf{w}}$ where $\mathbf{f}$ is the vector of sampled spatial frequencies and $\hat{\mathbf{w}}$ was obtained by normalizing $\mathbf{w}$ to sum 1. The ratio between spectral centroid and SD of $\hat{\mathbf{w}}$ provides a surrogate index of band-pass characteristics for bell-shaped $\hat{\mathbf{w}}$; the transition value from low-pass to high-pass (marked by vertical dashed line in Fig 4B) corresponds to a uniform spectrum, for which the centroid/SD ratio is $\sqrt{8}/2$ given the sampling rate used here. To establish a link with existing literature, in Fig 4B we plot this quantity in units of Q, the mean/width ratio for a rectangular shape; the conversion is obtained by approximating an assumed Gaussian shape with a rectangle of equivalent full-width at half-height (FWHH) [7], for which the conversion factor is FWHH $= \mathrm{SD} \times 2\sqrt{2\log(2)}$. AC/DC energy log-ratio (Fig 4C) was $\log[\mathbf{w}(\mathbf{f} > 0)/\mathbf{w}(0)]$ where $\mathbf{w}(0)$ is power at frequency 0 (DC) and $\mathbf{w}(\mathbf{f} > 0)$ is all remaining power in the spectrum. Spectral slope (Fig 4D) was the correlation coefficient of $\mathbf{w}$ across $\mathbf{f}$. The composite band-pass index in Fig 5D consisted of paired comparisons between 'early' and 'late' estimates of Q, AC/DC energy log-ratio and spectral slope; because Q is always positive the comparison involved log-ratios, while AC/DC energy log-ratios and spectral slopes (can be negative) were compared via subtraction.

Epoch definition and related analysis

Increment and decrement experiments were often run in alternate fashion across sessions (although they were never mixed within the same session/day); for exposure-related analyses (Fig 5), they were combined following sign inversion for noise modulations from decrement experiments to align them with the increment data. We restricted our analysis to the smallest number of trials collected by any listener (3.5k) to make the analysis comparable across listeners (35k trials contributed to Fig 5). We split this initial period of data collection into 'early'
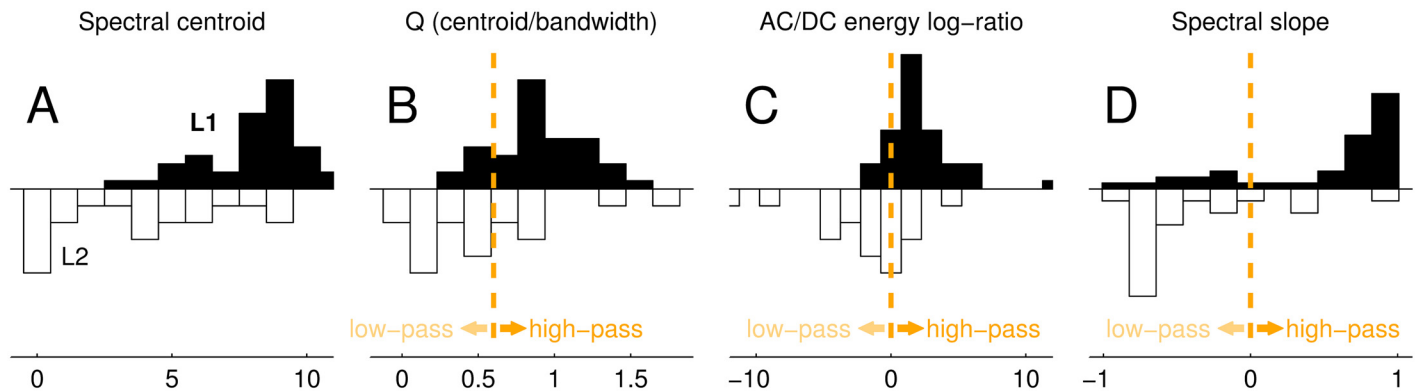
Fig 4. The band-pass character of L₁. The band-pass characteristics of the L_1 filter, contrasted with the low-pass L_2 read-out filter, are demonstrated by the distribution of four filter parameters (centroid in A, Q in B, AC/DC energy log-ratio in C, spectral slope in D) applied to all L_1/L_2 estimates. Parameters are gathered from across all experiments (increment/decrement detection) and listeners. Vertical dashed lines in panels (B,C,D) mark transition from low-pass to high-pass for each given parameter (see Materials and Methods for details).

and 'late' epochs by assigning the first 1750 trials to the former and the second 1750 trials to the latter. For the 10-epoch analysis (Fig 5E and 5F) we split the same period into 10 epochs of 350 trials each. Centroid drift (x axis in Fig 5F) is the correlation coefficient of centroid versus logarithm of epoch number. We logged epoch number before computing correlation because the semilogarithmic representation corresponded to an excellent linear fit of the aggregate data in Fig 5E (see solid gray lines).

Coupling between noise DC fluctuations and trial-by-trial responses

We averaged noise amplitude (as specified by the 9-element vector **n** detailed above) and used it as proxy for the noise-induced DC perturbation of each stimulus. We then took the difference in noise DC content between the two stimuli presented on each trial, and computed the biserial correlation coefficient between this differential DC content and the binary response returned by listeners (S4B Fig). We applied the above calculation only to the noise component of each stimulus (without target signal) because, when the signal is included, DC content is almost invariably greater for the stimulus containing the target signal; given that listeners performed above chance (i.e. their response was correlated with target presence), it is trivial that we should find a correlation (which we do find in all instances) between differential target-driven DC content and behavioural response. We therefore focused on the behavioural component of the response that was specifically driven by trial-to-trial random fluctuations of DC content (i.e. those induced by the noise), rather than the expected correlation with target presence.

Stimulus bandwidth characterization and correction for carrier-induced AM deviations

