version 31 Dec 2023

Predictive coding in the auditory brainstem

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

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, via a predictive model that is tuned with the objective of minimizing prediction error. This operation is performed independently within each peripheral channel, with parameters determined from information within that channel. The result is a sensory representation that is invariant to a certain class of interfering 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.

Introduction

Predictive coding is an influential idea in sensory and cognitive neuroscience (Attneave 1954, Barlow 1961, Srinivasan et al 1982, Rao and Ballard 1991, Clark 2013; Friston 2018). Incoming sensory patterns are compared with predicted patterns, and the outcome of this comparison – the error – serves to validate the predictive model, and as an input for the next stage of processing. Predictable patterns are redundant, the error may be less costly to represent and easier to process, and the parameters of the model, fit to the sensory input, may capture useful information about the structure of the stimulus from which to infer the state of the world. Predictive coding is one of a constellation of related ideas, including unconscious inference (Alhacen 1030; Helmholtz 1867), Bayesian inference (Aitchison and Lengyel 2017), redundancy reduction (Barlow 1961), compression (Schmidhuber, 2009), efficient coding (Shannon 1948), forward models in motor control (Wolpert et al 1995), efferent copy (Crapse 2008), homeostasis (Friston 2010), reverse hierarchy (Hochstein and Ahissar), and so on. In each case, an observation is compared against a target, predicted or desired, and the mismatch used to adjust the model to better attain the target.

The term "predictive coding" covers several quite different concepts (Spratling 2015). Given a hierarchy of processing levels leading from sensory patterns to higher-level abstractions, prediction can occur in the *ascending* (bottom-up Fig. 1 left) or *descending* direction (Fig. 1 right). Examples of the former are center-surround receptive fields in the retina (Srinivasan et al 1982), examples of the latter are top-down processing in the visual cortex (Rao and Ballard, 1990) and the general predictive coding framework discussed by Friston (2018).

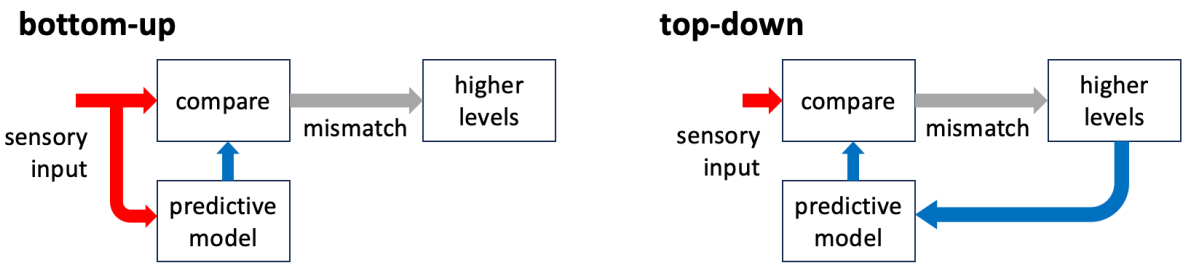

Fig. 1. Two forms of predictive coding. On the left, the prediction is derived from the same level or lower. On the right, it is derived from higher (more abstract) levels.

Bottom-up prediction is also embodied in the Linear Predictive Coding (LPC) method popular for spectral estimation in speech processing (Markel and Gray 1976). Top-down prediction is related to Helmholtz's notion of unconscious inference, and more generally 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 can take place within a hierarchy of processing stages, as are thought to exist along the sensory pathways of the brain.

Here, I discuss a recent model of low-level auditory processing that offers a textbook example of predictive coding of the bottom-up flavor. Low-level sensory input is compared, at each instant, with a predictor derived from its previous values, for the purpose of Auditory Scene Analysis. I take advantage of this concrete, bottom-up model to discuss some issues that arise when considering the alternative class of top-down predictive models, and discuss how the two approaches might coexist.

The in-channel cancellation model

This model addresses "cocktail party" situations involving complex acoustic scenes with multiple concurrent speakers, or multiple musical instrument lines. Specifically, it accounts for our ability to ignore a particular class of masking sounds that are *harmonic*, *quasi-harmonic*, or *spectrally sparse*. The structure and principle of the model are sketched here, more details and references are provided elsewhere (de Cheveigné 2021, 2023a).

The model is built around a "neural cancellation filter" hypothesized to be implemented in the auditory brainstem within temporally accurate neural circuits. This filter, which is fed by spike patterns carried from the cochlea by the auditory nerve, consists of a gating neuron with one excitatory synapse fed directly, and one inhibitory synapse fed via an adjustable delay of duration τ (Fig. 2 left, red). The properties of this neuron are assumed to be such that every spike on the direct pathway is transmitted *unless* a spike arrives simultaneously on the delayed pathway, allowing for some short time window of coincidence. This modifies the statistics of the spike train in such a way that it is depleted of inter-spike intervals of duration close to τ. The effects of this neural filter on instantaneous spike probability can be approximated by applying to that probability a delay-and-subtract filter of impulse response $h(t) = \delta(t) - \delta(0)$, where

$\delta(t)$ is the Dirac function translated to t (de Cheveigné 2023a). Based on this linear approximation, the effect of this filter is shown in Fig. 2 (right, red line)

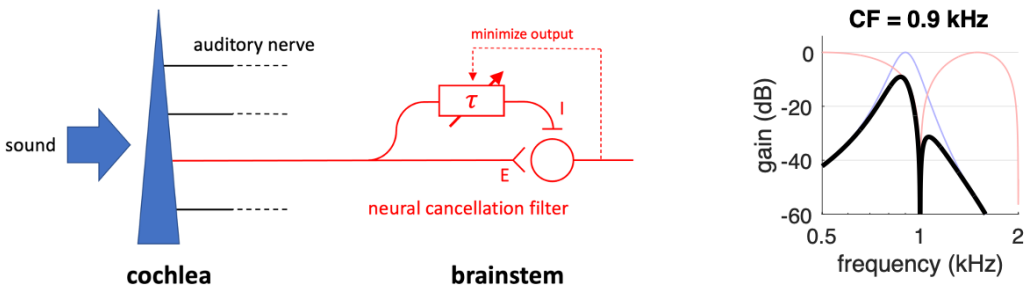

Fig. 2. Left: schematic representation of the cochlea as a bank of bandpass filters followed by a neural cancellation filter (one channel shown). Right: transfer functions of the cochlear filter (blue), a linear approximation of the cancellation filter (red) and the cascade of both filters (black).

