Does Loudness Represent Sound Intensity? (Preprint)

Kim Soland

Oct. 2021 revision
Mar. 2022 revision forthcoming in *Synthese*

Abstract

In this paper I challenge the widely held assumption that loudness is the perceptual correlate of sound intensity. Drawing on psychological and neuroscientific evidence, I argue that loudness is best understood not as a representation of any feature of a sound wave, but rather as a reflection of the salience of a sound wave representation; loudness is determined by how much attention a sound receives. Loudness is what I call a *quantitative character*, a species of phenomenal character that is determined by the amount of attention that an underlying perceptual representation commands. I distinguish quantitative from qualitative character; even qualitative characters that represent degrees of sensible magnitudes are phenomenally and functionally distinct from quantitative characters. A bifurcated account of phenomenal character emerges; the phenomenal is not exhausted by the qualitative.

Perceptual experience tells us about the things in our environment. My experience of my red coffee mug, for example, says that a certain object, the coffee mug, bears a certain feature, redness. Intuitively, perception gets things right at least some of the time; though sensory malfunction or poor lighting conditions might cause my mug to look some other color, my visual system accurately represents the color of my mug when it delivers an experience in which the mug looks to be red. If that is right, then there must be a correspondence between the reddish quality of my experience and some feature of the mug; reddishness represents that the mug has *that feature*, and whether the mug in fact bears *that feature* determines whether the experience is veridical.

The project of psychophysics is to determine the features of distal stimuli to which the qualities apparent in perception correspond. Color experiences, we are told, correlate with reflectance properties of object surfaces; pitch reports on sound wave frequency; odors tell us about chemical compounds in the air around us. In this tradition, is widely held that loudness is the perceptual correlate of sound wave intensity; the loudness or softness of a sound experience represents the intensity of the sound wave that causes it.

The problem with this view about loudness is the presupposition that loudness is a perceptual dimension of auditory experience; I argue that loudness does not represent a distal feature at all. Though it is true that loudness more-or-less covaries with sound intensity, the claim that loudness *represents* sound intensity is not supported by empirical evidence. I argue that the best way to make sense of the data about the relationship between sound intensity and loudness is to take loudness as a reflection of the *salience* of a sound – how much attention the sound commands – and sound wave intensity as a dominating but non-exclusive contributor to a sound's salience. In so doing, I champion a distinction between *qualitative* characters – perceptual correlates of distal qualities – and *quantitative* characters – dimensions of experience that reflect attentional processing of perceptual representations.

1. Loudness and Sound Intensity

In this section I describe the manner and extent to which sound wave intensity predicts the perceived loudness of a sound. To begin, I must introduce a bit more precision into the notions of sound intensity and loudness. There are three relevant senses of 'sound intensity'. The first is *source intensity*, the intensity of the sound wave produced at the site of a distal event. Sound waves lose intensity as they travel, so the intensity of a sound wave at its source is typically greater than at the location at which it is perceived. The intensity of a sound wave as it interfaces with the sensory receptors of the ear is its *proximal intensity*. Proximal intensity is encoded by the auditory system in a manner that I am reluctant to call representational for reasons that will soon emerge, but at any rate this encoding results in a state of the auditory nerve that is indisputably responsive to intensive auditory stimulation. I will call the neural state that reflects proximal intensity, however indirectly, *encoded intensity.*

Three corresponding senses of loudness must also be distinguished: *source loudness, subjective loudness*, and *encoded loudness*. The distinction between source and proximal intensity gives rise to that of source and subjective loudness. Subjective loudness is how loud a sound is *to a subject,* its volume on her auditory field. Source loudness is instead the loudness of a sound event out in the world, and is distinct from source intensity; source intensity is an objective feature of a sound event, whereas assessments of source loudness are both subjective and subjunctive, viz., how loud a sound *would be* to a subject were she at the location of the sound event. For example, the subjective loudness of a siren on yonder mountain is softer than the siren's source loudness because the intensity of the sound wave that carries information from its source degrades as a function of distance; it would be painfully loud if the subject were at the siren's location.