We estimated stimulus task-relevant modulation for different centroid values by extracting the AM content of signal+noise and noise-only stimulus waveforms via a bank of 1-octave AM filters centred at 2, 4, 8, 16 and 32 Hz, plus a lowpass filter with a cut-off at ∼1.5 Hz (approximating a DC-driven loudness estimator). Each simulation returned the difference between the AM content of the two traces over 1000 trials, and the average of 100 simulations is plotted in Fig 5E (orange shading; see [38] for related results). S5 Fig illustrates this procedure for 1 simulation (see caption for details of individual panels).
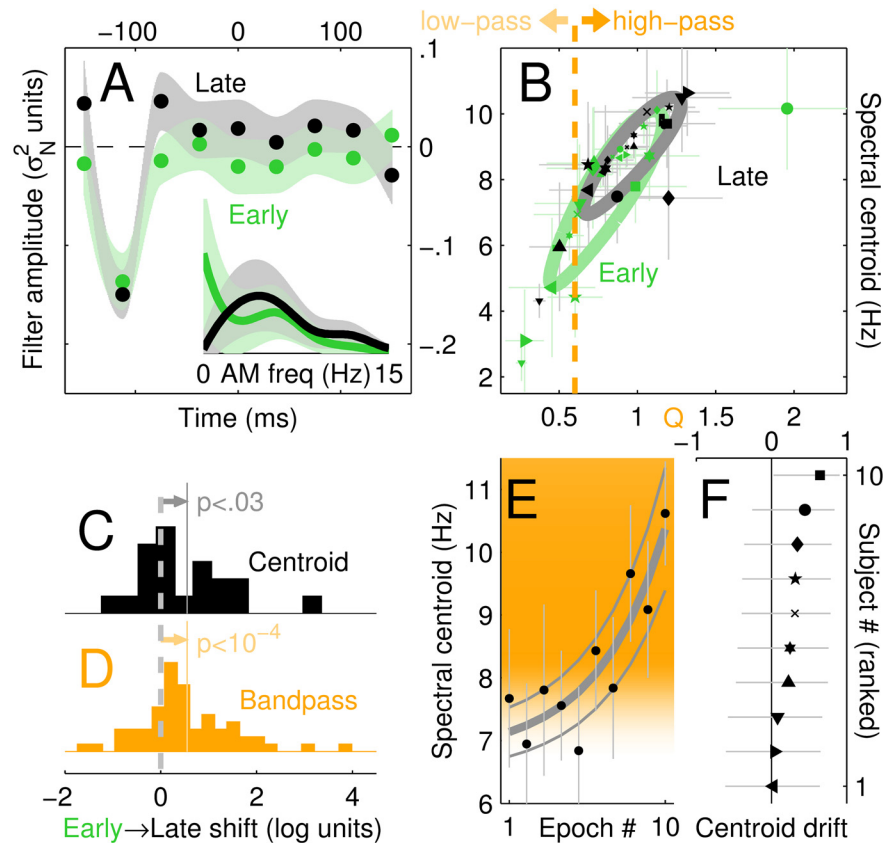
Fig 5. Front-end filter (L₁) shifts from low-pass to band-pass characteristics with exposure. (A) L_1 estimates as in Fig 3B, separately for 'early' (green) and 'late' (black) phases of the data collection epochs (see Materials and Methods). Inset curves show the corresponding low-pass (early) and band-pass (late) power spectra of the two L_1 estimates. (B) Quantifying the shift from low-pass to band-pass shapes of the L_1 filter. The spectral centroid (y axis) is plotted against the Q estimates using the same symbols as in Fig 4. Ovals are aligned with best-fit line, with radii matched to SD's for data projected parallel and orthogonal to line. (C) Distribution of late/early log-ratios for centroid values. (D) Distribution of late/early comparisons for all three band-pass metrics (Q, AC/DC energy log-ratio, spectral slope). (E) Gradual change in the spectral centroids in 10 successive epochs during the experiments (rather than just early/late); orange shading reflects task-relevant stimulus modulations across centroid values in the bandpass region (see Materials and Methods and S5I Fig for a full plot of these modulations spanning the low frequency region). Solid line shows semi-log linear fit (thin lines plot boundaries ±1 SEM around fit parameters). Associated correlation coefficient (centroid drift) is plotted separately for each subject in (F). Error bars/margins plot ±1 SEM in (A-B, E); ±95% confidence intervals in (F).

We also wished to verify that the slight amplitude deviations introduced by the randomly generated carrier (which was refreshed from block to block, see above) did not affect our calculations based on the notional AM perturbations specified before application of the carrier to generate the stimulus waveform. To this end, we applied the following energy-based recalibration to the stimulus samples in our dataset: each waveform as it was delivered to the listener was split into 9 segments, and RMS (root-mean-square) was computed from each segment to obtain a proxy AM vector equivalent to the 9-element vector specified by our amplitude modulation protocol. We then applied the same kernel estimation procedures used with the pre-specified noise samples to these RMS-corrected samples. The resulting kernel estimates are plotted in S6 Fig, where it can be seen that they demonstrate the same characteristics as obtained before RMS recalibration.

Results

Cascade approximation of auditory modulation filter models

Data from reverse correlation experiments are almost invariably interpreted with relation to a cascade that only incorporates one linear filter $\mathbf{L}_1$ applied via template-matching, followed by a threshold conversion to binary decision [49]. However, as discussed in Materials and Methods, this linear-nonlinear (LN) class of models is inadequate for our dataset (we return to this point in Discussion). Instead, we adopt the more general and highly successful linear-nonlinear-linear ($\mathbf{L}_1\mathbf{NL}_2$) cascade [64] which, in its most general formulation, serves as a functional approximator of wide applicability [75, 76] (notice that the $\mathbf{L}_1$ linear stage in the $\mathbf{L}_1\mathbf{NL}_2$ cascade involves convolution, not template matching; see Materials and Methods for further details). Qualitative inspection of the second-order kernels associated with the experiments described here appears consistent with this class of models [51, 64] (as discussed in Materials and Methods and demonstrated in S2C and S2D Fig).

Both LED and MFB can be cast in the form of $\mathbf{L}_1\mathbf{NL}_2$ cascades as illustrated in Fig 2. The only difference between the two models lies in the characteristics of the $\mathbf{L}_1$ stage (blue in Fig 2): low-pass for LED [6] (Fig 2A), primarily band-pass for MFB [16] (Fig 2B). Subsequent stages depend on relatively arbitrary choices of read-out rules [21], but in general they can all be approximated by the combination of a static nonlinearity ($\mathbf{N}$) and a subsequent linear stage ($\mathbf{L}_2$). Established techniques in nonlinear system identification [59, 64], combined with idiosyncratic features of their psychophysical variants [51], can be exploited to derive estimates for both $\mathbf{L}_1$ and $\mathbf{L}_2$ from first-order and second-order psychophysical kernels like those shown in Figs 3 and S2 (see Materials and Methods for a more detailed description of the connection between model components $\mathbf{L}_1/\mathbf{L}_2$ and psychophysical kernels).

The band-pass nature of AM filtering

Within the framework outlined above, $\mathbf{L}_1$ and $\mathbf{L}_2$ can be thought of as 'front-end' and 'read-out' filters. $\mathbf{L}_1$ is the component of primary interest for this study, as it supports the distinction between LED and MFB as being associated with low-pass versus band-pass characteristics respectively (see next section for discussion of recent variants of the MFB model incorporating a lowpass filter [35]). Estimates of this filter are depicted in Fig 3A and 3B (the black trace in A is an estimate of the filter autocorrelation, see Materials and Methods). They show that $\mathbf{L}_1$ presents band-pass characteristics with $Q \sim 1$ centred around 8 Hz (thus favouring the MFB model overall). This band-pass property is not an artefactual distortion induced by the target signal, because it is preserved when estimates are obtained from target-absent noise modulations alone (inset to Fig 3B). As we demonstrate in later sections of this study, it is also an evolving characteristic that may not be present at all stages of stimulus exposure.

