version 21 October 2025

Predictive coding in the auditory brainstem

Alain de Cheveigné,
Laboratoire des Systèmes Perceptifs,
Centre National de la Recherche Scientifique,
École normale supérieure (Paris),
and University College London

Abstract.

Predictive coding is an influential concept in sensory and cognitive neuroscience. It is often understood as involving top-down prediction of bottom-up sensory patterns, but the term also applies to feed-forward predictive mechanisms, for example in the retina. Here, I discuss a recent model of low-level predictive processing in the auditory brainstem that is of the feed-forward flavor. Past sensory input from the cochlea is delayed and compared with the current input, with a delay that is tuned with the objective of minimizing prediction error. This operation can be performed uniformly across peripheral channels, or independently within each peripheral channel with parameters local to that channel. The result is a sensory representation that is invariant to a certain class of masking sounds (harmonic, quasi-harmonic, or spectrally sparse), thus contributing to Auditory Scene Analysis. The purpose of this paper is to discuss that model in the light of predictive coding, and examine how it might fit within a wider hierarchical model that supports the perceptual representation of objects and events in the world.

I. Introduction

In predictive coding, an incoming sensory pattern is compared with a predicted pattern, and the outcome of this comparison – the error – serves both to tune the predictive model, and as an input for the next stage of processing. Predictive coding is an influential idea in sensory and cognitive neuroscience (Attneave 1954, Barlow 1961, Srinivasan et al 1982, Rao and Ballard 1999, Clark 2013; Friston 2018). A perfectly predictable pattern carries no information beyond that carried by the predictor (it is redundant). Even if the prediction is imperfect, the error may be less costly to represent and/or easier to process and/or more informative than the raw input. Furthermore, the *parameters* of the predictive model, fit to the stimulus, characterize the stimulus, while the model itself, tuned over a longer time scale, captures the statistics of the sensory environment. Predictive coding is one of a constellation of related ideas, including unconscious inference (Alhacen 1030; Helmholtz 1867), efficient coding (Shannon 1948), epistemological automaton (MacKay 1956), redundancy reduction (Barlow 1961), forward models in motor control (Wolpert et al 1995), reverse hierarchy (Hochstein and Ahissar 2002), efferent copy (Crapse 2008), compression (Schmidhuber, 2009), homeostasis (Friston 2010), Bayesian inference (Aitchison and Lengyel 2017), and so on. In each case, an incoming observation is compared against an internally-generated template, and the mismatch used to adjust the model and remove redundancies.

However, "predictive coding" covers two quite different concepts (Spratling 2015). Given a hierarchy of processing levels, prediction can occur in the *ascending* direction (bottom-up Fig. 1 A) or *descending* direction (Fig. 1 B). An example of the former is center-surround receptive fields in the retina (Srinivasan et al 1982; Hosoya et al 2005), an example of the latter is the recurrent processing proposed for the visual cortex by Rao and Ballard (1999). Top-down prediction is usually assumed to operate within a *hierarchy* of predictive stages ordered from concrete to abstract (Fig. 1C) (e.g., Rao and Ballard 1999; Spratling 2017; Friston 2018), with messages exchanged in both directions. Bottom-up prediction is embodied in the Linear Predictive Coding (LPC) method popular for spectral estimation in speech processing (Markel and Gray 1976), while top-down prediction is related to Bayesian models of perception that combine a prior (parameters of the predictive model) and evidence (the sensory pattern) to infer the causes of sensory input. Both flavors of predictive coding could play a role within the auditory system (Heilbron and Chait 2018).

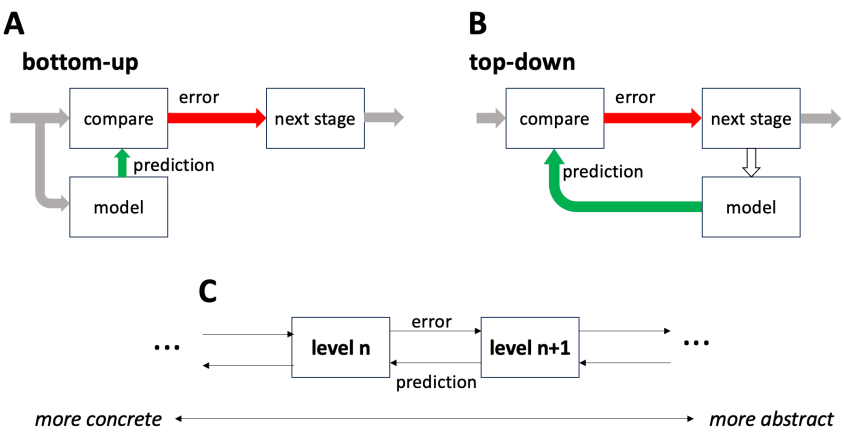

Fig. 1. Two forms of predictive coding. A: The prediction is produced by a feed-forward model (bottom-up) fed from incoming patterns. B: The prediction is produced by a feed-back model (top-down) fed from a higher-level representation. C: Hierarchy of predictive coding stages ordered from concrete to abstract.

Here, I reinterpret a recent model of low-level auditory processing as a form of predictive coding of the bottom-up flavor: low-level sensory input is compared, at each instant, with a prediction derived from its previous values. I then take advantage of this concrete, bottom-up model to explore issues that arise when considering the alternative hypothesis of top-down predictive coding, and discuss how the two approaches might complement each other.

II. The in-channel cancellation model

In-channel cancellation (ICC) is a hypothetical processing mechanism for auditory scene analysis, to help address a listener's need to handle “cocktail party” situations involving complex acoustic scenes with multiple concurrent speakers or polyphonic music. In particular, it accounts for our ability to ignore masking sounds that are *harmonic*, *quasi-harmonic,* or *spectrally sparse*. The structure and principle of the ICC model is outlined here, more details and references are provided elsewhere (de Cheveigné 2023a).

Patterns of acoustic vibration are transduced within the cochlea and carried by an array of auditory nerve fibers to the auditory brainstem (Palmer and Rees 2010), an auditory analog of the retina in terms of level of processing (Sitko and Goodrich 2021). The cochlea implements a

bank of frequency-selective filters, with each channel transduced by one of ~3000 hair cells into a pattern of spikes carried by a set of ~10 fibers (in human). In the ICC model, spikes within a channel feed a "neural cancellation filter" that consists of a gating neuron with one excitatory synapse fed directly, and one inhibitory synapse fed via an adjustable delay of duration τ_k where k is the index of the channel (Fig. 2A, red). The properties of this neuron are assumed to be such that every direct spike is transmitted *unless* a spike arrives simultaneously on the delayed pathway, within some short time window of coincidence. The effect is to modify the statistics of the spike train, depleting it of inter-spike intervals of duration close to τ_k.