Both subjective loudness and source loudness depend in part on encoded intensity. Nonetheless, encoded intensity is not a neural correlate of either source or subjective loudness. This is obvious with respect to source loudness; recovery of source loudness requires encoded intensity *and* some additional information about, e.g., the distance of the source of the sound from the subject.[1] So, source loudness (and source intensity) may be set aside. It is subjective loudness, not source loudness, that I claim reflects the attentional processing of sound representations. Encoded loudness, as I will call it, is the neural correlate of subjective loudness. In what follows, unqualified uses of 'loudness' should be taken to mean *subjective* loudness.

Encoded intensity is distinct from encoded loudness. In service of this point, I provide a sketch of the process by which sound intensity is encoded and demonstrate the extent to which encoded intensity departs from proximal intensity. I then describe various influences that transform encoded intensity into subjective loudness and describe the location and functional connectivity of encoded loudness as distinct from encoded intensity.

Sound perception begins with a vibration of the tympanic membrane by a sound wave. The tympanic membrane mechanically stimulates a series of small bones which themselves percuss on the entrance to the cochlea, a seashell-shaped organ lined with tiny hairs and filled with fluid that oscillates with vibrational input. Specific areas of the basilar membrane, a tonotopically organized rigid structure along the bottom of the cochlea, resonate according to the frequency of the input stimulation; the base of the cochlea resonates in response to the highest detectable frequencies, while the tip resonates with the lowest. Vibrations on the basilar membrane cause oscillations in the fluid of the cochlea that displace hairs in the vicinity of these vibrations. Displaced hairs transduce these mechanical stimulations into electrical ones; each hair in the cochlea synapses with a neuron that generates action potentials in response to

---

[1] Related to this is the phenomenon of loudness constancy; for at least some sounds, reports of source loudness remain constant – and accurate – even as distance from the source (and thereby proximal intensity) changes. The mechanism by which loudness constancy is achieved is not fully understood. One hypothesis is that source loudness is recovered from encoded intensity as a function of distance from the source, but this requires an accurate representation of source distance, which subjects systematically underestimate. Another possibility is that loudness constancy is achieved not as a function of distance information, but rather a function of information about reverberant sound energy recovered from the proximal stimulation (Zahorik & Wightman, 2001)

displacement, which then projects through a dedicated pathway through the auditory nerve.

Pitch is the perceptual correlate of sound wave frequency. Per the above, information about frequency is encoded by *which* auditory neurons fire. In addition, auditory neurons tend to phase-lock with the oscillations that trigger them, which means that sequences of action potentials by an auditory neuron are generated at the same "part" of each cycle of the stimulus wave; if an action potential occurs at a wave's peak, then subsequent potentials also occur at the peak of the cycle. Since auditory neurons are selectively responsive to specific frequencies, this means that the frequency of neural impulses from an auditory neuron matches the frequency of the tone to which it responds. So, information about pitch is also temporally encoded by the firing rate of auditory neurons. This temporal encoding has important implications for loudness for reasons that will emerge presently.

Loudness is typically held to represent sound intensity, a property of a sound wave closely related to its amplitude. There are no sound intensity receptors, which is to say receptors selectively responsive to sound wave amplitude; unlike pitch, the data relevant to sound intensity does not involve the direct transduction of any quality of a sound wave. Information roughly corresponding to sound intensity is instead recovered from total activity in the auditory nerve. That there is even an approximate correspondence between sound intensity and encoded intensity is a byproduct of pitch encoding: more intense sound waves cause more auditory neurons sensitive to the sound wave's component frequencies to be engaged as well as more frequent impulses from individual auditory neurons, resulting in more impulses being projected through the auditory nerve.

The more auditory nerve activity a sound generates, the greater its encoded intensity. The reason that encoded intensity is a systematically incorrect guide to sound intensity is that there are asymmetries in the potential of both individual and collections of frequencies to activate auditory neurons. Because action potentials from auditory neurons phase-lock to the frequency to which they are responsive, low-frequency tones do not generate as many action potentials as high-frequency ones. Suppose that a sound wave, by virtue of its very high intensity level, generates an action potential in a certain auditory neuron with every wave cycle. If that sound is a low frequency (and that neuron is thereby selectively responsive to that low frequency), then it produces only half the number of action potentials as a wave with double the frequency but the same intensity (and a corresponding auditory neuron) because the higher frequency tone has twice as many cycles in the same amount of time. In practice, individual neurons do not fire with every cycle of even very intense sounds, but collections of neurons that are all selective to a high frequency collectively fire more than collections of neurons selective to a lower frequency. Hence, higher frequency sounds are generally perceived as louder than lower frequency sounds when proximal intensity is held constant.[2]

By this sketch of auditory sensation, it is tempting to surmise that complex tones – sounds that are comprised of more than one frequency – should be systematically louder than equal-intensity pure tones. After all, if loudness is a matter of how many frequency-responsive neurons are activated, then a sound that involves more frequencies ought to activate more neurons, and thereby produce louder sound experiences, than one with fewer component frequencies.