We quantify this band-pass finding using four different metrics (Fig 4; see Materials and Methods). In *all* cases, $\mathbf{L}_1$ estimate distributions (solid histograms in Fig 4B–4D) fall within the highpass/band-pass range, while $\mathbf{L}_2$ estimate distributions (open histograms) fall within the low-pass range (the latter result points to a late temporal integration window of 50–100 ms consistent with independent estimates from previous studies [21]). More specifically, spectral centroids (Fig 4A) for $\mathbf{L}_1$ are larger than for $\mathbf{L}_2$ ($p < 10^{-5}$, unpaired two-tailed Wilcoxon test); Q values (Fig 4B) for $\mathbf{L}_1$ (but not $\mathbf{L}_2$) are significantly larger ($p < 10^{-5}$, two-tailed Wilcoxon test) than expected for a uniform spectrum (indicated by dashed lines in Fig 4B); the AC/DC energy ratio (Fig 4C) is larger than 0 ($p < 10^{-5}$) for $\mathbf{L}_1$ (indicative of band-pass/highpass characteristics) and smaller than 0 for $\mathbf{L}_2$ ($p < 0.05$); the spectral slope (Fig 4D) is positive for $\mathbf{L}_1$ ($p < 10^{-4}$) but negative for $\mathbf{L}_2$ ($p < 0.01$).

It is also noteworthy that results were comparable between datasets from increment detection and decrement detection (y axis): there was no statistically significant difference (at $p>0.05$) for any metric and for either L_1 or L_2 (data points scatter around solid unity lines in S3 Fig). Such convergence of independent datasets indirectly validates our estimation procedure and suggests that increments and decrements may be processed by the same perceptual mechanism (as also indicated by the similarity in performance metrics, see S1 Fig).

The overall conclusion from the above analyses is that the L_1 filter is band-pass; therefore, AM processing resembles the characteristics of the MFB more than the LED model. This conclusion is not the product of fitting either model to the data: it is based on non-parametric characterization of the front-end filter associated with a general framework cascade that encompasses both models (Fig 2).

Exposure-mediated retuning of AM filters

All estimates described above were obtained by pooling trials across the entire data collection period undertaken by each listener, spanning several sessions on different days. The characteristics of the perceptual process may have undergone substantial modifications over this extended period, particularly considering that listeners received trial-by-trial feedback and were therefore encouraged to optimize their strategy. To investigate this possibility we defined *early* versus *late* epochs for data collection (see Materials and Methods for definition). The L_1 estimates associated with the two epochs differed: only the 'late' estimate (black in Fig 5A) exhibited band-pass characteristics. In contrast, the 'early' estimate (green in Fig 5A) was closer to low-pass (see also AM frequency plots within inset). Similar exposure-mediated changes in kernel structure have been previously reported in the vision literature [31, 32]; Fig 5 offers the first demonstration for auditory processing.

The above result is supported by metric analysis of individual listener data: Q estimates (x axis in Fig 5B) are significantly larger (at $p<0.01$) than the lowpass/highpass cut-off point (orange vertical dashed line) for the late epoch (black), but not for the early epoch (green). We further probed this result with paired data analysis by computing a composite shift index for band-pass characteristics from early to late in each condition and each listener (see Materials and Methods); the resulting distribution (orange in Fig 5D) was significantly shifted away from 0 ($p<10^{-4}$) in the direction of greater band-pass for the late epoch. This shift in band-pass value was accompanied by a significant shift in filter centroid (Fig 5C).

Although more elaborate interpretations are possible, a parsimonious view of our measurements suggests that the shift involved an adjustment of the same underlying filter population, rather than ad-hoc neural assembly of a new filter bank: the 'late' dataset overlaps with the 'early' dataset in Q-centroid space, only restricted to a smaller region (compare black and green ovals in Fig 5B). When we compare the Q/centroid ranges spanned by the two epochs, we find that the lower percentile boundary (5%) shifts from 0.27 (early) to 0.44 (late) for Q and 2.7 to 5.1 for centroid, but the higher percentile boundary (95%) remains virtually unchanged at 1.5 (Q) and 10 Hz (centroid). A different but equivalent way of conceptualizing this result is to describe the early-late shift as reflecting differential weighting of two discrimination strategies: one relying on loudness, the other on the spectral shape of the modulation frequencies (temporal profile), both driven by valid cues for performing the task (see below for further discussion of this point). In the early phase, the two strategies would coexist and support discrimination to a roughly equal extent; in the late phase, the temporal profile strategy would play a more prominent role.

The above interpretation is consistent with an additional analysis where we estimated coupling between listeners' choices and the differential DC content (proxy for loudness) of the

noise samples presented on those same trials (see Materials and Methods). In the early phase, we found that correlation values across listeners were significantly different than 0 (data points in S4B Fig fall to the right of the vertical dashed line at p<0.02), indicating that the behavioural choices made by listeners were at least partly driven by stimulus loudness. In the late phase, correlation values did not demonstrate a significant shift away from 0 (data points in S4B Fig scatter around the horizontal dashed line at p<0.23), indicating that loudness did not play a significant role in driving behaviour during later phases of data collection.

The filter bank proposed by recent versions of the MFB model [35], encompassing a lowpass filter in the very low modulation range and bandpass filters at higher modulation rates, could accommodate our results when combined with appropriate weighting profiles. In this sense, our data provide support for a mixed lowpass/bandpass version of the MFB model combined with a flexible read-out stage that may undergo internally driven retuning. The lowpass filter recovered by our protocols should not be confounded with the processing stage preceding the filterbank in early formulations of the MFB model [39]; this stage does consist of a lowpass filter, but with a much higher cut-off frequency of 150 Hz. The lowpass filter of interest for the present discussion is therefore best viewed as a subcomponent of the filter-bank itself operating in the very low frequency range, rather than a separate earlier stage extending to the high frequency range.

To gain better insight into the temporal evolution of the exposure-mediated effects, we obtained centroid estimates across 10 different epochs of data collection. Centroid estimates drifted exponentially towards higher values (Fig 5E) matching closely the estimated modulation content of the stimulus (indicated by orange shading, see Materials and Methods and S5 Fig), and this effect was surprisingly robust across listeners: even though drift (see Materials and Methods for definition) returned a noisy measurement for individual listeners (see 95% confidence intervals in Fig 5F), it was consistently positive (symbols fall to the right of vertical line in Fig 5F) so that the overall trend across listeners was highly significant (p<0.005).

The above-detailed modifications of filter structure were associated with only mild improvements of absolute efficiency in some listeners (S4A Fig). This apparent decoupling between filter estimates and performance metrics is a well-documented finding in relation to various perceptual phenomena [77–79] including learning [33]. Direct coupling is theoretically expected only for LN models [80] which are not applicable to our experiments as pointed out earlier; therefore, the estimated filters cannot be transparently linked to discrimination performance. Even if they were, there are at least two reasons why one may not expect to see performance differences.

First, learning effects on AM discrimination are small and difficult to expose (often requiring >100 listeners, see [81]). Second, successful discrimination in our task was supported by both lowpass and bandpass stimulus power (see two peaks in S5I Fig); in this respect our protocol differs from the equally valid ones adopted by previous studies (e.g. [82]) with the specific goal of excluding loudness cues (see Materials and Methods for clarifications as to why we deliberately avoided a stimulus design that would invalidate loudness cues). Because filter structure shifted between these two equivalent sources of task-related information (see above), discrimination performance may well remain unchanged even though supported by different regions of AM frequency.