The delay parameter τ_k is determined automatically based on a criterion of minimum output, and it tends to lock on to the period (or pseudo period) of the signal that dominates the channel. In the presence of a masking sound that is *harmonic*, *quasi-harmonic*, or *spectrally sparse*, the filter cancels the correlates of that sound, potentially allowing components of a weaker concurrent sound to emerge. An earlier model (harmonic cancellation, HC, de Cheveigné 1993; 2021) used the same delay τ in all channels, making it effective only for the narrower class of harmonic maskers, whereas the ICC model is more flexible. Together, the HC and ICC models can account for a range of psychophysical results that involve masking with such sounds (as reviewed by de Cheveigné 2021, 2023a). The neural circuitry required by these models (fast excitatory-inhibitory interaction) is similar to that hypothesized by the classic equalization-cancellation (EC) model of binaural unmasking (Durlach 1963), and the within-channel version of that model (Culling and Summerfield 1995), for which physiological support has been found within the lateral superior olive (LSO) (Biederbeck 2018; Franken et al 2021).

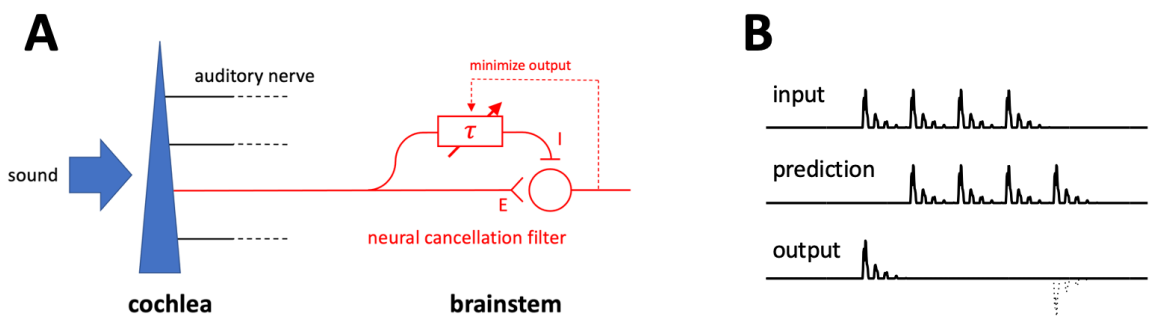

Fig. 2. A: Schematic representation of the cochlea as a bank of bandpass filters followed by a neural cancellation filter (one channel shown). The neural cancellation filter depletes the auditory-nerve spike train of intervals equal to τ, thus attenuating correlates of sound components with that period. B: instantaneous spike probability at the excitatory synapse (top), inhibitory synapse (middle) and at the

output of the gating neuron (bottom) in response to a short periodic stimulus. Only the first period reaches the output.

In a linear approximation, the effects of this neural filter can be described by a delay-and-subtract filter of impulse response $h(t) = \delta(t - \tau_k) - \delta(t)$, where $\delta(t)$ is the Dirac function, followed by a half-wave rectifier that embodies the fact that probability is non-negative (de Cheveigné 1993, 2023a). If the delay equals the period of a sound, the response to that sound is suppressed, as illustrated in Fig. 2B: the predicted signal (middle) is subtracted from the input signal (top), suppressing all but the first period (bottom). In the frequency domain, the effect is to suppress all harmonics of $1/\tau$ (Fig. 3A, top). The cascade of cochlear and cancellation filters has a frequency characteristic that attenuates both power that is distant from the characteristic frequency (CF) *and* power that is harmonic with period τ (Fig. 3A, bottom).

As an example, Fig. 3B shows a cochleagram, a time-frequency representation based on a modeled cochlear filterbank (Slaney 1983) in response to a short snippet of speech ("wow cool", male voice). Panel C shows the gain of the time-varying cancellation filters applied to the output of the filterbank, also as a function of channel and time. The gain is reduced (by as much as -20dB) within the time-frequency pixels corresponding to high-amplitude components of this sound. Thus, if this speech sound were to mask a weaker target, that target might be detected or partially recovered by applying this array of time-varying filters to suppress the masker. This suggests that this hypothetical mechanism might effectively deal with highly non-stationary “real-world” sounds such as speech.

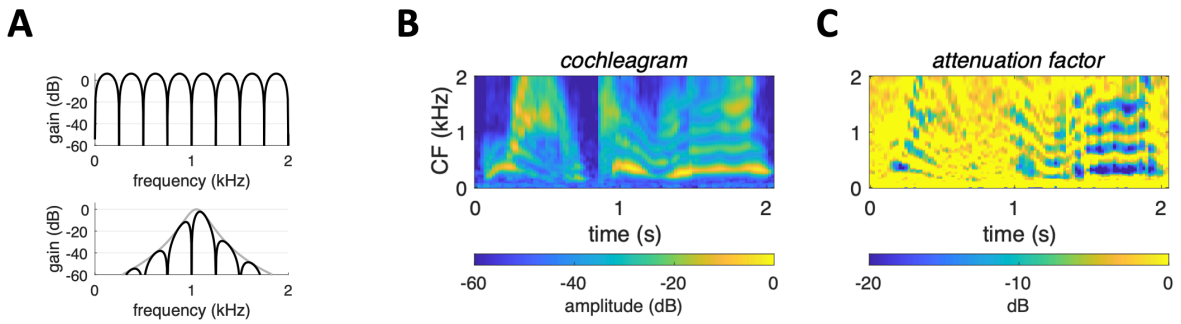

Fig. 3. A top: transfer function of delay-and-subtract filter for τ=2.5 ms, bottom: transfer function of cascade of this filter with a cochlear filter with CF=1.05 kHz. B: Cochlear excitation pattern in response to a short snippet of speech ("wow, cool!"). C: Attenuation factor of the cancellation filter array at each time and CF for the speech snippet in A.

III. ICC as predictive coding?

A. Bottom-up

The structure of the ICC model (Fig. 2) evokes predictive coding of the bottom-up flavor, similar to the model of Srinivasan et al (1982) in which the center of a retinal patch is predicted from its neighbors and/or previous values. Here, the incoming auditory nerve spike probability is predicted from its past values (Fig. 2, left). The model is also reminiscent of LPC (Gray and Markel 1996) in which each new sample of a sampled representation is predicted as a linear combination of N previous samples (where N is the order of the LPC model). The ICC model achieves both *redundancy reduction*, in that the cancellation filter strips the representation of redundant, periodic patterns (Fig. 2B), and *redundancy exploitation* (Attneave 1954; Barlow 1961), in that the optimal delay is the period of the sound, predictive of its pitch (de Cheveigné 1998; 2023). The period is an *abstract* quantity and, in this sense, ICC may contribute to the concrete-to-abstract hierarchy depicted in Fig. 1C (see also Sect. IV, B).