The combination of cochlear and neural cancellation filter can thus be approximated as the cascade of two linear filters, as illustrated in Fig. 2 (right) for a cochlear filter of characteristic frequency CF = 0.9 kHz and a delay τ = 1 ms (inverse of 1 kHz). The cochlear filter has a bandpass characteristic with strong attenuation far from its characteristic frequency (blue line), whereas the delay-and-subtract filter has a deep zero at a frequency inverse of the delay $1/\tau$ and all of its multiples (red line). The cascade inherits both of these characteristics (black).

One such filter is applied to each peripheral channel (represented by the ~10 fibers of the auditory nerve issued from one of the ~3000 inner hair cells of the organ of Corti). The delay τ_k applied within channel k is determined independently for each of the K channels, with the objective to minimize the output of the compound filter (variance in its linear approximation). There are thus K parameters, all determined automatically based on a criterion of minimum output. Simulations show that the delay τ_k tends to lock on to the period of whichever signal component dominates the channel. In the presence of a strong interfering sound that is either *harmonic*, *quasi-harmonic*, or simply *spectrally sparse*, the filter tends to cancel the correlates of that sound, allowing components of a weaker concurrent sound to emerge. The model can account for a range of psychophysical results that involve masking with such sounds (de Cheveigné 2021, 2023a).

Figure 3 illustrates two examples. Panel A shows a cochleagram (time-frequency representation based on a modeled cochlear filterbank, Slaney 1983) in response to a pure tone masker of duration 0.5 s, upon which is superimposed a weak narrowband noise probe, at a relative amplitude (-10 dB) for which it is detectable by listeners (Hellman 1972) albeit not salient in this representation. Panel B shows the same at the output of an array of in-channel cancellation filters. A glitch is observed at the onset and offset of the masker, but between these the response is near zero in all channels, except at the position of the probe. The presence of the probe is now salient (visually) in B, consistent with its detectability in this condition. Accurate cancellation depends on the periodicity of the masker: if it were narrowband noise rather than a sinusoid, cancellation would be less effective and the probe less salient. This explains, according to the model, Hellman's (1972) observation that a pure tone masker is less effective (by up to 20 dB) than a narrowband noise masker (see de Cheveigné 2023a for a more complete discussion).

As another example, Fig. 3 C shows a cochleagram in response to a short snippet of speech ("wow cool", male voice), and Fig. 3D shows the degree of attenuation (in dB) of an array of cancellation filters at each time (resolution 30 ms) within each cochlear channel. Here, the cancellation filters adapt rapidly, independently in each channel, to track the local periodicities in this highly non-stationary stimulus. At most times, and in most channels, the attenuation is small (yellow color, close to 0 dB), but in certain spectro-temporal “pixels” the attenuation is greater, up to -20 dB (blue color). High attenuation pixels line up with high amplitude pixels (compare panels C and D), implying that attenuation may be significant overall, despite the sparsity. Thus, if this speech were to play the role of masker, cancellation-filtering might allow a weaker target (for example voice) to be glimpsed. The fact that the model can handle non-stationary sounds such as speech suggests that it might contribute to account for the well-known “cocktail party effect”.

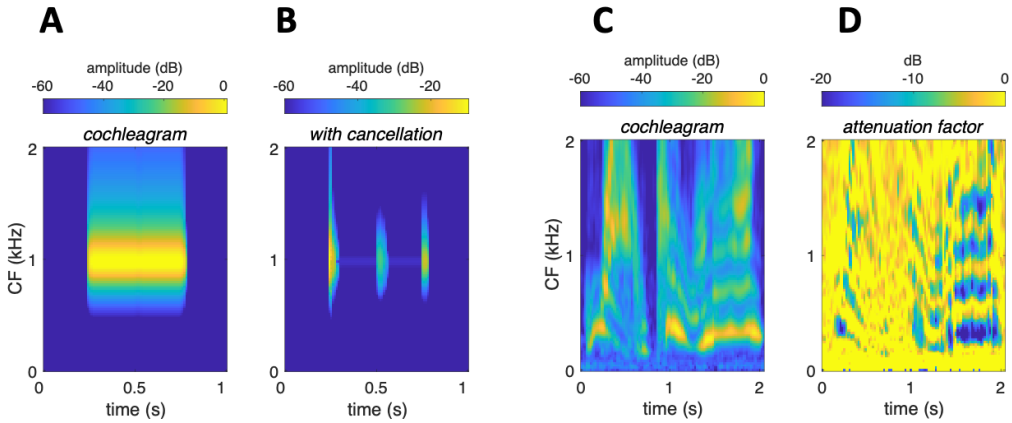

Fig. 3. A: Cochlear excitation pattern in response to a pure tone masker of frequency 1 kHz on which is superimposed a weak narrowband noise probe (duration 30 ms, onset circa 0.5 s, amplitude −10 dB relative to the masker). B: Same, at the output of the cancellation filter array. C: Cochlear excitation pattern in response to a short snippet of speech ("wow, cool!"). D: Attenuation factor of the cancellation filter array at each time and CF for the stimulus in C.