Reality is a bit more complicated. Multiple frequencies within the same *critical band* – a range of audible frequencies about 1/3 of an octave in range, phenomenally speaking – activate some of the same neurons by virtue of vibrating the same area of the basilar membrane. For example, the frequency corresponding to middle-C activates some of the same neurons as the frequency corresponding to the adjacent C♯; neurons responsive to both frequencies do not consistently phase-lock to either.[3] Information about auditory receptor activity within the same critical band is pooled together and projected through channels that are bandwidth-specific, not frequency specific, meaning that total auditory nerve activity is determined by action potentials within critical bands rather than by individual frequencies. To the original point, it *is* the case that wide-bandwidth sounds – sounds that activate a wide range of critical bands – are systematically louder than sounds featuring fewer critical bands when sound wave intensity is held constant. This is because auditory neuron activity grows non-linearly with intensity; a tenfold gain in intensity produces only about a doubling in neural activity within the relevant channel, so spreading intensity out across more critical bands produces more activity overall (McDermott, 2013). For example, a 30 dB sound is ten times as intense as a 20 dB sound. So, ten pure tones in distinct critical bands of 20 dB each, played simultaneously, generate a wave with the same intensity as one pure tone of 30 dB. Nonetheless, the ten-tone wave will sound much louder than the pure tone because the neural activity produced in the pure tone's frequency

---

[2] See, e.g., Suzuki and Takeshima (2004). This trend reverses with very high frequency sounds for reasons that will take us too far afield to explain; for present purposes, it is enough to appreciate that encoded intensity, and thereby loudness, depends not only on proximal intensity but also, in part, on frequency.

[3] This is why playing both keys together on the piano sound like a single rough and beat-y tone, rather than two distinct tones as with C and E.

channel is only double what it would have been at 20 dB, but the ten-tone wave features ten of that 20 dB amount of activity, which sums to a total amount of activity five times the amount of the 30 dB pure tone.[4]

So, even at the level of encoded intensity there is a significant departure from proximal intensity; encoded intensity depends both on the frequency and bandwidth of the sound encoded. This is not a novel observation about sound intensity encoding; it is discussed in, e.g., O'Callaghan (2007) and Pautz (2015). While Pautz uses these facts as evidence against a tracking representationalist view of loudness, O'Callaghan instead posits that loudness represents the complicated property that encoded intensity in fact responds to – something that incorporates frequency, bandwidth, etc. But the discrepancy between loudness and proximal intensity is not exhausted by that of proximal and encoded intensity; loudness is also distinct from encoded intensity.

Evidence suggests that loudness may be subject to modulation by top-down influences. For example, Siegel and Stefanucci (2011) find that anxious subjects perceive tones to be louder that subjects in a neutral mood. It is not unusual for researchers to uncover evidence that negative emotional states exert some influence on perception, particularly with respect to "enhancing" sensation; for example, Siegel and Stefanucci note that researchers have discovered links between negative affect and, e.g., sensitivity to changes in spatial frequency, efficiency of visual search, and overestimation of vertical distances. It *is* unusual for emotional states to be found influencing an ostensibly primary sensory dimension; anxious subjects might overestimate heights and perform more efficient visual searches, but this does not require that anxious subjects undergo fundamentally different experiences than emotionally neutral counterparts, i.e., that anxiety alters the appearance of hue, luminance, etc. Perhaps, unlike all the other primary sensory dimensions, loudness may be modulated by emotional state. Or, more elegantly, perhaps loudness is not a primary sensory dimension at all.