B. Top-down

There is a very different class of models in which incoming patterns are predicted from abstract internal models, according to a principle akin to "analysis by synthesis" (Helmholtz 1867; Braitenberg 1984; Kawato et al 1993; Friston 2002, 2018; Spratling 2017). There is keen interest in such top-down models, in particular due to theoretical properties of certain formulations (Friston 2010, 2018; Friston et al 2009; Bastos et al 2013; see also Sims 2017; Litwin and Miłkowski 2020; Andrews 2021; Williams 2021; Hemmatian 2023). Re-casting ICC as a top-down predictive model might allow us to "hitch onto that wagon" and reap its theoretical benefits. An additional motivation is the need to explain massive efferent pathways that link most stages of the auditory system between cortex and cochlea (Schofield 2012; Romero and Trussell 2022). Together, these reasons encourage us to consider whether it is possible to translate, or insert, ICC into a top-down model. Here I review three simple schemes, numbered for easy reference.

(1) Splitting the ICC model of Fig. 2A in two, parameters τ_k can be treated as message from a higher level that evaluates the filter output, to a lower level that implements the filter. This reformulation makes the model seem "top-down", but with little added value.

(2) The predictive filter of Fig. 4A can be modified by deriving the "prediction" from its output (Fig. 4B) rather than input, thus making it "top-down." A moment's thought tells us this is a bad idea: if prediction were effective, i.e., the error was zero, there would be nothing to feed the prediction. If we persisted, we would discover that the filter has an infinite, non-decreasing impulse response (Fig. 5B, bottom). This could be alleviated by introducing a gain factor in the feedback loop, with a value less than one, but the impulse response would still be long relative to that in Fig. 5A for no clear benefit.

(3) A third scheme, inspired from the concept of "analysis by synthesis" is to derive the prediction via a generative model fed from a higher, abstract stage (Fig. 5C). This too cannot work, because the temporal (or phase) information needed to correctly align the prediction with the incoming pattern is not available. As pointed out by Mumford (1992), the higher, abstract stage speaks a different "language" from the lower concrete layer. Abstraction is a many-to-one transform, not invertible, and thus an abstract trace cannot translate unambiguously to a concrete one.

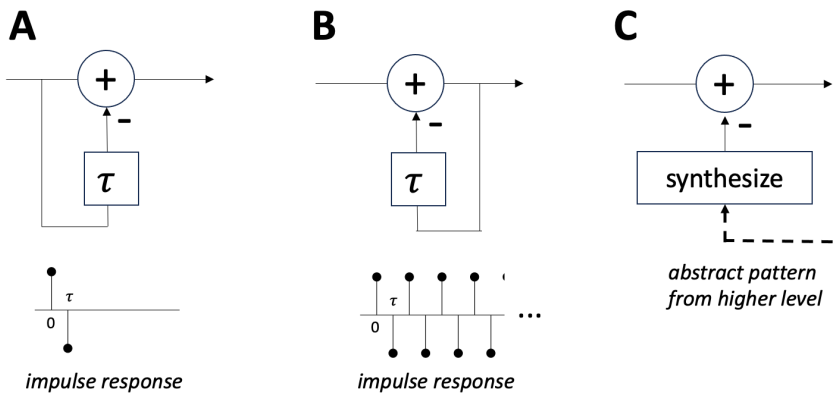

Fig. 4. A: The bottom-up predictive coding model embodied by ICC implements a filter with a very compact impulse response that suppresses periodic input. B: Deriving the prediction from from the output leads to a filter with an infinite, non-decreasing impulse response. C: Top-down prediction from an abstract trace cannot work because temporal (or phase) information to align it with the input is lacking.

Interim summary. The ICC model looks like a text-book example of bottom-up predictive coding, but straightforward attempts to reformulate it as a top-down model are not fruitful. The

rest of this paper tries to better understand why, and salvage the idea in view of its multiple benefits invoked at the beginning of this section.

IV. Let's try again

A. Top-down predictive coding revisited

Theories of perception mostly agree on a hierarchy of representations going from concrete (sensory) to abstract (perceptual, or conceptual), with messages passed between levels in the ascending and (for predictive coding) descending directions (Fig. 1C). As pointed out by Mumford (1992), messages passed between different levels of abstraction require *translation*, which could occur *before* the message is sent, or *after* it is received. Referring to Fig. 1C, the ascending error might be translated at either level n or level $n + 1$, and the same options are available for the descending prediction, implying four different interpretations of Fig. 1C, each plausible but with different implications and constraints. This is an example of questions that may arise when trying to make sense of predictive coding.

Top-down translation presents two difficulties. One is the difference in representation format (e.g., a retinotopic sensory pattern in a sensory area versus an abstract representation within a face area), the other is that abstraction is not invertible. For the first, Mumford (1992) suggested that a "template neuron" might code a detailed sensory pattern in the form of synaptic weights. Bottom-up translation would be performed by weights between axons from neurons in level n onto the dendrites of a template neuron in level $n + 1$, whereas top-down translation from that template neuron is performed by weights of on its axonal terminals in layer n. Incidentally, both sets of weights define the same "template," which raises the question of how they are learned in sync. For the second difficulty, he proposed that a *family* of templates collectively represents the set of inputs that map to the abstract trace. Top-down translation then involves comparing all of these templates, sequentially or in parallel, to the input pattern, the error (or residual) being presumably the vector difference between input and closest template.

Mumford (1992) further suggested that each template might be "flexible", for example defined by a mean and covariance (analogous to a Gaussian mixture model), or alternatively as a collection of extreme values (analogous to support vectors). This adds flexibility to the representation of the distribution of observations at level n that map to an abstract trace at level $n + 1$, easing bottom-up translation, but it is less clear how the residual is defined. Is it the difference

from the mean of the nearest template? Is it the difference from the nearest "support vector"? He further suggested that observations might be *transformed* before prediction (for example by time-warping), increasing flexibility. As yet another alternative, Mumford (1992) also suggested that each template might be formed as linear combinations of basis-functions, on the lines of Poggio's (1990) "Hyper Basis Function" theory. An abstract trace is then a distribution of sets of *coefficients* of such combinations, flexibility coming from the spread of this distribution rather than from the diversity of a set of templates, and/or their individual flexibilities.