Downstream processing stages are assumed to have access to both the output of the cancellation filters and their input (i.e. output of the cochlea). Each peripheral channel is thus split into two subchannels, one filtered and the other not, effectively doubling the number of channels (possibly contributing to the “neural fanout” observed in the brainstem). Downstream processing can also benefit from the automatically-determined delay parameters within each channel, τ_k, for example as cues to musical pitch (tone height) (de Cheveigné 1998, 2023a). Alternatively, delay parameters might be set in a top-down fashion, overriding their automatic estimation, for example to hear each note of a melody in the context of previous notes, or to tune a musical instrument relative to a reference. The model may also be extended to provide a measure of *period change* (rather than period) to directly support the sensation of pitch change that is sometimes felt in the absence of a clear pitch (Demany and Semal 2018, de Cheveigné 2023a)

The neural circuitry required by the model (fast excitatory-inhibitory interaction, Fig. 2 left, circuit in red) is similar to that hypothesized by the classic equalization-cancellation model of binaural unmasking (Durlach 1963; Culling and Summerfield 1995) for which support has been found within the lateral superior olive (LSO) in the brainstem (Biederbeck 2018; Franken et al 2021). Similar fast excitatory-inhibitory interaction is also found at other levels within the auditory brainstem, as reviewed by de Cheveigné (2022, 2023a,b). However, positive evidence

for an in-channel cancellation circuit is still lacking, so this neural processing mechanism remains hypothetical.

In-channel cancellation as bottom-up predictive coding

The structure of the in-channel cancellation model (Fig. 2) clearly evokes predictive coding. The incoming sensory pattern (direct pathway) is predicted by an internal predictive model (delay). This is similar to the predictive coding model of Srinivasan et al (1982) in which the center of a retinal patch is predicted from its neighbors and/or previous values. The result of subtraction is zero for the class of stimuli that fit that model, whereas non-zero output implies a mismatch with the model, for example indicating the presence of an additional, weaker sound source. The cancellation filter achieves both *redundancy reduction*, using stimulus regularity (periodicity and/or spectral sparsity) to simplify the sensory code and remove structured interference, and *redundancy exploitation* (Attneave 1954; Barlow 1961), in that delay estimates τ_k characterize the stimulus structure.

The model is related to linear predictive coding (LPC) (Gray and Markel 1996). Whereas for LPC each sample is predicted as a linear combination of N successive samples with coefficients adjusted for minimum error variance (where N is the order of the LPC), for in-channel cancellation it is predicted by a single sample with a delay also adjusted for minimum error variance. In the case of LPC one predictive model is fit to the wideband audio signal, whereas the in-channel cancellation model posits multiple predictive circuits, one for each channel of the peripheral filter bank.

The output of this predictive coding stage (cancellation filter) is a fast, multi-channel time-domain pattern available for further low-level processing (Fig. 4) that eventually results in a more slowly-varying abstract representation handed over to the next-higher processing level. Higher levels might also benefit from the time-varying vector of local period estimates and/or period-change estimates (Fig. 4). Within the auditory nervous system, fast processing might occur in brainstem, and slower processing within inferior colliculus and beyond.

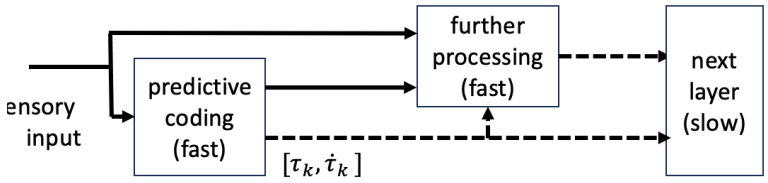

Fig. 4. The output of the cancellation filter (predictive coding) consists of a fast-varying pattern of spikes within each channel, that presumably undergoes further processing before being handed over as a more slowly-varying signal to the next layer in the sensory processing hierarchy. The dashed pathways carry slowly-varying messages, abstracted from the sensory patterns. Values τ_k and $\dot{\tau}_k$ represent the period and period change, respectively, estimated within each channel k.

In-channel cancellation as top-down predictive coding

Recent accounts of predictive coding invoke a top-down structure in which predicted patterns are inferred from more abstract internal models, according to a principle akin to "analysis by synthesis" that allows inference of the cause that produced the stimulus (Braitenberg 1984; Kawato et al 1993; Friston 2002). It is worth considering whether the present model could be re-cast as a top-down predictive model, so as to "hitch onto that wagon" and reap its theoretical benefits.

I'll first consider two naive ideas, seemingly reasonable but that turn out to be unworkable. The first idea is to feed the predictive model from the error (output of the filter, Fig. 5B) rather than the sensory input (Fig. 5A). A moment's thought tells us this might be a bad idea: if prediction were effective, i.e. the error were zero, there would be nothing to feed the predictive model. If we persisted, we would discover that such a filter is unstable, with an infinite, non-decreasing impulse response (Fig. 5B). This instability 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. A more sophisticated idea, inspired from the concept of "analysis by synthesis" is to derive the prediction from a generative model fed with abstract patterns from a higher stage (Fig. 4D). A moment's thought, again, suggests that this won't work, because that model has no access to the temporal (phase) information needed to correctly align the prediction with the incoming pattern. Both of these straightforward-seeming ideas are unworkable.