Loudness is also modulated by the context in which a sound arises. Induced loudness reduction (ILR) is an auditory phenomenon in which identical sounds – tones featuring the same frequency, intensity, and duration – are not perceived as being equally loud when one of them follows an "inducer" tone of the same frequency and duration but greater intensity; the tone following the inducer sounds less loud than the

---

[4] These numbers are idealized for sake of demonstration. The amount of neural activity in a critical band is determined by myriad factors. For example, as I have already noted, the amount of neural activation caused by a tone depends in part on the tone's frequency, so the lower-frequency critical bands contribute less neural activity than the mid-range ones.

other tone by as much as half (Arieh & Marks, 2010). This effect cannot be attributed to a low-level perceptual process (e.g. receptor fatigue), for, among other things, the effect persists in individuals with cochlear implants whose auditory processing bypasses the cochlea (Wang et al., 2015). Identical stimuli that cause identical encoded intensity states may nonetheless sound loud to different degrees.

A recent study by Schmidt, Mauermann, and Kollmeier (2020) exploited ILR to investigate the location of loudness encoding in the brain. It is usually difficult to distinguish whether a brain state corresponds to encoded intensity or encoded loudness since the former strongly influences the latter, but because ILR effects involve identical encoded intensities, any brain region that differs in activation between contexts – that is, whether or not the target tone is preceded by an inducer tone – is a candidate for a loudness-encoding state. Schmidt *et al.* took EEG measurements of subjects during each context to determine where (and when) loudness is encoded during auditory processing. A state correlating with subjective loudness is not observed until late in auditory processing but nonetheless within its bounds, in the auditory cortex. All things considered, this means that ILR is not an effect of a high-level judgment, and hence cannot be attributed to a decision bias; loudness encoding is genuinely distinct from the intensity encoding that occurs early in auditory processing. This finding is consistent with results from studies that observe the fMRI correlate of loudness in the auditory cortex. In one study by Röhl and Uppenkamp (2012), subjects were asked to evaluate the categorical loudness of pink noise stimuli of various intensities from within an MRI scanner. Activity in the ascending auditory pathway increases as a function of sound intensity, but a linear covariation with loudness is detected only in the auditory cortex. The researchers find no evidence that loudness is represented in the brain at any point in auditory processing prior to the auditory cortex which, it bears repeating, conducts *high-level* auditory processing.

An unexpected result of this fMRI study is significant. In their concluding remarks, Röhl and Uppenkamp observe that

> a range of sound pressure levels of almost 30 dB was rated as similarly loud across the normal-hearing listeners. This range of individual differences in loudness sensation within the examined subject group is by far larger than would be expected for the results from other controlled psychoacoustic experiments…None of the variables we registered in our study (e.g., hearing thresholds, age, personality traits, musical ability, and experience) seemed to be able to explain a significant amount of the observed variance in loudness

> sensation…On the other hand, it is very unlikely that this variance is only a result of methodological issues of categorical loudness scaling, since these individual differences were shown to be linked to individual differences in neural activation of the [auditory cortex], as reflected by the BOLD signal in auditory fMRI. (p. 378)

So, there is huge variation across subjects in the intensity of sounds described as being similarly loud, but the extraordinary variability of subjects' loudness reports is substantiated by regularity in loudness ratings and brain activity corresponding to encoded loudness. In other words, sound intensity is a poor predictor of encoded loudness, but encoded loudness is a good predictor of subjective loudness. Thirty dB represents a *thousand-fold* difference in intensity between the top and the bottom of the range. This degree of variability is inexplicable if loudness represents sound intensity, *especially* since subjects start out with more-or-less the same encoded intensity from the proximal stimulus.

This concludes my survey of evidence against an informational relationship between sound intensity and loudness. In brief, information about sound intensity is lost in the process of intensity encoding, encoded intensity does not fully determine loudness, and there are vast interpersonal differences in loudness experiences that are difficult to account for.

2. Loudness and Salience

Nonetheless, it must be respected that encoded intensity, and by extension sound intensity, is the principal driver of loudness. In this section, I present evidence in support of the idea that the best way to account for the fact that loudness heavily depends on but does not represent encoded intensity is to identify loudness with the salience of a sound representation and encoded intensity as a feature that influences auditory salience. I will say more about salience and attention in subsequent sections; for now it is enough to think of the salience of a sound representation as the amount of attentional processing it receives.