In a similar direction, Rao and Ballard (1999) suggested that an abstract trace be defined based on a linear combination of basis-functions u_j, plus a stochastic noise term:

$$\mathbf{I} = f\left(\sum_{j=1}^{k} u_j r_j\right) + \mathbf{n} \qquad (1)$$

where **I** is an input that maps to the abstract trace, f is a function (e.g., sigmoid), r_j is a coefficient ("cause") and **n** is a noise process. Here too, the abstract trace is defined as a set of coefficients, but flexibility is now determined by the noise term, rather than by the distribution of admissible sets of coefficients. Friston's (2009) "free energy principle" requires further that this noise term be *Gaussian*: in the "Laplace approximation," an abstract trace translates to (or is translated from) an ensemble of observations that is distributed normally in some space. The prediction residual ("error" in Fig. 1) is defined as the vector difference between observation and mean, possibly weighted by "precision" derived from the covariance matrix of the abstract trace (e.g., Friston 2002), or of the observation (e.g., Heldman and Friston 2010), or perhaps both. Redundancy reduction is achieved if fewer bits are required to represent the narrower range of the error than of the observation.

The Gaussian assumption is restrictive compared to the arbitrarily-shaped distribution of Mumford (1992), but Friston (2009, p298) argues that it can easily be satisfied by applying the appropriate *non-linear transform* to the observations. This shifts the "heavy-weight lifting" of modeling a complex distribution (e.g., visual patterns that match the concept "dog") to that bottom-up transform. ICC might be interpreted as implementing one such transform, as it projects a diverse set of (quasi-)harmonic sounds to a space where they are distributed near zero.

Spratling (2017) pointed out the similarity between bottom-up predictive coding (e.g., LPC) and top-down predictive coding (e.g., Rao and Ballard 1999) which can both be expressed by an equation analogous to Eq. 1, in which the basis-functions u_j are intervals of recent

observations for LPC, while for top-down predictive coding they belong to an internal generative model, fixed or slowly-varying. This opens the possibility of *hybrid* ascending and descending prediction, again suggesting a bottom-up component in top-down predictive coding.

If one insists on a "free-form" distribution (rather than Gaussian), the "error" cannot take the form of a simple difference. For a single observation, it might be quantified by a measure of *surprise* (a scalar), and for a set of observations (or an uncertain observation), by a measure such as *KL divergence*, $D_{KL}(\mathcal{O} \parallel \mathcal{A})$, where $\mathcal{O}$ represents the distribution of the observation(s) and $\mathcal{A}$ that of the abstract trace (e.g., Pinchetti et al 2002). If divergence is small, the abstract trace is a good model of the observations, and so the error message can be empty. If divergence is large, a useful message might consist of the surprising observations themselves, or some transform thereof (e.g., statistic). Redundancy reduction is achieved if the latter event is rare. Note, however, that KL divergence requires *parametrization* of the distributions, which removes some (but not all) "freedom" from the free-form assumption.

It is worth noting also that KL divergence is not symmetric. $D_{KL}(\mathcal{O} \parallel \mathcal{A}) \approx 0$ means that *the observations fit the abstraction*, whereas $D_{KL}(\mathcal{A} \parallel \mathcal{O}) \approx 0$ means that *the abstraction fits the observations*. The former might be true and the latter not if recent observations are more tightly distributed than expected, in which case a new abstraction might be considered, possibly nested within the old (e.g., a category "poodle" within the category "dog"). For this to occur, the distribution $\mathcal{O}$ must be reliable, which may require gathering multiple observations over some interval of time. This, in turn, requires *chunking* the observation stream, yet another bottom-up process. This suggests a different narrative: predictive coding achieves redundancy reduction for the purpose of *efficient memory* of the stream of observations (de Cheveigné 2025).

Finally, in the "forward-reverse optics" model of Kawato et al (1993), a complex visual scene is analyzed by a combination of two processes, one top-down, the other bottom-up. The former attempts to generate retinal patterns from an internal representation of the scene, as in other analysis-by-synthesis schemes. The authors refer to it as a "direct optics" model because it mimics, within the brain, the optical pathway from scene to retina, along the lines of a forward model in motor control (Wolpert and Kawato 1998). The bottom-up process, referred to as "reverse optics," attempts to translate retinal patterns into a representation of the scene, as in a classic computer vision system. This second, bottom-up, process does not need to perform

perfectly, and indeed, it cannot, given the ambiguity of scenes with occlusion, etc. Its role is to provide an initial "guess" of the scene that allows the search space of the top-down process to be pruned. Transposing to audition, we could hypothesize that ICC contributes to a "reverse acoustics" model, applicable to parse scenes that include harmonic or quasi-harmonic sources.

In sum:

(1) The term "predictive coding" has been applied to a variety of very different concepts.

(2) "Analysis-by-synthesis" is problematic because abstraction is not invertible. "Prediction error" may make sense in terms of surprise or KL divergence, but not in terms of a vector *difference* unless restrictive assumptions apply. While a *property* of an observation may be predicted from an abstract trace, the value of the observation may not.

(3) Top-down predictive coding relies on bottom-up processing for much of the "heavy lifting," in particular to perform abstraction between layers of a hierarchical representation.

B. In-channel cancellation contributes to abstraction

ICC was earlier presented as a kind of adaptive filter, the output of which has the same level of abstraction as its input (i.e., a rapidly-varying pattern of firing probability). Here, I look at how it might produce slowly-evolving abstract quantities, thus contributing to a first stage in a hierarchy of abstractions.

A first such quantity is the *period*, estimated by fitting the ICC or HC model with a criterion of minimum output (Fig. 5A). It is an abstraction in that different periodic sounds may share the same period. A wider class of sounds are *approximately periodic* in various ways: limited duration, slow amplitude or frequency changes, periodicity within a restricted spectral region, etc. (as in the speech example of Fig. 3B). For these, a measure of *periodicity* may be informative, for example the cancellation filter output/input power ratio (as plotted in Fig. 3C for individual channels over time). Period and periodicity are perceptually relevant, the former as a cue to *pitch*, the second as a possible contributor to an aspect of *timbre*.