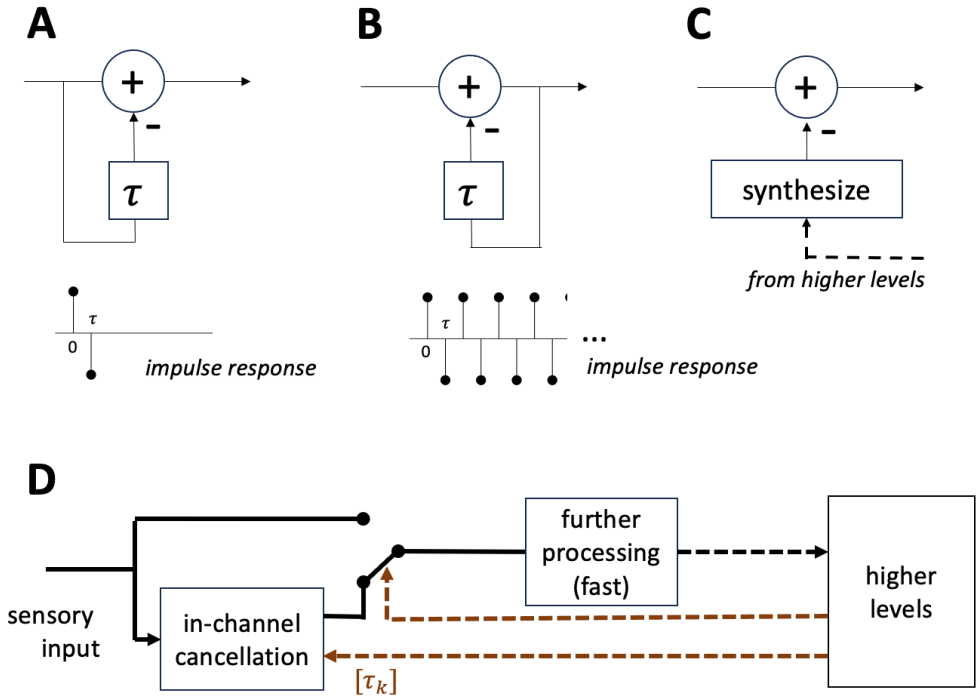

Fig. 5. A: Feed-forward predictive coding suppresses periodic input. It implements a filter with a very compact impulse response. B: A naive feed-back predictive coding model. It implements a filter with an infinite impulse response, non-decreasing. C: Another naive feed-back predictive model based on "analysis-by-synthesis". D: A more abstract top-down predictive coding scheme.

A third, less naive approach is to insert in-channel cancellation within a looser, top-down predictive coding loop (Fig. 5D). A selling point of in-channel cancellation is that it allows certain masking sounds (harmonic or spectrally sparse) to be suppressed, allowing weaker sounds to be perceived. As such, it might have its place as part of a bottom-up "reverse acoustics model", analogous to the "reverse optics" model of Kawato et al (1993) that approximately reversed the scene-to-retina transform. Here, the purpose would be to approximately reverse the physical mixing of sounds. That bottom-up model would work hand-in hand with a top-down "forward acoustics model", analogous to the "forward optics" model of Kawato et al (1993), embodied by descending pathways in the auditory nervous system. Forward and reverse models are required parts of predictive coding (Friston 2002).

This combination of an ascending "reverse model" and a descending "forward model" seems counterintuitive and might need more explaining. The *ascending* part makes sense as an attempt to unmix the sensory evidence to better observe its causes, but it cannot succeed perfectly

due to the non-reversible nature of acoustic mixing. The *descending* part can be understood as a means to fill the inevitable gaps left by that unmixing process. Via this descending pathway, both the tentative representation being elaborated, and the forward model, are gradually refined under the guidance of the mismatch signal. This combination of analysis (through ascending pathways) and synthesis (through descending pathways) learns an internal representation of the external causes (Friston 2002). Synthesis is easier than analysis (Braitenburg 1984).

As Kawato et al (1993) pointed out, the reverse (unmixing) model does not have to be perfect, and indeed it certainly is not perfect here. For one thing, in-channel cancellation is effective only for harmonic/spectrally sparse maskers, so it would need to be complemented by other mechanisms to handle a wider range of sounds. For another, it impoverishes the representation of the weaker, non-cancelled sounds, that then need to be inferred according to a Helmholtzian inference process (de Cheveigné 2021), hopefully subsumed by the descending part of the loop.

What are the descending pathways in our case? One path might control a hypothetical switch that determines whether in-channel cancellation will be applied or not (Fig. 5D). Another might impose the delays τ_k to override the automatic in-channel estimates. It requires some mental flexibility to see all this as a "forward acoustics model", but roughly speaking: flipping the switch to turn on the filtering corresponds to the assertion "a harmonic masker was added to the mix" and setting the delays all to T corresponds to the assertion "its period is T". Imagining this bidirectional process in action, the onset of a musical note might trigger the automatic estimation of its period, which is then sent to a higher level to derive a musical pitch sensation characteristic of the note. That level might then send the period T back down to the filters for confirmation. A mismatch might then signal a pitch change (Demany and Semal 2018, de Cheveigné 2023a) that contributes to the percept of a melody. Alternatively, a wider loop might include a hand turning a key to tune a guitar, in the spirit of sensorimotor interpretations of predictive coding (Friston 2018).

The reader may feel dissatisfied at this point. Two seemingly textbook renderings of "analysis-by-synthesis" were unworkable, whereas a third accomplishes "synthesis" but in an abstract and hand-waving sense. Is it that top-down predictive coding doesn't work for a task that seems ideally cut for it? In the rest of the paper, I will attempt to flesh out the notion of sensory

predictive coding, and identify its place within prediction-based mechanisms of perception, decision, and action. But first, let us step back for a wider view.

Predictive coding, revisited

The concept of predictive coding has evolved over the years, not always in the direction of clarity. Writings of early authors (e.g. Craik 1943, Attneave 1954, Barlow 1961, Mumford 1991, 1992) are limpid, but more recent discussions are more technical and challenging, possibly as a consequence of their greater ambitions (Friston 2009, 2010; Buckley et al 2017).

What to predict, and why? We might predict (1) optimal *future actions* so as to survive within the world, or (2) the *future state of the world*, or (3) the *future state of sensory stimulation* (i.e. predictive coding). Arguably, we really only need (1) to survive and multiply. If we can plan all future actions to satisfaction, nothing more is required. However, that goal may be assisted by a maintaining a model of the world that includes the ability to predict what will happen next within that world, in the same way as an internal model of a limb assists motor control of that limb (Wolpert et al 1995). Predicting the world (2) is thus a useful subgoal of action planning (1). Our knowledge of the world in turn is based on our senses, and these may plausibly benefit from the ability to predict at each instant the future of the stimulus itself, so as to improve its processing (Attneave 1948; Barlow 1961). Thus, predicting the future of the stimulus (3) may help predict the future state of the world (2), which helps plan good actions (1).

These goals are nonetheless distinct. I might form a model of what I hear (e.g. music) with little knowledge or care about what produced it, or I might maintain a model of the world but not know what to do next. Conversely, I might know what to do next (e.g. flee) despite an incomplete world model.

Objectives at a lower level (e.g. sensory processing) must fit longer-term requirements of higher levels (e.g. inference of a representation, or stability over the lifespan of the individual or the evolutionary span of a species). However, factoring this into the analysis of lower-level processing introduces a teleological component that may interfere with a mechanistic understanding. Here, I will set aside higher-level goals (e.g. representation, decision, action, homeostasis, etc.) and focus on the lowest level of a predictive processing hierarchy. I will argue two main points: (1) the need for a just-in-time statistical representation of sensory input, and (2) the potential role of predictive coding for the chunking of this representation.