Prompted by the above results, we re-analyzed data from a prior published study [58] to determine whether similar effects could be exposed for an independent dataset collected using substantially different stimulus/task designs. At the time when this dataset was published, the exposure-mediated effects reported in the present study were not known. We converted perceptual filters from the previous study into a format comparable with the one adopted here and

applied the same analysis; as demonstrated in S7 Fig, we obtained remarkably similar results, including lowpass-to-bandpass retuning during the first ∼4K trials.

Discussion

This study represents the first targeted application of psychophysical reverse correlation to AM processing. Although related tools have been applied successfully in auditory neuroscience [83] and psychoacoustics (see [84] and [85] for the case of spectral processing and [48] for an application to a loudness illusion), they have not addressed the specific case of modulation perception [86], possibly due to multiple challenges associated with this question. First, there is the critical issue of which stimulus dimension should be perturbed by the noisy process in order to provide meaningful and feasible leverage for tapping into the mechanisms responsible for analyzing AM signals. Adding acoustic white-noise is inappropriate, partly because the envelope fluctuations it induces are difficult to control and exercise adequately, and partly because the dimensionality of the space needing characterization is impractically large to measure [86]. Second, there is the question of whether the analytical toolkit associated with reverse correlation is sufficiently flexible to accommodate both LED and MFB: in the vast majority of its applications, reverse correlation is tightly coupled with the assumption of the linear-nonlinear (LN) cascade [49, 87]. Neither LED nor MFB can be correctly approximated by this model (and our data fail to comply with its basic prediction that target-present first-order kernels must match corresponding target-absent estimates, compare black versus orange traces in Fig 3A), requiring more elaborate analytical tools.

To overcome these challenges, we exploited techniques from nonlinear system identification analysis [59] where the coupling between input noise and output response is used not only to compute linear descriptors of the sensory process [55, 85, 87], but also nonlinear (second-order) descriptors that afford the opportunity to characterize more complex cascades than LN [64]. In particular, these additional tools can effectively constrain linear-nonlinear-linear (LNL) cascades [51] to which both LED and MFB likely conform when formulated with reference to the stimulus dimension perturbed in the experiments described here (Fig 2). By combining these system identification tools with AM-tailored stimulus perturbations, we gained sufficient insight into the perceptual process to constrain its properties via data-driven characterization. To appreciate the significance of the analysis adopted here, it is instructive to consider first-order filter estimates from target-absent noise modulations alone (orange trace in Fig 3A). This measurement is similar to the decision weight profiles reported by loudness studies [44–48], including evidence for a primacy effect (larger weights during early phase of the stimulus [47], see orange trace in Fig 3A), and is often regarded as a more appropriate description of the filtering process [88, 89]; more importantly, these measurements (from previous studies as well as our own) present *lowpass* characteristics, with no evidence of bandpass AM filtering. Signatures of bandpass processing are exposed specifically by second-order kernels (Fig 3B). Indeed, when the dataset from [48] is re-analyzed using the nonlinear tools described here, bandpass filtering becomes evident at the level of second-order kernels in the presence of decisively lowpass first-order kernels [90] (see [91, 92] for related examples from the vision literature).

A further enabling factor in the experiments reported here is the stability of task rules and impenetrability of cognitive introspection. Specifically, listeners in these experiments carried out the same task throughout data collection and it is extremely unlikely that from one trial to the next they could explicitly monitor all small deviations introduced by the noisy process and introspect cognitively on those to reach a decision. They therefore operated under relatively stable conditions (except for potential intrinsic changes in adaptive state), allowing us to

confidently treat our dataset as reflecting the properties of the same perceptual machinery throughout [31, 32] (even though specific parameters within that machinery may change with exposure). These advantages carry the cost of unusually large data mass (for this study we collected ~100k trials), restricting our investigation to only a limited ecologically relevant portion (3–12 Hz) of the system's operating regime, however they allow us to exclude stimulus-driven alterations. More specifically, previous literature has shown that auditory perceptual templates depend not only on signal spectrotemporal structure [26] but also on signal intensity [27] (see [92–94] for related results in the vision literature). Furthermore, strategy shifts not dissimilar from those we report here can be triggered by simple stimulus modifications such as signal-masker asynchrony [28]. It is therefore critical to use stable task rules and a statistically invariant stimulus throughout (as we have done in this study) if one is to ascribe spontaneous filter changes to the perceptual system alone.

Over the past decades, evidence in favour of LED/MFB models [7, 20] has been interpreted in the light of fitting procedures around specified computational implementations and often requiring ad-hoc adjustments for different datasets, weakening the associated conclusions regarding the applicability of one model over the other. Our approach did not favour any specific model, nor did it involve explicit implementation of those or other computational schemes. At the same time, it enabled model selection while retaining close proximity with data structure. These methods uncovered a more complex picture than initially suggested by the LED/MFB dichotomy. Although the low-pass/band-pass distinction retains its descriptive power in relation to our dataset, the underlying mechanism displays *dynamic* adaptive properties that potentially depend on various factors, most notably learning-mediated plasticity [95–99] (see also expectation effects on AM processing [29]). This raises the possibility that previous diverse interpretations of the results (with respect to lowpass versus bandpass filtering) may in fact both provide adequate representations of the underlying process, albeit under different learning states [18]. For example, long-term release from adaptation of postulated bandpass filters [18] may need to be re-interpreted in the context of exposure-mediated reweighting across the filter bank, rather than evidence against low-level adaptation within the MFB (as originally hypothesized in [18]). This interpretation may be verified/falsified using the tools developed and validated in this study.

A further unresolved issue concerning the LED/MFB distinction pertains to the specific Q value associated with AM filtering. It was originally hypothesized that the filterbank associated with the MFB exhibits slightly different Q's for low versus high modulation rates [16]: above ~10 Hz, filters would span a bandwidth that increased with center frequency; below ~10 Hz, bandwidth would remain roughly constant regardless of center frequency. Our dataset presents sufficient variability of estimated central spectroid to span the 3–10 Hz range (y axis in Fig 5B), allowing us to test the latter hypothesis directly. We find that the hypothesized trend is well supported by data: if filter bandwidth is relatively independent of center frequency, the ratio between center frequency and bandwidth (Q) should scale with center frequency; consistent with this prediction, we measured a strong correlation between spectral centroid and Q for both early ($r = 0.89, p < 10^{-7}$) and late ($r = 0.9, p < 10^{-6}$) epochs (see tilted ovals in Fig 5B). However, the specific values hypothesized for bandwidth ($Q \approx 2$; [16]) are higher than suggested by relevant studies ($Q \approx 1$; [24, 25]) and by some of the values we measured in this study (Fig 4B). Our results indicate that, even when restricted to comparable regions of AM rates, aggregate Q values may span a 0.3–1.7 range depending on the degree of recalibration undergone by the system via exposure/learning (see data scatter across x axis in Fig 5B), consistent with the latest estimates [24, 25] but substantially lower than the values employed by recent modelling work [35, 39].