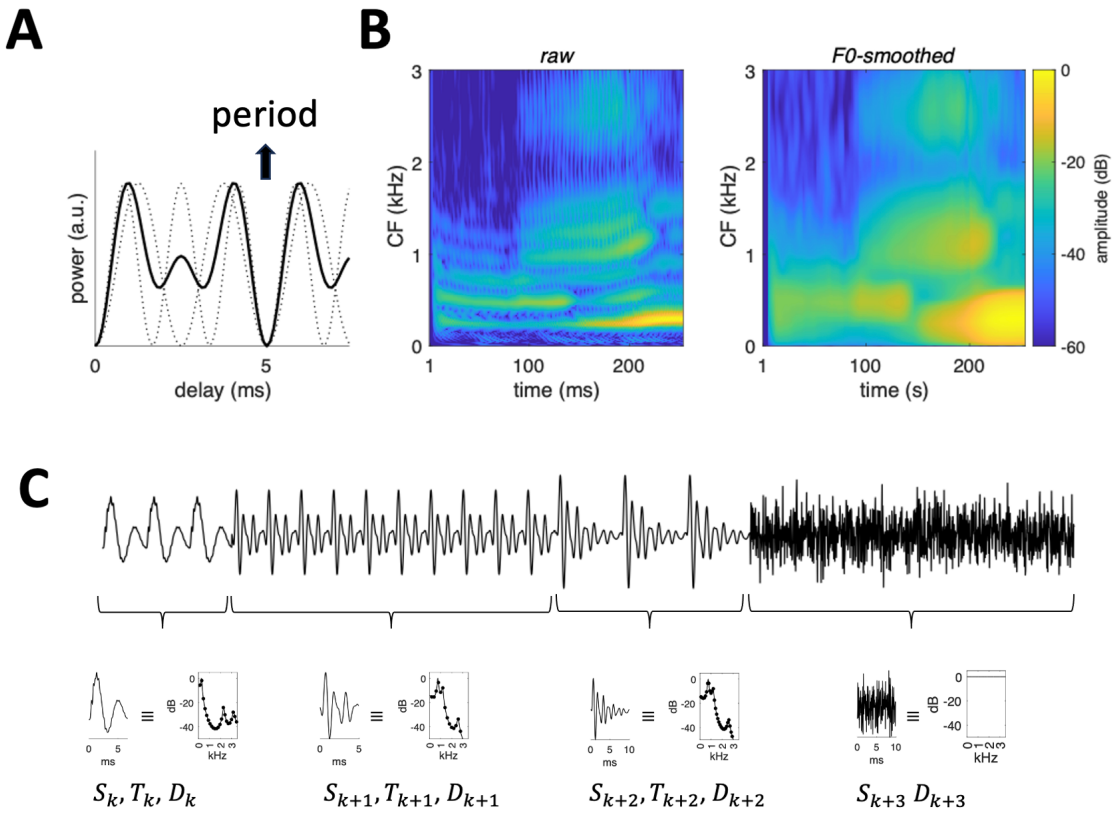

Fig. 5. ICC as abstraction. A: The period of a 200 Hz complex tone is estimated as the first non-zero lag that yields zero output, common over channels. The black line is for CF=500 Hz (this channel responds to both harmonics 2 and 3), the dotted lines are for CF=400 Hz (harmonic 2) and CF=600 Hz (harmonic 3). B: Optimal smoothing of a time-frequency representation. Left: raw cochleagram, right: cochleagram smoothed in time (kernel duration $T = 1/F_0$) and frequency (kernel width $F_0 = 1/T$). C: Chunking for redundancy reduction. Each stationary segment might be coded by the *waveform during one period* (or its Fourier transform), the value of the *period* (for periodic sounds) and the *duration* of the segment.

Another abstraction is the *spectrum*, which comes in many variants according to how it is calculated. A time-frequency representation can be calculated based on the short-term Fourier transform (*spectrogram*), or as instantaneous power at the output of a cochlear filterbank (*cochleagram*, e.g. Fig. 3B), or instantaneous probability at the output of a more realistic peripheral model. Again, it is abstract in that many different sounds share the same spectrum or spectrogram, and perceptually relevant as a predictor of *timbre*, for example that of a vowel. Unfortunately, this representation is not invariant relative to the fundamental frequency $F_0 = 1/T$ of a sound: individual harmonics may emerge at low frequencies, and a pulsed temporal structure may emerge at high frequencies (Fig. 5B, left), whereas timbre is usually thought to reflect the smooth time-frequency *envelope*. Here again, the ICC/HC model contributes by allowing

period-adaptive smoothing to approximate that envelope (Fig. 5B, right). This is the principle of the well-known STRAIGHT method of spectro-temporal estimation (Kawahara et al 1999). It offers yet another level of abstraction, as sounds with different spectra/spectrograms (due to different fundamental frequency) map to the same smooth spectro-temporal envelope.

Yet another popular abstraction is the *modulation spectrum* (Viemeister 1979; Dau et al 1997; Jepsen 2003; Joris et al 2004; Anden and Mallat 2014; Koumura et al 2019), defined as the spectrum of the *temporal envelope* within peripheral channels, possibly averaged over channels. It accounts for many aspects of sound, in particular speech (Elhilali et al 2003; Hermansky 2010) and, together with other statistics, captures the perceptually-relevant *texture* of many natural sounds (McDermott and Simoncelli 2011; McDermott et al 2013). Local fluctuations may serve also as cues to spectrally local features (Carney 2024). The quality of this abstraction is improved if the temporal envelope is first smoothed with a kernel tailored to the period of the sound (as described in the previous paragraph), to factor out periodicity-related fluctuations from slower perceptually-relevant modulations, thus avoiding interference between the two (Stein et al 2005).

Related to abstraction is *redundancy reduction*, and *summarization*. A benefit of detecting periodicity and estimating the period is that periodic segments can be cheaply represented by the waveform over one period, or its spectral equivalent (Fig. 5C). The ICC model contributes by detecting such periodic segments and their boundaries, and estimating the periods. This combination of *segmentation* and summarization may be seen as an initial layer in a hierarchical memory representation (de Cheveigné 2025).

This section hinted at various ways in which ICC *might* contribute to the elaboration of abstract representations of sound that *might* contribute to some of the top-down prediction schemes reviewed earlier. These ideas are presented schematically, obviously a more precise formulation is required to take them seriously as models of sensory processing.

V. Discussion

This paper presented the low-level auditory processing model ICC as a form of *bottom-up* predictive coding. However, much of the appeal of the concept of predictive coding is specific to its *top-down* flavor; a reader who expected that flavor might feel short-changed. For this

reason, I also investigated how ICC might relate to, or contribute to top-down predictive coding. At this point, the following hypothetical roles emerge:

1. The ICC model contributes an early link to the chain of abstractions postulated by most (all?) predictive coding models. One cannot achieve a hierarchy of abstraction if the output of each stage belongs to the same concrete space as its input, as is the case if each level sends a residual to the next. In contrast, the *parameters* of a model fit to a concrete pattern might be predicted, but this requires that that model (e.g., ICC) is defined.
2. The ICC model contributes to acoustic scene analysis, by its ability to suppress sounds with a harmonic, quasi-harmonic, or sparse spectrum. In this way, it might contribute to a “reverse acoustics” model on the lines of Kawato et al (1993).
3. The ICC model transforms input patterns that belong to the class of periodic sounds in such a way that they are distributed around zero, plausibly according to a Gaussian distribution. In this, it might play the role of a "Laplacifier” transform so that a predictive coding model on the lines of Friston (2009) meets its assumptions (Sect. IV, A).
4. The parameters of the ICC model (delays τ_k) might be under top-down control. For example, a musical context might prime a listener as to the next pitch, or a decision to listen to source might translate into top-down imposition of delays that suppress competing sources. Multiple top-down parameters might be tried in sequence (or in parallel) to find one for which ICC best “fits” the input.
5. If predictive coding is reinterpreted as forming a compact statistical model of the stream of observations, ICC can contribute to the initial “chunking” by detecting intervals of quasi-periodicity and their boundaries, and by removing redundancy within such intervals. The magnitude of the ICC output serves as a measure of surprisal relative to the assumption of periodicity.