Predictive coding of the stimulus offers benefits of four sorts. First, it offers a road to *redundancy reduction* (Attneave 1954, Barlow 1961), among other roads, since predictable bits don't need representing. Redundancy reduction offers potential advantages for computation ("economy of thought") and metabolism ("economy of impulses") (Barlow 1961). Second, irrelevant offsets, trends, or patterns may be predicted and factored out, as in the present in-channel cancellation model. Third, the parameters of the predictive model, obtained by fitting it to sensory input, reflect regularities within the world (Attneave 1954). Regularities are distributed unevenly in space and time, and their extent and boundaries may provide useful information, for example to delimit an object and segregate it within a scene. For Barlow (2001), redundancies are to be *exploited*, not just reduced. The in-channel cancellation model operates in this spirit by providing, within each channel, an estimate of the period of a periodic sound. Fourth, predictive coding supports inference and representational learning, as mentioned earlier (Friston 2002).

In the literature, the term "predictive coding" carries several meanings (Spratling 2015). Among them, we can distinguish *bottom-up* prediction from *top-down* prediction, with assumptions, constraints, and ambitions that are very different. Of the two, top-down prediction is arguably the more "interesting" because it relates to many important concepts such as Helmoltzian or Bayesian inference, representational learning, the largely top-down circuitry in the cortex, and the concept of Free Energy (Friston 2002, Clark 2013). Much recent literature focuses on this aspect of predictive coding (also referred to as "predictive processing"), but the concepts are sometimes obscure and there is some controversy (Litwin and Miłkowski 2020; Andrews 2021; Hemmatian 2023).

Carving sensation at its joints

In discussions of predictive coding, sensory input is treated as if it were a well-defined quantity, but in reality, its nature is blurred by *time*. If we define "sensory input" as the pattern on the sensory epithelium, input at time t no longer qualifies a "input" by time $t + 1$: it has been replaced by something new. If we agree that that information nonetheless retains some value past the time it was gathered, we must acknowledge that sensory input has a temporal dimension, in other words it consists of *memory*. Furthermore, sensory patterns may be characterized by how they unfold over time (as in the case of motion in vision) and in audition essentially every

acoustic pattern of interest has a temporal dimension. Thus, when we speak of a sensory input "layer", we must keep in mind that the representation it carries (a) extends over time, and (b) is constantly renewed.

This must be taken into account when considering the idea of analysis-by-synthesis, by which an input pattern is resolved within a reverberating predictive loop. At the same time as this occurs, new input is at hand and competes for processing. Even if the task requires waiting for "prediction" to complete, a channel must be kept open for ongoing, newer sensory input. It is this "task-independent" channel that I wish to focus on here. It operates under several constraints: any processing must be carried out in time to be ready for new input, the representation must include a temporal axis (memory), and the information must cater for a range of *future* needs. When (and if) a need arises for that information, the quality of the organism's response depends on how it was gathered and processed, before that need was fully known. The process can only be tuned by extrapolating from previously-arisen needs.

Taking a different perspective, sensory processing is sometimes seen as a pipeline that feeds a representation of *causes*, possibly leading to *action*. Instead, one may argue, its foremost role is to learn and recognize the vast class of stimuli that are *not* worthy of representation or action. We refrain from acting more often than we act, and ignore much more than we attend. This is not for lack of stimulation: the ongoing sensory input may salient and variegated, and yet discounted. We must learn to recognize when this should be the case. Gardner-Medwin and Barlow (2001) point out that statistics of *all* sensory events must be tallied, particularly those that are common and unimportant, so as to allow significant conjunctions ("suspicious coincidences") to be spotted. A conjunction AB is "suspicious" if $p(AB) \gg p(A)p(B)$, which can only be spotted if A and B have been counted. This counting occurs outside the context of any particular action, and one can argue that learning these statistics, and learning to ignore them, is the bulk activity of the sensory brain.

In the last section of this paper, I will attempt to cast the in-channel cancellation model as contributing a temporally-extended, time-varying sensory representation that, to paraphrase Plato, "carves sensation at its joints" (Friston and Buzsaki 2016).

Prediction and error

The words "prediction" and "error" are used inconsistently in the literature. A simple interpretation of "prediction" is that it is a *value* P within the same space as the sensory observation O (Fig. 6A). The "error" then is the vector difference $O - P$ between observation and prediction, or possibly its norm $\|O - P\|$. Predictive coding then involves the minimization of the error (norm), possibly using the error (vector) to nudge the predictive model in the right direction.

This interpretation cannot be applied to prediction from a higher, abstract layer. As pointed out by Mumford (1992), abstraction is a many-to-one transform, not invertible, and thus a top-down prediction translates as a *class* of values, rather than a single value (Fig. 6B). A generative model might produce an exemplar (for example a retinotopic pattern from a texture or face), but it is one of many and thus unlikely to fit exactly the sensory input. Closer to the focus of this paper, a slow auditory representation from the inferior colliculus or cortex cannot translate unambiguously to a fast representation in the brainstem.