To what extent do our results depend on the stimulus parameters and task specifications selected for this study? As explained earlier, we were constrained in our ability to test a wide range of configurations, however we did perform measurements using increment as well as decrement target signals. The spectral profile of task-relevant stimulus information is substantially different between these two configurations: in the increment case, the useful bandpass region lies between 8 and 20 Hz (S5I Fig), while the decrement configuration mostly targets the 4–8 Hz region (S5J Fig). As for the lowpass (loudness) cue, it required opposite read-out rules for the two configurations (target signal is louder in the increment configuration, and softer in the decrement configuration; see peaks of opposite signs in the lowpass region of S5I and S5J Fig). Despite these differences in the stimulus, we observed no difference between increment and decrement estimates of bandpass tuning (S2 and S3 Figs), and we found no correlation/relationship between the exposure-mediated effects (Fig 5E and 5F) and the relative exposure to increment versus decrement signals. Furthermore, the very fact that these characteristics (both centroid and Q values, Fig 5B–5D) changed with exposure in the face of an unchanging stimulus indicates that they are not solely driven by stimulus specification. This is not to say that our estimates are completely decoupled from the chosen stimulus/task parameters: by requiring observers to detect a specific signal, we are implicitly prompting them to calibrate their available perceptual resources in relation to the assigned task and signal [92, 94]; for example it is conceivable that, in extreme cases such as detection of very narrow AM pulses with a broad modulation spectrum, the measured filters may become broader reflecting stimulus characteristics [93]. However, beyond the inevitable structure imposed by task instructions and signal specification on listeners' selection of perceptual resources, our measurements appear to reflect properties that are intrinsic to the perceptual process and informative of its inherent characteristics. Finally, targeted re-analysis of an earlier published dataset [58] exposed structure entirely consistent with the results reported here (S7 Fig), providing strong validation of our findings: due to numerous design differences between the two studies, it is not trivially expected that we should find similar overall characteristics.

In summary, the experimental approach adopted in this study has enabled us to examine outstanding issues in the AM processing literature from a different perspective, and clarify important aspects of this phenomenon, enabled by a set of tools that has not been previously applied with relation to this phenomenon. We have delineated the relationship between LED and MFB models within the context of a prominent theoretical cascade framework [100], we have refined and further constrained previous estimates of channel selectivity for processing amplitude modulations, and we have demonstrated its spontaneous adaptive nature in the context of active listening tasks (see also [31, 32]). Further research and additional characterization will be necessary to establish the applicability of our findings across a wider range of tasks and determine the exact nature and functional purpose of exposure-dependent adaptive processes [101, 102].

Supporting Information

S1 Fig. Performance metrics are similar for detecting increments and decrements. (A) Sensitivity is in the $d' \sim 1$ range (unity is indicated by dashed lines). (B) Bias is not statistically different from 0 (indicated by dashed lines). (C) Internal noise is within expected range [56] (grey shaded area). (D) Absolute efficiency is within expected range. Ovals are aligned with best-fit line, with radii matched to 0.5× (thickest), 1× and 1.5× (thinnest) SD's for data projected parallel and orthogonal to line. Different symbol shapes refer to different listeners. Error bars plot ±1 SEM. There was no statistically significant difference between increment (x axis) and

decrement data (y axis) for any of these metrics.
(EPS)

S2 Fig. Detailed overview of kernel measurements. (A,B) Target-absent and target-present first-order kernels are plotted in A and B respectively (see Materials and Methods for detailed description of how these kernels were computed), for both increment (black) and decrement (red) detection tasks (shaded regions show ±1 SEM). Target-absent kernels differ markedly from target-present kernels (among other features, notice different scaling of y axis between A and B), ruling out the linear-nonlinear model typically assumed by reverse correlation studies. Inset in (B) shows the autocorrelation of the kernels in E. (C,D) Second-order kernels for increment and decrement detection (plotted as upper/lower triangular matrix respectively). (E) First row (or equivalently column) from second-order kernels (indicated by black/red rectangles in C-D). Inset shows same measurements but only from target-absent second-order kernels.
(EPS)

S3 Fig. Expanded view of bandpass characteristics for front-end filter (L_1) and read-out filter (L_2). Panels (A-D) display estimates from increment detection on x axis against those from decrement detection on y axis. Solid symbols refer to L_1 (bigger symbols for estimates obtained from first row/column of second-order kernels (S2E Fig) and smaller symbols for estimates obtained from target-present first-order kernels (S2B Fig)). Open symbols refer to L_2. Different symbol shapes refer to different listeners. Error bars plot ±1 SEM. As for performance metrics (S1 Fig), there was no statistically significant difference between increment (x axis) and decrement data (y axis) for any of these estimates.
(EPS)

S4 Fig. (A) Absolute efficiency [61] was similar during early (x axis) and late phases. (B) Coupling between noise-driven DC fluctuations and behavioural responses, however, was only present during the early phase (see Materials and Methods and main text).
(EPS)

S5 Fig. Bandwidth characteristics of signal and noise. The stimulus waveform (A) is mapped to its spectrotemporal representation (B) using standard cochlear parameters [103]. It is then processed by a bank of modulation 1-octave filters with central frequency ranging from lowpass to 64 Hz (C); the scalar output of each filter (plotted in D) is the mean/maximum (orange/black) absolute value across the corresponding spectrogram (mean trace has been rescaled by 3× to span range comparable to maximum trace). This procedure is applied to both signal +noise (A-D) and noise-only (E-H) stimuli (example shown for increment detection in A-I); the difference between the filter-bank output from the two stimuli provides an indication of the modulation frequency bands that supported above-chance discrimination (the region around lowpass (lp) and 8–16 Hz for increment in I, and 4–8 Hz for decrement in J). Solid lines in D,H-J show mean across 20 iterations of 5K trials each (±10 SD indicated by shaded region).
(EPS)

S6 Fig. Fine temporal structure does not impact our results. As detailed in Materials and Methods, a new carrier was generated at the beginning of each blok of data collection. This additional source of randomness in the stimulus caused slight amplitude deviations from those specified by the AM noise deliberately introduced and controlled by the experimenter. To verify that these deviations had no impact on our results, we reconstructed the amplitude content of individual noise samples while taking into account the specific carrier waveform presented on those specific trials (see Materials and Methods). In this figure, we reproduce the critical

measurements supporting our conclusions (target-present first-order kernels (left column) and L_1 estimates from second-order kernels) before and after correcting for carrier-induced AM deviations (black/red versus blue/yellow respectively). More specifically, A-B plot the quantities estimated in Fig 3A and 3B for increment detection; C-D plot the same for decrement detection. Shaded regions show ±1 SEM. The overlap between corrected and uncorrected estimates in A is such that the two traces (blue versus black) are barely distinguishable in the plot. (EPS)

S7 Fig. Confirmation of lowpass-to-bandpass shift from post-hoc analysis of prior dataset. We re-analyzed data from a previously published study [58] in light of the results obtained here. More specifically, we converted spectrograms from reference [58] (defined across the dimensions of both time and frequency) into vectors defined along time alone by extracting values only at the target frequency, to bring them into coarse alignment with the perceptual filters used in the present study. Except for small differences in total duration and sampling rate (7 samples spanning 280 ms in the prior study versus 9 samples spanning 300 ms in the present study), results from the two studies are therefore comparable in data format (even though different stimuli (in particular target signals) and tasks were employed for the two studies). A is plotted to the same conventions adopted in Fig 3A. B is plotted to the conventions adopted in the inset to Fig 5A. To make the early/late analysis comparable to the analysis used in Fig 5, we defined the 'early' phase as spanning the first 2000 trials (from trial 1 to 2000) and the 'late' phase as spanning the second 2000 trials (from trial 2001 to 4000) for each of the 8 observers tested in the prior study (main two-tone condition). This analysis exposes structure in the previously published dataset consistent with the effects reported in the present study, further supporting their validity and applicability to other stimuli/tasks. (EPS)

Author Contributions

Conceived and designed the experiments: PN ERMJ. Performed the experiments: ERMJ PN. Analyzed the data: PN SAS ERMJ. Contributed reagents/materials/analysis tools: SAS CL PN. Wrote the paper: SAS CL PN.