A stumbling block is the notion of prediction error as a *difference*, or *residual*. It would seem that Mumford's (1992) warning about the need for translation between levels of abstraction was not heeded (including by Mumford himself). Objections to that notion are logical, as reviewed earlier, and also empirical. In a recent study, Furutachi et al (2024; Lowet and Uchida 2024) probed responses of putative “error neurons” in mouse primary visual cortex. With conditions carefully designed to test the difference hypothesis, they found that a surprising observation elicited an *enhanced* stimulus-specific response, rather than a difference. It is plausible

that Mumford and others and others were aware of the issue, but that its implications got "lost in translation" when trying to make the model intelligible (as in Fig. 1C).

A prime motivation of top-down predictive coding is the need to explain efferent pathways within the brain. From the roles just enumerated it is not clear why they would need to be massive. One might speculate that massive efferent pathways are required by the need to propose, at level $n + 1$, a large number of alternative abstract explanations to be tested at level n.

Focusing on the auditory brainstem, the working hypothesis of a *single* delayed feedforward pathway within a channel might be over-restrictive. More plausible is a combination of multiple, within-channel predictors (as in LPC), possibly complemented by cross-channel interactions (e.g. Shamma 1984; Nelson and Carney 2004; de Cheveigné and Pressnitzer 2006; Carney 2002, 2024) and binaural interactions (Durlach 1964; Culling and Summerfield 1995; Beiderbeck et al 2018; Franken et al 2021). The multiple parameters involved (weights, delays) might help account for multiple efferent pathways.

An explanatory, reactive, or compressive model can better be fit and updated based on a buffer that extends into the past (Schmidhuber 2009). Predictive coding may contribute to *memory* of past observations (as opposed to a model of causes, decisions or actions) by helping to *segment* the input stream, and more generally helping to code observations in terms of already-acquired statistics, or internal explanatory models of the data. Rather than the message-passing hierarchy of Fig. 1C, memory would form a hierarchy of increasingly abstract statistics (de Cheveigné 2025). "Prediction error" might take one of the following forms: (1) a simple binary flag (fit vs no-fit) or a scalar (surprise), (2) a concise account of how the observation can be made to fit (e.g., "apply delay τ to the ICC model"), (3) a more exhaustive record or summary of the misfitting observations, "memorized" for future resolution.

To summarize, the simple bottom-up ICC model of auditory brainstem processing served as a pretext to review issues related to top-down predictive coding, highlighting the conceptual difficulties involved, and outlining approaches to dealing with them.

Conclusion

The in-channel cancellation model describes a hypothetical form of low-level processing within the auditory brainstem, that offers a text-book example of predictive coding of the feed-forward, bottom-up flavor, similar to earlier proposals for retinal processing (Srivinasan et al 1982). This paper briefly reviewed the model and its relations to bottom-up predictive coding, and examined how it might form part of an early layer within a predictive coding model of the feed-back, top-down flavor. Specifically, its segregation properties might contribute to a bottom-up "analysis" pathway that works hand-in-hand with a top-down predictive "synthesis" pathway (Kawato et al 1993). It could also participate to an early stage of statistical chunking of the sensory input that reorganizes it for more effective predictive coding. It might contribute to transform a non-Gaussian distribution of sensory regularities to better fit the Gaussian approximation assumed by many models of predictive coding.

Acknowledgments

This work was supported by grants ANR-10-LABX-0087 IEC, ANR-10-IDEX-0001-02 PSL, and ANR-17-EURE-0017. Malcolm Slaney made helpful comments on an early draft.

Declaration of competing interests

none

References:

Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, *46*, 219–227. https://doi.org/10.1016/j.conb.2017.08.010

al Haytham I. (1030) Book of optics (in Hatfield 2002).

Anden, J., & Mallat, S. (2014). Deep Scattering Spectrum. *IEEE Transactions on Signal Processing*, *62*(16), 4114–4128. https://doi.org/10.1109/TSP.2014.2326991

Andrews, M. (2021). The math is not the territory: Navigating the free energy principle. *Biology & Philosophy*, *36*(3), 30. https://doi.org/10.1007/s10539-021-09807-0

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3), 183–193. https://doi.org/10.1037/h0054663

Barlow, H B. (1961). Possible principles underlying the transformation of sensory messages. In Rosenblith, W A, *Sensory Communication*. MIT Press.

Barlow, H. (2001). The exploitation of regularities in the environment by the brain. *Behavioral and Brain Sciences*, *24*(4), 602–607. https://doi.org/10.1017/S0140525X01000024

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Bein, O., Gasser, C., Amer, T., Maril, A., & Davachi, L. (2023). Predictions transform memories: How expected versus unexpected events are integrated or separated in memory. *Neuroscience & Biobehavioral Reviews*, *153*, 105368. https://doi.org/10.1016/j.neubiorev.2023.105368

Biederbeck, B., Myoga, M. H., Müller, N. I. C., Callan, A. R., Friauf, E., Grothe, B., and Pecka, M. (2018). "Precisely timed inhibition facilitates action potential firing for spatial coding in the auditory brainstem," Nat. Commun. 9(1), 1771.

Braitenberg, V. (1984). Vehicles: Experiments in synthetic psychology. Cambridge, MA: MIT Press.

Buckley, C. L., Kim, C. S., McGregor, S., & Seth, A. K. (2017). The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*, *81*, 55–79. https://doi.org/10.1016/j.jmp.2017.09.004

Buesing, L., Bill, J., Nessler, B., & Maass, W. (2011). Neural Dynamics as Sampling: A Model for Stochastic Computation in Recurrent Networks of Spiking Neurons. *PLoS Computational Biology*, *7*(11), e1002211. https://doi.org/10.1371/journal.pcbi.1002211

Carney, L. H., Heinz, M. G., Evilsizer, M. E., Gilkey, R. H., & Colburn, H. S. (2002). Auditory Phase Opponency: A Temporal Model for Masked Detection at Low Frequencies. *ACTA ACUSTICA UNITED WITH ACUSTICA*, *88*, 15.