The first and simplest point in favor of this view about loudness is that experimental results, to the extent that we have them, suggest that loudness is highly correlated with salience. The methodology for studying auditory salience is still under development. Compared to visual attention, which involves eye movements to fixated locations and hence can be studied via eye tracking, it is difficult to establish when an auditory stimulus successfully captures attention. Captures of auditory attention have not been demonstrated to involve any obvious physiological change that can be exploited in auditory attention experiments. As a result, studies of auditory salience tend to rely on subjective reports of salience.

That said, one potential guide to auditory attentional capture is pupillary response; a handful of studies indicate that pupils dilate in response to deviant auditory stimuli.[5] In a series of three experiments designed to investigate this connection, Liao *et al.* (2015) recorded subjects' pupillary responses to various sounds and asked subjects to evaluate, among other features, the salience and loudness of each sound. The sounds had equal intensity in the first experiment, equal encoded intensity in the second experiment, and variable intensity in the third experiment. Across all three experiments, judgments of salience and loudness are found to be highly correlated with one-another, leading the researchers to conclude that "[a]lthough the aim of the study was to investigate the effects of subjective salience on pupillary responses, the results demonstrated that salience is indicative of, or is heavily influenced by, loudness," (p. 423).

Another aspect of this study is worth remarking on. Pupillary response is found to covary with loudness/salience judgments in the first and third experiment. In the second experiment, pupillary response covaries with encoded intensity, but not with loudness/salience judgments. Notably, what I call "encoded intensity" is called "loudness" in Liao *et al.*'s (2015) study, a convention in the literature on auditory attention that surely arises from the implicit assumption that loudness represents sound intensity. Liao *et al.* claim to have "adjusted the sound pressure level of each sound so that all of the sounds would have the same loudness on the basis of the loudness model developed by Glasberg and Moore," (p. 418). Because of the idiosyncratic way that early auditory processing encodes intensity, equal loudness of very different sounds must be achieved by making their objective intensities *un*equal. Glasberg and Moore's (2002) model reproduces the effects of early auditory encoding on sound waves and hence may be used to determine the expected *encoded intensity* of a sound, not the sound's loudness and, per the previous section, not its encoded loudness. Calling encoded intensity 'loudness' introduces some confusion into Liao *et al.*'s conclusions. For example, they claim that "[i]n Experiment 2, when the sounds were equally loud, the pupillary dilation responses were similar to each other and *did not correlate with salience or loudness*" and yet that "the overall results [of all three

---

[5] E.g. Wang and Munoz (2014), Liao *et al.* (2016), Huang and Elhilali (2017).

experiments] suggest that the subjective judgment of salience is more or less equivalent to loudness, and that *the pupillary dilation response reflects both*," (p. 421, emphasis added). Pupillary dilation cannot both "reflect" and "not correlate" with salience/loudness.

Indeed, in light of Experiment 2, the overall results of Liao *et al.* suggest that pupillary dilation is a reflection not of loudness/salience, but of encoded intensity. Encoded intensity is, of course, a principal driver of loudness/salience, but, as we have seen, encoded intensity is non-identical to encoded loudness. If pupillary response covaries with encoded intensity, then the results of all three experiments conform to expectations: pupillary response covaries with both encoded intensity *and thereby* loudness/salience in conditions of varying encoded intensity (experiments 1 and 3) since encoded intensity exerts a dominating influence on loudness/salience such that the two often covary. However, pupillary response does *not* covary with loudness/salience in conditions of equivalent encoded intensity (experiment 2) since, as with Schmidt *et al.*'s ILR experiment, it is under these conditions that distinctions between encoded intensity and loudness emerge. This interpretation is consistent with a later study by Huang and Elhilali (2017) which found that, though all salient auditory events correlated with pupil dilations, only about 29% of pupil dilations correspond to salient auditory events. In other words, auditory-event-caused pupil dilations include but are not limited to salient sound events. However, an analysis of all pupil dilations finds that they are "correlated significantly with increases in acoustic loudness," where acoustic loudness refers again to the output of a function that models the computational behavior of early audition – encoded intensity, that is.