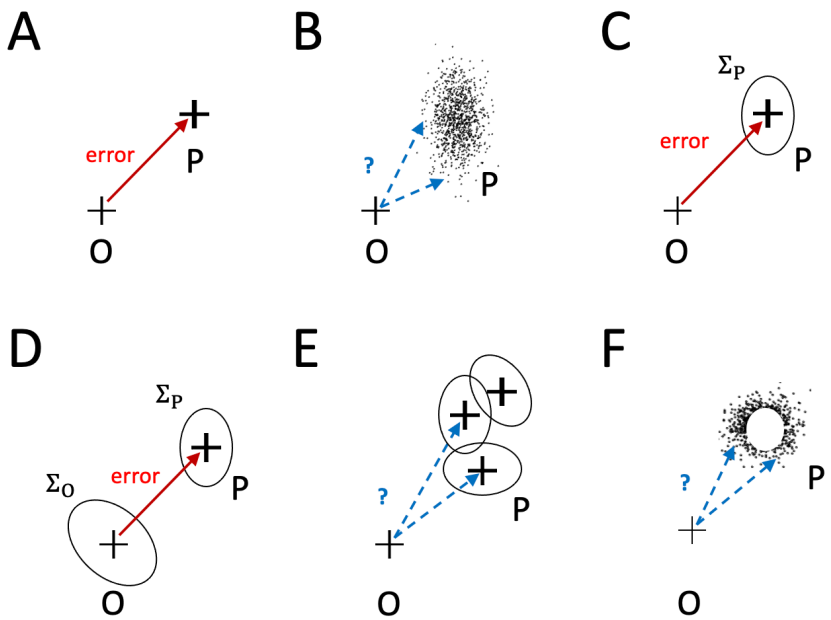

Fig. 6. A: If the prediction is a *value* (point in the space of observations), the error can be defined as the difference between observation O and prediction P, or its norm $\|O - P\|$. B: If the prediction is a *class*, the error is not well defined. C: If the prediction can be approximated by a Gaussian distribution, the error may be defined as the difference between the observation and the mean of the distribution (or its norm), possibly weighted by the variance or covariance. D: Multiple or uncertain

observations might also be modeled as a Gaussian, in which case there are two covariances to consider. E and F: If the prediction is a mixture-of-Gaussian or arbitrary distribution, the error is not well defined. It is also not defined for a non-statistical predicted class, for example "waveforms of period T" as exploited by the in-channel cancellation model.

However, if the class can be approximated as a Gaussian distribution parametrized by its *mean* ("template") P and *variance* or *covariance* Σ_P, the error again can be defined as the difference between observation and mean, or the norm of that difference, possibly weighted by its covariance (Fig. 6C), similar to the single-value case. The Gaussian assumption is widely made in the literature, and underlies a construct such as "precision-weighted error". An observation that is uncertain, or a group of observations can also be represented as a Gaussian distribution characterized by its covariance Σ_O. It is sometimes unclear whether "precision" refers to that of the prediction (e.g. Friston 2002), or that of the observation (e.g. Heldman an Friston 2010), or both. The precision Π_P (inverse of Σ_P) is presumably part of that prediction, and the precision Π_O(inverse of Σ_O) part of the observation. However, Π_O might also be inferred from prior knowledge about observational reliability, or attentional state (e.g. Heldman and Friston 2010), and in that case part of the prediction. Whether or not the predictive model parameters are updated based on the observation, and by how much, depends on these precisions, but it also depends on priors on those parameters (not captured by these precisions). In terms of message-passing, the content of the "error" depends on whether the update occurs at a lower level (it consists in new parameters P and Σ_P) or at the upper level (it consists of $O - P$ and Σ_O).

The Gaussian assumption is restrictive, as it excludes for example multimodal distributions (Sanborn and Chater 2017), and the Laplace approximation imposes a further constraint that ties the precision to the mean (Friston 2009). Arbitrary distributions might be catered for by applying transform into a space within which they is Gaussian (Friston 2009, p298), however this begs the question as to how the transform is learned or evolved. The in-channel cancellation model might be construed as such a transform (see Discussion).

If the predictive model is not Gaussian, the concept of "error" is not straightforward (Fig. 6E,F), but fit can still be evaluated on the basis of the predicted probability density at the observation, or the Kullback-Liebler (KL) divergence between the distribution $\mathcal{O}$ of the observation and the distribution $\mathcal{P}$ of the prediction. KL divergence is not symmetrical, and thus we have a choice between the divergence of *the observation relative to the prediction,* $D_{\mathrm{KL}}(\mathcal{O} \parallel \mathcal{P})$, or of *the*

prediction relative to the observation, $D_{\mathrm{KL}}(\mathcal{P} \parallel \mathcal{O})$ (analogous to Π_{P} and Π_{O} respectively in the Gaussian case). A small value of $D_{\mathrm{KL}}(\mathcal{O} \parallel \mathcal{P})$ suggests that the observation is unsurprising, and a small value of $D_{\mathrm{KL}}(\mathcal{P} \parallel \mathcal{O})$ that the observation is imprecise. KL divergence is scalar, and thus does not indicate the *direction* of the mismatch between prediction and observation. An option is to use it to control a "gate" that only transmits surprising input. Those signals then form an "error" signal, possibly recoded in a form that allows updating the predictive model (e.g. a new mode in a Gaussian mixture model).

A further issue is how to code a non-Gaussian prediction for transmission from a higher to a lower level. It has been suggested that complex distributions may be represented in the brain by sampling (e.g. Sanborn and Chater 2016), in which case the top-down prediction could consist of an ensemble of such samples, possibly consistent with the large number of top-down connections observed between levels in the brain. It may also happen that the "prediction" is not a statistical distribution, but rather as an algorithm or filter applied to incoming data (as the in-channel cancellation model), in which case its output can be treated similarly to KL-divergence as it is small if the data fit the prediction. A non-Gaussian, KL-divergence-based version of predictive coding is described by Pinchetti et al (2022).

To summarize, predictive processing addresses a continuum of tasks from sensory coding to perceptual representation, decision and action, and beyond. The first stage, sensory coding, operates under stringent time constraints, and must cater to needs that are not fully defined at the time of coding. Sensory representations may involve complex regularities or distributions that are not Gaussian, requiring tools such as the Kullback-Liebler divergence to measure the match between observation and prediction. The next section attempts to bring these ideas together, moving from the in-channel cancellation model to a wider perspective of predictive sensory processing.

Predictive chunking in sensory memory

A sensory representation is the first layer in the predictive processing hierarchy from which a perceptual representation is derived (Friston 2002). As argued earlier, it must include time as a dimension to allow the perceptual representation to be calculated over a temporally extended pattern of features. It thus must take the form of a stream of feature samples (spectrum, etc.) indexed by time relative to the present, going backwards. This temporally-extended

representation itself evolves over time as new sensory data come in, and old features are (presumably) forgotten.