Carney, L. H. (2024). Neural Fluctuation Contrast as a Code for Complex Sounds: The Role and Control of Peripheral Nonlinearities. *Hearing Research*, *443*, 108966. https://doi.org/10.1016/j.heares.2024.108966

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Crapse, T. B., & Sommer, M. A. (2008). Corollary discharge across the animal kingdom. *Nature Reviews Neuroscience*, *9*(8), 587–600. https://doi.org/10.1038/nrn2457

Culling, J. F., and Summerfield, Q. (1995). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am. 98, 785-797.

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. 102, 2892–2905.

de Cheveigné, A. (1998) Cancellation model of pitch perception. *J Acoust Soc Am* 103:1261-1271.

de Cheveigné, A. (1999) Pitch shifts of mistuned partials: a time-domain model. *J Acoust Soc Am* 106:887-897.

de Cheveigné, A., Kawahara, H. (1999) Missing data model of vowel identification. *J Acoust Soc Am* 105:3497-3508.

de Cheveigné, A., Kawahara, H. (2002) YIN, a fundamental frequency estimator for speech and music. *J Acoust Soc Am* 111:1917-1930.

de Cheveigné, A. (2006) Multiple F0 estimation. In: *Computational Auditory Scene Analysis - Principles, Algorithms and Applications* (Wang C, Brown G, eds). IEEE press, Wiley, 45-79.

de Cheveigné, A., and Pressnitzer, D. (2006). "The case of the missing delay lines: Synthetic delays obtained by cross-channel phase interaction," J. Acoust. Soc. Am. 119, 3908–3918.

de Cheveigné, A (2021). Harmonic Cancellation - A Fundamental of Auditory Scene Analysis. *Trends in Hearing*, 25, https://doi.org/10.1177/23312165211041422.

de Cheveigné, A. (2023a) In-channel cancellation, a model of early auditory processing, J. Acoust. Soc. Am., 153, 3361–3372, https://doi.org/10.1121/10.0019752.

de Cheveigné, A. (2023b) Why is the perceptual octave stretched? An account based on mismatched time constants within the auditory brainstem, J. Acoust. Soc. Am. 153, 2600. https://doi.org/10.1121/10.0017978.

de Cheveigné (2025) Graceful forgetting: memory as a process. https://doi.org/10.48550/arXiv.2502.11105

Demany, L., and Semal, C. (2018). "Automatic Frequency-Shift Detection in the Auditory System: A Review of Psychophysical Findings," Neuroscience 389, 30-40. https://doi.org/10.1016/j.neuroscience.2017.08.045.

Den Ouden, H. E. M., Kok, P., & De Lange, F. P. (2012). How Prediction Errors Shape Perception, Attention, and Motivation. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00548

Durlach, N. (1963). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. 35, 1206–1218.

Elhilali, M., Chi, T., & Shamma, S. A. (2003). A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. *Speech Communication*, *41*(2–3), 331–348. https://doi.org/10.1016/S0167-6393(02)00134-6

Franken T. P., Bondy B. J., Haimes D. B., Goldwyn J. H., Golding N. L., Smith P. H., and Joris P. X. (2021). "Glycinergic axonal inhibition subserves acute spatial sensitivity to sudden increases in sound intensity," eLife 10, e62183. https://elifesciences.org/articles/62183 https://doi.org/10.7554/eLife.62183

Friston, K. (2002). Functional integration and inference in the brain. *Progress in Neurobiology*, *68*(2), 113–143. https://doi.org/10.1016/S0301-0082(02)00076-X

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement Learning or Active Inference? *PLoS ONE*, *4*(7), e6421. https://doi.org/10.1371/journal.pone.0006421

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K., & Buzsáki, G. (2016). The Functional Anatomy of Time: What and When in the Brain. *Trends in Cognitive Sciences*, *20*(7), 500–511. https://doi.org/10.1016/j.tics.2016.05.001

Friston, K. (2018). Does predictive coding have a future? *Nature Neuroscience*, *21*(8), 1019–1021. https://doi.org/10.1038/s41593-018-0200-7

Fountas, Z., Sylaidi, A., Nikiforou, K., Seth, A. K., Shanahan, M., & Roseboom, W. (2022). A Predictive Processing Model of Episodic Memory and Time Perception. *Neural Computation*, *34*(7), 1501–1544. https://doi.org/10.1162/neco_a_01514

Furutachi, S., Franklin, A. D., Aldea, A. M., Mrsic-Flogel, T. D., & Hofer, S. B. (2024). Cooperative thalamocortical circuit mechanism for sensory prediction errors. *Nature*, *633*(8029), 398–406. https://doi.org/10.1038/s41586-024-07851-w

Gardner-Medwin, A. R., & Barlow, H. B. (2001). The Limits of Counting Accuracy in Distributed Neural Representations. *Neural Computation*, *13*, 477–504.

Hartmann WM (1997) Signals, sound and sensation. Woodbury, N.Y.: AIP.

Hatfield G. (2002) Perception as unconscious inference. In Heyer D., & Mausfeld R. (Eds) Perception and the physical world: Psychological and philosophical issues in perception (pp. 113–143). John Wiley and Sons.

Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience*, *389*, 54–73. https://doi.org/10.1016/j.neuroscience.2017.07.061

Hellman, R.P. (1972). "Asymmetry of masking between noise and tone," Perception and Psychophysics 11, 241-246.

Helmholtz H. (1867). Handbuch der Physiologischen Optik (English tranl.: 1924 JPC Southall as Treatise on Physiological Optics) Voss.

Hemmatian, B., Varshney, L. R., Pi, F., & Barbey, A. K. (2023). The Utilitarian Brain: Moving Beyond the Free Energy Principle. *Cortex*, S0010945223003076. https://doi.org/10.1016/j.cortex.2023.11.013

Hermansky, H (2010) History of modulation spectrum in ASR, IEEE International Conference on Acoustics, Speech and Signal Processing. DOI:10.1109/ICASSP.2010.5494907

Hochstein, S., & Ahissar, M. (2002). View from the Top: Review Hierarchies and Reverse Hierarchies in the Visual System. *Neuron*, *36*, 791–804.

Hosoya, T., Baccus, S. A., & Meister, M. (2005). Dynamic predictive coding by the retina. *Nature*, *436*(7047), 71–77. https://doi.org/10.1038/nature03689

Hoyer, P. O., & Hyvärinen, A. (2002). Interpreting Neural Response Variability as Monte Carlo Sampling of the Posterior. *Proceedings NIPS 2002*.

Jaynes ET. Probability theory: The logic of science. Cambridge university press; 2003.