In §1 I presented evidence that loudness does not covary with sound intensity in a manner that suggests a representational relationship between the two; intensity is encoded by piggybacking on processes that subserve pitch encoding, and as such returns a value of encoded intensity that at best approximates sound intensity. Indeed, loudness does not even covary with *encoded* intensity, as is demonstrated by EEG and fMRI studies that identify loudness encoding at the high level of the auditory cortex. In the fMRI study, it is discovered that there is a high degree of correlation between categorical loudness ratings and encoded loudness; correspondence between categorical loudness and encoded intensity are much more variable. The EEG study finds that the ILR phenomenon is first reflected by electrical activity corresponding to activity in the auditory cortex. Liao *et al.*'s results confirm that loudness and salience covary even if the link they set out to demonstrate, between pupillary dilation and salience, is less direct. To my knowledge, no other study has directly investigated the relationship between subjective loudness and salience.

In the remainder of this section I demonstrate that a parsimonious account of all these data points is available if loudness a reflection of a sound representation's salience. I begin with a discussion of the feature integration theory of attention, and how it has been adapted to model auditory salience.

Feature Integration Theory ("FIT"), Treisman and Gelade's (1980) influential model of bottom-up visual attention, has inspired some highly effective computational models of visual attention.[6] On these models, visual attention takes its inputs from "feature maps," two-dimensional topographic maps of the visual field for visible features like color, luminance, and orientation. Each of these maps encodes the intensity of a certain feature across the visual field. Information about locations on each feature map with the highest amount of contrast – that is, relatively large local differences in activation on a given feature map – is passed on for the construction of a master "saliency map" that encodes conspicuous locations. More contrast detected at a location across various feature maps results in a more conspicuous location on the saliency map. Attention then processes locations in order of their priority on the saliency map.[7]

Recent interest in developing a computational model of auditory attention has resulted in several models based on saliency maps. The earliest of these, from Kayser *et al.* (2005)*,* proposes that the features contributing to the auditory salience map include intensity, frequency contrast, and temporal contrast. To test this model, recordings of complex auditory scenes were evaluated for saliency on the presumption that saliency is determined by extraction and analysis of the features just noted. For each auditory scene, researchers extracted each of these features from waveforms in a manner that emulates the behavior of auditory neurons. The level of contrast apparent in the representation of each feature was analyzed and an aggregate assessment of saliency surmised. Next, participants were presented with pairs of these auditory scenes – one in each ear. They were prompted to indicate whether one scene featured a more salient sound event than the other and, if so, which one. The intensity filter by itself is correlated with participant responses, but not as strongly as the predictions of all three features together. This supports the hypothesis that encoded intensity is an

---

[6] See, e.g. Itti and Koch (2000), Itti (2005).

[7] This simplified description eschews interesting and important details regarding, e.g., whether and how top-down attention influences this process, how bottom-up inputs from various feature maps are aggregated, and whether visual attention is subserved by one or more than one saliency map. For further discussion, see e.g. Treisman and Sato (1990), Itti and Koch (2000), and Burrows and Moore (2009), respectively.

important but non-exclusive contributor to auditory salience. If there is a saliency map for audition, then one of the feature maps on which it depends is surely drawn from encoded intensity. [8]

During Kayser *et al.*'s trials, participants were also asked to indicate which feature of the sound judged as more salient drove their judgment: frequency structure, temporal structure, or intensity. The contribution of frequency contrast predicted by the saliency map model was significantly larger in auditory scenes for which frequency structure was reported as the basis of subjects' judgments, and the same is true, *mutatis mutandis,* for temporal contrast and structure. Put another way, sounds judged as more salient in virtue of their frequency or temporal structure have higher levels of frequency or temporal contrast, respectively, than sounds judged as salient by virtue of another feature. Not so for intensity; there is no significant difference in the intensity of stimuli reportedly selected based on their intensity than for stimuli selected on some other grounds.

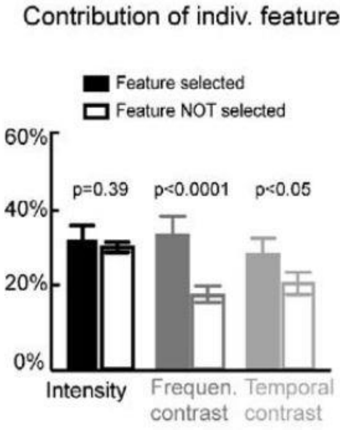

Figure 1. Contribution of individual features to saliency. *Note.* Per Kayser *et al.*, "Solid bars indicate the contribution of each feature to the total saliency for trials on which the subject indicated a rely on that feature, and open bars indicate the contribution on all other trials. Bars show the mean and s.d. across subjects. P values refer to t tests." Importantly, intensity was

---

[8] Since Kayser *et al.*'s study, several saliency map models of auditory attention have demonstrated improved predictive power by incorporating additional feature maps. Even in these improved models, intensity (sometimes under the guise of 'envelope' – the "shape" of a waveform from which intensity information may be gleaned) – remains a dominant factor in salience prediction. See, e.g., Kaya (2012).

variable across trials. This figure appears in Kayser *et al.* (2005). Copyright Elsevier 2005. Reprinted with permission.