References

1. Singh NC, Theunissen FE (2003) Modulation spectra of natural sounds and ethological theories of auditory processing. J Acoust Soc Am 114: 3394–3411. doi: 10.1121/1.1624067 PMID: 14714819

2. Drullman R, Festen JM, Plomp R (1994) Effect of reducing slow temporal modulations on speech reception. J Acoust Soc Am 95: 2670–2680. doi: 10.1121/1.409836 PMID: 8207140

3. Drullman R, Festen JM, Plomp R (1994) Effect of temporal envelope smearing on speech reception. J Acoust Soc Am 95: 1053–1064. doi: 10.1121/1.409836 PMID: 8132899

4. Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270: 303–304. doi: 10.1126/science.270.5234.303 PMID: 7569981

5. Christiansen C, MacDonald EN, Dau T (2013) Contribution of envelope periodicity to release from speech-on-speech masking. J Acoust Soc Am 134: 2197–2204. doi: 10.1121/1.4816409 PMID: 23967949

6. Viemeister NF (1979) Temporal modulation transfer functions based upon modulation thresholds. J Acoust Soc Am 66: 1364–1380. doi: 10.1121/1.383531 PMID: 500975

7. Moore BCJ (2003) An introduction to the psychology of hearing. Academic Press.

8. McDermott JH, Simoncelli EP (2011) Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. Neuron 71: 926–940. doi: 10.1016/j.neuron.2011.06.032 PMID: 21903084

9. Elhilali M, Chi T, Shamma S (2003) A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. Speech Communication 41: 331–348. doi: 10.1016/S0167-6393(02)00134-6

10. Jörgensen S, Ewert SD, Dau T (2013) A multi-resolution envelope-power based model for speech intelligibility. J Acoust Soc Am 134: 436–446. doi: 10.1121/1.4807563 PMID: 23862819

11. Shamma S, Lorenzi C (2013) On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. J Acoust Soc Am 133: 2818–2833. doi: 10.1121/1.4795783 PMID: 23654388

12. Bruna J, Mallat S (2013) Invariant Scattering Convolution Networks. IEEE Trans Pattern Anal Mach Intell 35: 1872–1886. doi: 10.1109/TPAMI.2012.230 PMID: 23787341

13. Mallat S (2012) Group invariant scattering. Communications on Pure and Applied Mathematics 65: 1331–1398. doi: 10.1002/cpa.21413

14. Andén J, Mallat S (2011) Multiscale Scattering for Audio Classification. ISMIR.

15. Strickland EA, Viemeister NF (1996) Cues for discrimination of envelopes. J Acoust Soc Am 99: 3638–3646. doi: 10.1121/1.414962 PMID: 8655796

16. Dau T, Kollmeier B, Kohlrausch A (1997) Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. J Acoust Soc Am 102: 2906–2919. doi: 10.1121/1.420345 PMID: 9373977

17. Wojtczak M, Viemeister NF (2005) Forward masking of amplitude modulation: basic characteristics. J Acoust Soc Am 118: 3198–3210. doi: 10.1121/1.2042970 PMID: 16334900

18. Bruckert L, Herrmann M, Lorenzi C (2006) No adaptation in the amplitude modulation domain in trained listeners. J Acoust Soc Am 119: 3542–3545. doi: 10.1121/1.2200696 PMID: 16838495

19. Plack CJ, Moore BC (1990) Temporal window shape as a function of frequency and level. J Acoust Soc Am 87: 2178–2187. doi: 10.1121/1.399185 PMID: 2348022

20. Plack CJ (2010) Hearing. Oxford: Oxford University Press.

21. Lorenzi C, Berthommier F, Demany L (1999) Discrimination of amplitude-modulation phase spectrum. J Acoust Soc Am 105: 2987–2990. doi: 10.1121/1.426911 PMID: 10335649

22. Viemeister NF, Stellmack MA, Byrne AJ (2004) The role of temporal structure in envelope processing. Auditory Signal Processing: Physiology, Psychoacoustics, and Models (Pressnitzer D, de Cheveigne A, McAdams S and Collet L eds), New York: Springer-Verlag: 221–229.

23. Sheft S, Yost WA (2007) Discrimination of starting phase with sinusoidal envelope modulation. J Acoust Soc Am 121: L84–89. doi: 10.1121/1.2430766

24. Ewert SD, Dau T (2000) Characterizing frequency selectivity for envelope fluctuations. J Acoust Soc Am 108: 1181–1196. doi: 10.1121/1.1288665 PMID: 11008819

25. Lorenzi C, Soares C, Vonner T (2001) Second-order temporal modulation transfer functions. J Acoust Soc Am 110: 1030–1038. doi: 10.1121/1.1383295 PMID: 11519571

26. Huang R, Richards VM (2008) Estimates of internal templates for the detection of sequential tonal patterns. J Acoust Soc Am 124: 3831–3840. doi: 10.1121/1.2967827 PMID: 19206809

27. Lentz JJ (2007) Variation in spectral-shape discrimination weighting functions at different stimulus levels and signal strengths. J Acoust Soc Am 122: 1702. doi: 10.1121/1.2756800 PMID: 17927430

28. Qian J, Richards VM (2010) The effect of onset asynchrony on relative weights in profile analysis. J Acoust Soc Am 127: 2461–2465. doi: 10.1121/1.3314251 PMID: 20370029

29. Wright BA, Dai H (1998) Detection of sinusoidal amplitude modulation at unexpected rates. J Acoust Soc Am 104: 2991–2996. doi: 10.1121/1.423881 PMID: 9821343

30. Patterson RD (1976) Auditory filter shapes derived with noise stimuli. J Acoust Soc Am 59: 640–654. doi: 10.1121/1.380914 PMID: 1254791

31. Li RW, Levi DM, Klein SA (2004) Perceptual learning improves efficiency by re-tuning the decision'-template' for position discrimination. Nat Neurosci 7: 178–183. doi: 10.1038/nn1183 PMID: 14730311

32. Li RW, Klein SA, Levi DM (2008) Prolonged perceptual learning of positional acuity in adult amblyopia: perceptual template retuning dynamics. J Neurosci 28: 14223–14229. doi: 10.1523/JNEUROSCI.4271-08.2008 PMID: 19109504