Jepsen, M. L., Ewert, S. D., & Dau, T. (2008). A computational model of human auditory signal processing and perception. *J. Acoust. Soc. Am.*, *124*(1), 422–438.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural Processing of Amplitude-Modulated Sounds. *Physiological Reviews*, *84*(2), 541–577. https://doi.org/10.1152/physrev.00029.2003

Kawahara, H., Masuda-Katsuse, I., de Cheveigné, A. (1999) Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication* 27:187-207.

Kawato, M., Hayakawa, H., & Inui, T. (1993). A forward-inverse optics model of reciprocal connections between visual cortical areas. *Network: Computation in Neural Systems*, *4*(4), 415–422. https://doi.org/10.1088/0954-898X_4_4_001

Koumura, T., Terashima, H., & Furukawa, S. (2019). Cascaded Tuning to Amplitude Modulation for Natural Sound Recognition. *The Journal of Neuroscience*, *39*(28), 5517–5533. https://doi.org/10.1523/JNEUROSCI.2914-18.2019

Litwin, P., & Miłkowski, M. (2020). Unification by Fiat: Arrested Development of Predictive Processing. *Cognitive Science*, *44*(7), e12867. https://doi.org/10.1111/cogs.12867

Lowet, A. S., & Uchida, N. (2024). Predictive coding: A distinction — without a difference. *Current Biology*, *34*(20), R926–R929. https://doi.org/10.1016/j.cub.2024.09.026

Mackay, D. M. (1956x). TOWARDS AN INFORMATION-FLOW MODEL OF HUMAN BEHAVIOUR*. *British Journal of Psychology*, *47*(1), 30–43. https://doi.org/10.1111/j.2044-8295.1956.tb00559.x

MacKay, D. M. (1956). The Epistemological Problem for Automata. In C. E. Shannon & J. McCarthy (Eds.), *Automata Studies. (AM-34)* (pp. 235–252). Princeton University Press. https://doi.org/10.1515/9781400882618-012

Markel J.D. and Gray, A.H. Linear Prediction of Speech, volume 12 of Communication and Cybernetics. Springer-Verlag, New York, 1976.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman and Co.

McDermott, J. H., & Simoncelli, E. P. (2011). Sound Texture Perception via Statistics of the Auditory Periphery: Evidence from Sound Synthesis. *Neuron*, *71*(5), 926–940. https://doi.org/10.1016/j.neuron.2011.06.032

McDermott JH, Schemitsch M, Simoncelli EP (2013) Summary statistics in auditory perception. Nature neuroscience 16:493–498.

McWalter, R., & McDermott, J. H. (2018). Adaptive and selective time averaging of auditory scenes. *Current Biology*, *28*(9), 1405-1418. e1410.

Millidge, B., Seth, A., & Buckley, C. L. (2021). *A Mathematical Walkthrough and Discussion of the Free Energy Principle* (arXiv:2108.13343). arXiv. http://arxiv.org/abs/2108.13343

Mumford, D. (1991). On the computational architecture of the neocortex. I. The role of the thalamo-cortical loop. *Biological Cybernetics*, *65*, 135–145.

Mumford, D. (1992). On the computational architecture of the neocortex. II The role of cortico-cortical loops. *Biological Cybernetics*, *66*, 241–251.

Nelson, P. C., & Carney, L. H. (2004). A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. *The Journal of the Acoustical Society of America*, *116*(4), 2173–2186. https://doi.org/10.1121/1.1784442

Palmer A. R, Rees A (2010). The Oxford Handbook of Auditory Science: The Auditory Brain. Oxford University Press.

Pinchetti, L. Tommaso Salvatori, Yordan Yordanov, Beren Millidge, Yuhang Song, & Thomas Lukasiewicz. (2022). Predictive Coding beyond Gaussian Distributions. *arXiv*, arXiv:2211.03481.

Poggio T (1990) A theory of how the brain might work. In: The Brain. Proc Cold Spring Harbor Symp 55

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Romero, G. E., & Trussell, L. O. (2022). Central circuitry and function of the cochlear efferent systems. *Hearing Research*, *425*, 108516. https://doi.org/10.1016/j.heares.2022.108516

Sanborn, A. N., & Chater, N. (2016). Bayesian Brains without Probabilities. *Trends in Cognitive Sciences*, *20*(12), 883–893. https://doi.org/10.1016/j.tics.2016.10.003

Sanborn, A. N., & Chater, N. (2017). The Sampling Brain. *Trends in Cognitive Sciences*, *21*(7), 492–493. https://doi.org/10.1016/j.tics.2017.04.009

Schmidhuber J (2009) Driven by Compression Progress: A Simple Principle Explains Essential Aspects of Subjective Beauty, Novelty, Surprise, Interestingness, Attention, Curiosity, Creativity, Art, Science, Music, Jokes. arXiv:0812.4360 [cs] arXiv: 0812.4360

Schofield, B.R. (2012). Structural organization of the descending auditory pathway. In *The Oxford handbook of auditory science—The auditory brain* (pp. 43–64). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199233281.013.0003

Shannon CE. A mathematical theory of communication. Bell Syst Tech J. 1948; 27: 379–423.

Sims, A. (2017). Problems with Prediction: The Dark Room Problem and the Scope Dispute. *Philosophy and Predictive Processing*. https://doi.org/10.15502/9783958573246

Sitko, A. A., & Goodrich, L. V. (2021). Making sense of neural development by comparing wiring strategies for seeing and hearing. *Science*, *371*(6525), eaaz6317. https://doi.org/10.1126/science.aaz6317

Slaney M. (1993). "An efficient implementation of the Patterson-Holdsworth auditory filter bank," technical report No. 35. Apple Computer.

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, *112*, 92–97. https://doi.org/10.1016/j.bandc.2015.11.003

Srinivasan, M. V., Laughlin, S B, & Dubs, A. (1982). Predictive Coding: A Fresh View of Inhibition in the Retina. *Proceedings of the Royal Society B: Biological Sciences*, *216*, 427–459.

Stein, A., Ewert, S. D., & Wiegrebe, L. (2005). Perceptual interaction between carrier periodicity and amplitude modulation in broadband stimuli: A comparison of the autocorrelation and modulation-filterbank model. *The Journal of the Acoustical Society of America*, *118*(4), 2470–2481. https://doi.org/10.1121/1.2011427

Viemeister, N. F. (1979). “Temporal modulation transfer functions based upon modulation thresholds,” J. Acoust. Soc. Am. 66, 1364–1380.

Williams, D. (2022). Is the brain an organ for free energy minimisation? *Philosophical Studies*, *179*(5), 1693–1714. https://doi.org/10.1007/s11098-021-01722-0

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An Internal Model for Sensorimotor Integration. *Science*, *269*(5232), 1880–1882. https://doi.org/10.1126/science.7569931

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. Neural Networks, 11, 1317–1329.