The physical stimulus is variable on scales long and short, so the raw features fluctuate over time, i.e. are *noisy*. However, not all fluctuations are worth representing, in particular there may be intervals of relative stability bounded by transitions from one state to the next. Conserving a time series of states, rather than instantaneous features, allows the sensory representation to cover a longer time span for the same storage cost. However, the mapping from instantaneous features to states is imperfect, and thus each state sample reflects a set of values that it summarizes by a set of *statistics* that parametrize their empirical distribution. The mapping is thus an abstraction, that might be seen as an early layer within a hierarchical predictive coding model. It involves *chunking* the raw feature stream in such a way that the distribution within each chunk is tight and distinct from neighbours. The assignment of boundaries between chunks is crucial to the quality of the representation. A plausible criterion is KL divergence: states S_t and S_{t+1} can be merged if $D_{\mathrm{KL}}(S_t \parallel S_{t+1})$ and $D_{\mathrm{KL}}(S_{t+1} \parallel S_t)$ are both small (if either were large there would be grounds to keep the states distinct).

In the rest of this section, I sketch four ways in which in-channel cancellation might contribute to these goals. Each might benefit from further elaboration, which is outside the scope of this paper.

Proposition 1. In-channel cancellation contributes to chunking

A periodic signal is concisely specified by its period T and the shape of the waveform during the period. A natural representation of a piecewise periodic stimulus is as a time series of such statistics, possibly associated with an error term that summarizes the deviation from that specification over a chunk. Chunk boundaries may be set at instants when the error term exceeds a threshold. Figure 3B offers an illustration: cancellation filter output is small during the periodic masker, and large at its onset and offset (it is also momentarily increased at the position of the probe). A chunk may also be split if the mean error over the chunk exceeds a threshold, and an additional state, *aperiodic*, may be used to concisely represent intervals for which no periodic model applies. The model parameter T (if periodic) may be determined by a bottom-up fit to the data as described in section “The in-channel cancellation model”, or based on a top-down prediction from a higher level as described in section “In-channel cancellation as top-down

predictive coding". The "in-channel" property allows a different fit in different time-frequency regions, allowing for a richer sensory representation.

Proposition 2. In-channel cancellation contributes to smoothing

Within each state, sensory input is represented by the period T and the shape of the signal during the period, or a spectral representation thereof. The time series of spectral representations (for example defined as a tonotopic pattern of activity over peripheral channels, or a Fourier transform within channels) is less noisy, and thus more useful, if it is *smoothed* over a period, or period multiple. Period-tailored smoothing avoids the "moiré" effect that arises when a regularly-fluctuating quantity is sampled at an unrelated rate, that adds unnecessary noise to the spectro-temporal representation (Kawahara et al. 1999). Again, the "in-channel" property allows for different smoothing in different channels. Channels that carry a periodic signal are smoothed according to that period, channels that respond to noise are usually dominated by the period 1/CF, where CF is the center frequency of the channel. Period-specific smoothing has an effect similar to adding instantaneous power of signals in quadrature as in the Hilbert transform.

Proposition 3. In-channel cancellation contributes to perception of concurrent sounds

In the example illustrated in Fig. 3B, in-channel cancellation allowed a weak (-10dB) probe to emerge from a periodic (pure tone) masker. That plot showed power, but more generally the output of the cancellation filter is a rapidly-varying signal that could be further analyzed to characterized whatever sound source was present concurrently with the masker. Alternatively, the standard cancellation filter can be replaced by a more complex circuit that performs joint estimation of both sound sources. Examples of such schemes can be found in de Cheveigné (2006, 1999, 2023b). If multiple sources within a complex scene can be suppressed (for example if they all are periodic but with different periods), it may be necessary to switch between processing options, for example under the direction of higher levels in the processing hierarchy (as illustrated in Fig. 5D). This is another way in which in-channel cancellation might participate within a top-down predictive processing structure.

Proposition 4. Cancellation filtering may operate hierarchically

The in-channel cancellation model operates on a rapidly-varying sensory pattern such as carried by the auditory nerve, and thus is applicable to processing at the lowest level. However,

the same principle is applicable to slower modulations, for example those typically modeled by a "modulation filter bank" (Viemeister 1979; Dau et al 1997), or rhythmic patterns as in music, speech, or certain environmental sound textures (McDermott et al 2013; McWalter and McDermott 2018). The combination of feed-forward predictive coding and top-down predictive control might be deployed at multiple levels (Fig. 7).