33. Dobres J, Seitz AR (2010) Perceptual learning of oriented gratings as revealed by classification images. J Vis 10: 1–11. doi: 10.1167/10.13.8

34. Stein A, Ewert SD, Wiegrebe L (2005) Perceptual interaction between carrier periodicity and amplitude modulation in broadband stimuli: a comparison of the autocorrelation and modulation-filterbank model. J Acoust Soc Am 118: 2470–2481. doi: 10.1121/1.2011427 PMID: 16266168

35. Dau T, Piechowiak T, Ewert SD (2013) Modeling within- and across-channel processes in comodulation masking release. J Acoust Soc Am 133: 350–364. doi: 10.1121/1.4768882 PMID: 23297908

36. Bacon SP, Grantham DW (1989) Modulation masking: effects of modulation frequency, depth, and phase. J Acoust Soc Am 85: 2575–2580. doi: 10.1121/1.397751 PMID: 2745880

37. Houtgast T (1989) Frequency selectivity in amplitude-modulation detection. J Acoust Soc Am 85: 1676–1680. doi: 10.1121/1.397956 PMID: 2708683

38. Gallun FJ, Hafter ER (2006) Amplitude modulation sensitivity as a mechanism increment detection. J Acoust Soc Am 119: 3919–3930. doi: 10.1121/1.2200136 PMID: 16838535

39. Piechowiak T, Ewert SD, Dau T (2007) Modeling comodulation masking release using an equalization-cancellation mechanism. J Acoust Soc Am 121: 2111–2126. doi: 10.1121/1.2534227 PMID: 17471726

40. Schreiner CE, Urbas JV (1988) Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. Hear Res 32: 49–63. doi: 10.1016/0378-5955(88)90146-3 PMID: 3350774

41. Eggermont JJ (1994) Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. Effects of carrier type, modulating waveform and intensity. Hear Res 74: 51–66. doi: 10.1016/0378-5955(94)90175-9 PMID: 8040099

42. Liegeois-Chauvel C, Lorenzi C, Trebuchon A, Regis J, Chauvel P (2004) Temporal envelope processing in the human left and right auditory cortices. Cereb Cortex 14: 731–740. doi: 10.1093/cercor/bhh033 PMID: 15054052

43. Wakefield GH, Viemeister NF (1984) Selective adaptation to linear frequency-modulated sweeps: evidence for direction-specific FM channels? J Acoust Soc Am 75: 1588–1592. doi: 10.1121/1.390808 PMID: 6736421

44. Dai H, Nguyen Q, Green DM (1996) Decision rules of listeners in spectral-shape discrimination with or without signal-frequency uncertainty. J Acoust Soc Am 99: 2298–2306. doi: 10.1121/1.415417 PMID: 8730076

45. Lutfi RA, Gilbertson L (2011) Effect of decision weights and internal noise on the growth of d' with N. J Acoust Soc Am 130: L329–333. doi: 10.1121/1.3645968

46. Oberfeld D, Plank T (2011) The temporal weighting of loudness: effects of the level profile. Atten Percept Psychophys 73: 189–208. doi: 10.3758/s13414-010-0011-8 PMID: 21258919

47. Oberfeld D, Heeren W, Rennies J, Verhey J (2012) Spectro-temporal weighting of loudness. PLoS ONE 7: e50184. doi: 10.1371/journal.pone.0050184 PMID: 23209670

48. Ponsot E, Susini P, Saint Pierre G, Meunier S (2013) Temporal loudness weights for sounds with increasing and decreasing intensity profiles. J Acoust Soc Am 134: L321–326.

49. Murray RF (2011) Classification images: A review. J Vis 11 (5): 1–25. doi: 10.1167/11.5.2

50. Steeneken HJ, Houtgast T (1980) A physical method for measuring speech-transmission quality. J Acoust Soc Am 67: 318–326. doi: 10.1121/1.384464 PMID: 7354199

51. Neri P (2010) Stochastic characterization of small-scale algorithms for human sensory processing. Chaos 20: 045118. doi: 10.1063/1.3524305 PMID: 21198130

52. Berg BG, Nguyen QT, Green DM (1992) Discrimination of narrow-band spectra. I: Spectral weights and pitch cues. J Acoust Soc Am 92: 1911–1918. doi: 10.1121/1.405238 PMID: 1401535

53. Rabinowitz NC, Willmore BD, Schnupp JW, King AJ (2011) Contrast gain control in auditory cortex. Neuron 70: 1178–1191. doi: 10.1016/j.neuron.2011.04.030 PMID: 21689603

54. Rabinowitz NC, Willmore BD, Schnupp JW, King AJ (2012) Spectrotemporal contrast kernels for neurons in primary auditory cortex. J Neurosci 32: 11271–11284. doi: 10.1523/JNEUROSCI.1715-12.2012 PMID: 22895711

55. Murray RF, Bennett PJ, Sekuler AB (2002) Optimal methods for calculating classification images: weighted sums. J Vis 2: 79–104. doi: 10.1167/2.1.6 PMID: 12678598

56. Neri P (2010) How inherently noisy is human sensory processing? Psychon Bull Rev 17: 802–808. doi: 10.3758/PBR.17.6.802 PMID: 21169572

57. Geisler WS (2011) Contributions of ideal observer theory to vision research. Vision Res 51: 771–781. doi: 10.1016/j.visres.2010.09.027 PMID: 20920517

58. Joosten ERM, Neri P (2012) Human pitch detectors are tuned on a fine scale, but accessed on a coarse scale". Biological Cybernetics 106: 465–482. doi: 10.1007/s00422-012-0510-x PMID: 22854977

59. Westwick DT, Kearney RE (2003) Identification of Nonlinear Physiological Systems. Piscataway: New Jersey: Wiley IEEE Press.

60. Ahumada AJ (2002) Classification image weights and internal noise level estimation. J Vis 2: 121–131. doi: 10.1167/2.1.8 PMID: 12678600

61. Green DM, Swets JA (1966) Signal Detection Theory and Psychophysics. New York: Wiley.

62. Fitzgerald JD, Rowekamp RJ, Sincich LC, Sharpee TO (2011) Second order dimensionality reduction using minimum and maximum mutual information models. PLoS Comput Biol 7: e1002249. doi: 10.1371/journal.pcbi.1002249 PMID: 22046122

63. Neri P, Heeger DJ (2002) Spatiotemporal mechanisms for detecting and identifying image features in human vision. Nat Neurosci 5: 812–816. PMID: 12101403

64. Marmarelis VZ (2004) Nonlinear Dynamic Modeling of Physiological Systems. Piscataway: New Jersey: Wiley IEEE Press.

65. Brunelli R, Poggio T (1997) Template matching: matched spatial filters and beyond. Pattern Recognition 30: 751–768. doi: 10.1016/S0031-3203(96)00104-5

66. Hauske G, Wolf W, Lupp U (1976) Matched filters in human vision. Biol Cybern 22: 181–188. doi: 10.1007/BF00365084 PMID: 953076

67. Simpson WA, Manahilov V (2001) Matched filtering in motion detection and discrimination. Proc Biol Sci 268: 703–709. doi: 10.1098/rspb.2000.1406 PMID: 11321058

68. Solomon JA (2002) Noise reveals visual mechanisms of detection and discrimination. J Vis 2: 105–120. doi: 10.1167/2.1.7 PMID: 12678599

69. Abbey CK, Eckstein MP (2006) Classification images for detection, contrast discrimination, and identification tasks with a common ideal observer. J Vis 6: 335–355. doi: 10.1167/6.4.4 PMID: 16889473

70. Bussgang JJ (1952) Cross-correlation functions of amplitude-distorted Gaussian signals. MIT Research Laboratory Electricity Technical Report 216: 1–14.

71. Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. Network 12: 199–213. PMID: 11405422

72. Oberfeld D (2008) Does a rhythmic context have an effect on perceptual weights in auditory intensity processing? Can J Exp Psychol 62: 24–32. doi: 10.1037/1196-1961.62.1.24 PMID: 18473626

73. Tjan BS, Nandy AS (2006) Classification images with uncertainty. J Vis 6: 387–413. doi: 10.1167/6.4.8 PMID: 16889477

74. Bracewell RN (1965) The Fourier Transform and its Applications. Singapore: McGraw-Hill.

75. Funahashi K (1989) On the approximate realization of continuous mappings by neural networks. Neural Networks 2: 183–192. doi: 10.1016/0893-6080(89)90003-8

76. Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. Neural Networks 2: 359–366. doi: 10.1016/0893-6080(89)90020-8

77. Nagai M, Bennett PJ, Sekuler AB (2008) Exploration of vertical bias in perceptual completion of illusory contours: Threshold measures and response classification. J Vis 8: 1–17. doi: 10.1167/8.7.25

78. Neri P (2011) Global properties of natural scenes shape local properties of human edge detectors. Front Psychol 2: 172. doi: 10.3389/fpsyg.2011.00172 PMID: 21886631

79. Neri P (2014) Semantic control of feature extraction from natural scenes. J Neurosci 34: 2374–2388. doi: 10.1523/JNEUROSCI.1755-13.2014 PMID: 24501376

80. Murray RF, Bennett PJ, Sekuler AB (2005) Classification images predict absolute efficiency. J Vis 5: 139–149. doi: 10.1167/5.2.5 PMID: 15831074

81. Füllgrabe C, Moore BCJ (2007) A perceptual-learning investigation of auditory amplitude-modulation detection: Testing the existence of frequency-selective mechanisms in the temporal-envelope domain. SfN Posters.

82. Sek A, Moore BC (2003) Testing the concept of a modulation filter bank: the audibility of component modulation and detection of phase change in three-component modulators. J Acoust Soc Am 113: 2801–2811. doi: 10.1121/1.1564020 PMID: 12765397

83. Klein DJ, Depireux DA, Simon JZ, Shamma SA (2000) Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. J Comput Neurosci 9: 85–111. doi: 10.1023/A:1008990412183 PMID: 10946994

84. Berg BG, Green DM (1990) Spectral weights in profile listening. J Acoust Soc Am 88: 758–766. doi: 10.1121/1.399725 PMID: 2212301

85. Shub DE, Richards VM (2009) Psychophysical spectro-temporal receptive fields in an auditory task. Hear Res 251: 1–9. doi: 10.1016/j.heares.2009.02.007 PMID: 19249339

86. Ardoint M, Mamassian P, Lorenzi C (2007) Internal Representations of Amplitude Modulation Revealed by Reverse Correlation. ARO Abstracts 358.

87. Ahumada AJ, Marken R, Sandusky A (1975) Time and frequency analyses of auditory signal detection. J Opt Soc Am A Opt Image Sci Vis 57: 385–390.

88. Eckstein MP, Shimozaki SS, Abbey CK (2002) The footprints of visual attention in the Posner cueing paradigm revealed by classification images. J Vis 2: 25–45. doi: 10.1167/2.1.3 PMID: 12678595

89. Neri P (2004) Attentional effects on sensory tuning for single-feature detection and double-feature conjunction. Vision Res 44: 3053–3064. doi: 10.1016/j.visres.2004.07.014 PMID: 15474578

90. Ponsot E (2015) Global loudness processing of time-varying sounds. Paris 6 University.

91. Neri P (2009) Nonlinear characterization of a simple process in human vision. J Vis 9: 1–29. doi: 10.1167/9.12.1 PMID: 20053092

92. Neri P (2010) Visual detection under uncertainty operates via an early static, not late dynamic, non-linearity. Front Comput Neurosci 4: 151. doi: 10.3389/fncom.2010.00151 PMID: 21212835

93. Taylor CP, Bennett PJ, Sekuler AB (2009) Spatial frequency summation in visual noise. J Opt Soc Am A Opt Image Sci Vis 26: 84–93. doi: 10.1364/JOSAA.26.000B84

94. Taylor CP, Bennett PJ, Sekuler AB (2014) Evidence for adjustable bandwidth orientation channels. Front Psychol 5: 578. doi: 10.3389/fpsyg.2014.00578 PMID: 24971069

95. Kilgard MP, Merzenich MM (1998) Plasticity of temporal information processing in the primary auditory cortex. Nat Neurosci 1: 727–731. doi: 10.1038/3729 PMID: 10196590

96. Kilgard MP, Pandya PK, Vazquez J, Gehi A, Schreiner CE, et al. (2001) Sensory input directs spatial and temporal plasticity in primary auditory cortex. J Neurophysiol 86: 326–338. PMID: 11431514

97. Fitzgerald MB, Wright BA (2005) A perceptual learning investigation of the pitch elicited by amplitude-modulated noise. J Acoust Soc Am 118: 3794–3803. doi: 10.1121/1.2074687 PMID: 16419824

98. Fitzgerald MB, Wright BA (2011) Perceptual learning and generalization resulting from training on an auditory amplitude-modulation detection task. J Acoust Soc Am 129: 898–906. doi: 10.1121/1.3531841 PMID: 21361447

99. Sabin AT, Eddins DA, Wright BA (2012) Perceptual learning evidence for tuning to spectrotemporal modulation in the human auditory system. J Neurosci 32: 6542–6549. doi: 10.1523/JNEUROSCI.5732-11.2012 PMID: 22573676

100. Korenberg MJ, Hunter IW (1986) The identification of nonlinear biological systems: LNL cascade models. Biol Cybern 55: 125–134. PMID: 3801533

101. Jaaskelainen IP, Ahveninen J, Belliveau JW, Raij T, Sams M (2007) Short-term plasticity in auditory cognition. Trends Neurosci 30: 653–661. doi: 10.1016/j.tins.2007.09.003 PMID: 17981345

102. Shamma S, Fritz J (2014) Adaptive auditory computations. Curr Opin Neurobiol 25: 164–168. doi: 10.1016/j.conb.2014.01.011 PMID: 24525107

103. Carlyon RP, Shamma S (2003) An account of monaural phase sensitivity. J Acoust Soc Am 114: 333–348. doi: 10.1121/1.1577557 PMID: 12880045