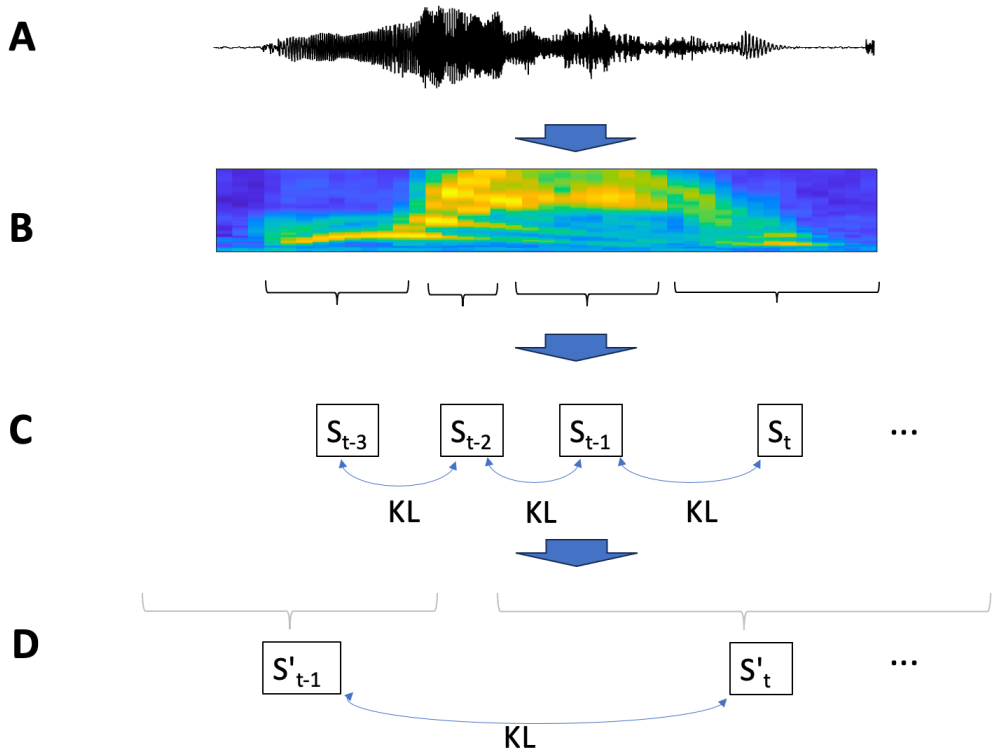

Fig. 7. The sound (A) is transduced and represented as features (B) that are chunked and summarized by statistics (C). The time series of first-order statistics may be further chunked and summarized by second-order statistics (D), and so on. Boundaries between chunks are positioned to maximize KL divergence between distributions of features in neighboring chunks.

To summarize, this section briefly examined four ways in which in-channel cancellation, a bottom-up form of predictive coding, might contribute to a more general top-down predictive coding framework.

Discussion

The in-channel cancellation model was developed to account for our ability to hear a faint target sound within a background comprising maskers that are harmonic, quasi-harmonic, or spectrally sparse (de Cheveigné 2023a). The processing principle (delay and subtract) is a prime example of predictive coding of the feed-forward type (Srinivasan et al 1982). More

recently there has been intense interest in predictive coding models of the feed-back type, with the wider perspective of representational learning, decision, action, and beyond (Attneave 1954; Mumford 1992; Kawato et al 1993, Rao and Ballard 1999; Friston 2002, 2009, 2018, Clark 2013). Thus, it seemed interesting to also consider if the in-channel cancellation model can be reformulated as a feed-back predictive model, or if not, if it could usefully contribute to one.

Feed-forward prediction cannot be replaced by feed-back prediction from a higher layer. As Mumford (1992) pointed out, top-down prediction requires message-passing between layers that do not "speak the same language". While a generative model might "translate" the abstract prediction to the concrete language of a lower layer, that translation is not unique. In our case, the specific temporal (or phase) pattern required for effective cancellation cannot be obtained from a more sluggish layer (e.g. inferior colliculus or cortex). Thus, the useful properties of the feed-forward cancellation model cannot all be attained by a feed-back model.

If a prediction from a more abstract layer maps to a *class* of patterns, there is no unique "error". This issue can be finessed by assuming that the abstract class follows a *Gaussian distribution*. The error signal then may be defined as the difference between the observation and the *mean*, possibly weighted according to the "precision" (inverse of the covariance). However, this is a strong assumption, reasonable in the absence of better knowledge (Jaynes 2003), and given the far-reaching theoretical benefits it procures, but that might not capture every interesting aspect of the world. Going back to the example in this paper, the periodic structure of a sound is not easily cast as a Gaussian distribution. Friston (2002) suggested that any distribution may be made Gaussian by a *coordinate transform*, however, it is not clear how an appropriate transform can be found or learned in every situation.

The in-channel cancellation model might be viewed as supporting one such transform. The reasoning is as follows: consider a bank of J cancellation filters indexed by delay τ_j and denote the power across the output as e_j. The minimum of e_j is distributed close to zero for the class of periodic sounds, while the position at which it is attained (see de Cheveigné and Kawahara 2002 for details) is distributed close to T for the class of sounds of period T. Periodic properties might thus be cast in statistical terms.

An alternative is to give up the Gaussian assumption and compare prediction and observation using KL divergence. This yields a scalar rather than a difference vector as in the Gaussian case, a less useful guide for the update of a predictive model. However, a plausible strategy might be to gate observations according to KL divergence, and forward those observations for which KL divergence is large (in lieu of "error signal"), possibly after recoding. For example, if the predictive distribution were a mixture of Gaussians, the error might signal a new component of that mixture. If a complex distribution were represented within the brain by *sampling* rather than parametrically (e.g. Hoyer and Hyvarinen 2002; Buesing 2011; Sanborn and Chater 2016; Aitchison and Lengyel 2016), a top-down prediction message might consist of multiple samples, possibly justifying a large number of top-down neural connections. In-channel cancellation suggests yet another, non-statistical alternative.

The sensory representation must extend over time to capture context effects, or temporal regularities, etc., and its content is incessantly renewed as time passes. In-channel cancellation might participate in a chunking mechanism by which sensory input is grouped over intervals for which its statistics are (approximately) stationary. Chunking, which might itself be hierarchical, would improve representational efficiency and increase the time span that can be represented for a given cost.

In-channel cancellation contributes to the analysis of an acoustic scene by allowing it to be stripped of sources that are harmonic or spectrally sparse. It thus can be viewed as forming part of a bottom-up analysis hierarchy, analogous to the "reverse optics" model of Kawato et al (1993), that together with a top-down synthesis hierarchy elaborates a perceptual representation of the sources (Friston 2002). Cancellation ensures invariance with respect to a masking source, however it comes with a side effect: a target source that accompanies the masker is distorted by the filter: any component of the target that coincides with a component of the masker is suppressed. To perceive that source as if it were whole, perception must rely on a Helmholtzian process of unconscious inference, hopefully subsumed by the descending predictive hierarchy.

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 most models of predictive coding.

References:

Aitchison, L., & Lengyel, M. (2016). The Hamiltonian Brain: Efficient Probabilistic Inference with Excitatory-Inhibitory Neural Circuit Dynamics. *PLOS Computational Biology*, *12*(12), e1005186. https://doi.org/10.1371/journal.pcbi.1005186

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

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

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

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

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. (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 (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.

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.

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. (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

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

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

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.

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.

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

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.

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

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

Mackay, D. M. (1956). 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

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 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. *Biological Cybernetics*, *65*, 135–145.

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*, *66*, 241–251.

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

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

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

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

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

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.

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

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