Kayser *et al.* suggest that this asymmetry is due to the fact that

> any feature is dependent on intensity (zero intensity implies no other features exist). Thus, a feature like frequency or temporal contrast will always be somewhat confounded with intensity. (p. 1944)

In other words, they speculate that stimuli identified as having intensity as their most salient feature are, on average, no more intense than stimuli identified as having a different most-salient feature because detection of those other features requires that they have some level of intensity. So, a sound that is identified as having, say, temporal contrast as its most salient feature might also be rather intense, inflating the average intensity of sounds not identified as having intensity as their most salient feature.

This explanation is unsatisfactory. Though it is true that "zero intensity" implies that that there are no other features of a sound (indeed, that there is no sound wave), this is also true of frequency and temporal structure. Frequency is as much a fundamental feature of any wave as intensity, and by virtue of having frequency a sound wave is necessarily temporally extended. This is true also with respect to phenomenal dimensions of sound: apparent pitch/timbre, temporal structure, and loudness are mutually entailing. To the point in the preceding paragraph, sounds identified as featuring intensity as their most salient feature might also have a high degree of temporal and/or frequency contrast, artificially raising the average contribution of those features in cases for which they were not selected as the most salient feature. So, the fact that intensity is confounded with frequency and temporal structure does not explain the asymmetry, for the same may be said of the other two features.

Summing up so far, auditory attention relies heavily on encoded intensity; a saliency model that relies only on intensity as encoded by the auditory system performs nearly as well as one that incorporates other features. Nevertheless, subjects perform no better than chance when reporting whether intensity is the primary driver of their assessments of salience despite a capacity to successfully do so with respect to other salient auditory features; the actual contribution of intensity to salience does not predict subjects' reports of what they believe drives their salience judgment.

One possible explanation for this asymmetry is a bias for choosing 'intensity' when subjects are not sure which feature of a sound is responsible for its salience. Kayser *et al.*'s method does not include an "unsure" option for subjects who do not know which feature drove their salience judgment. Perhaps subjects answered "frequency structure" when frequency contrast was especially apparent, "temporal structure" when temporal contrast was especially apparent, and "intensity" otherwise. If that is right, then "intensity" judgments would apply not only to sounds they felt were salient in virtue of their intensity, but also to those sounds for which no aspect of the sound was especially apparent. However, this explanation raises a different question: why would subjects asymmetrically choose 'intensity' when they are not sure?

It seems to me that either asymmetry may be explained in the same way. Perhaps subjects genuinely believe that intensity drives their judgements in all such reports, or perhaps some of those reports are disguised 'unsure' responses. In either case, subjects surely use loudness as a guide to intensity; no other feature of auditory experience is even a candidate for the job. Nothing other than loudness can explain why subjects report that the intensity of a sound drives its salience. This means that loudness, unlike frequency structure and temporal structure, is not a good guide to whether the feature it ostensibly represents drives a sound's salience. As I have discussed, *some* processing beyond intensity encoding can affect loudness, but surely not so much that subjects should fail so spectacularly at identifying events that are salient in virtue of their intensity when they are evidently capable of doing so on behalf of other auditory features. Loudness, in some fundamental way, is *unlike* pitch or tempo.[9]

Suppose, then, that loudness is not a representation of intensity, but rather reflection of a sound's salience. On this view, there is an available explanation for subjects' unreliability with respect to whether their auditory salience judgments are based on intensity. If, unlike pitch and temporal structure, we do not phenomenally represent sound intensity, then we should *expect* that subjects perform no better than chance at judging whether intensity is a salient feature of a sound. We cannot accurately report on qualities that we cannot perceive.

This cannot be the entire explanation, of course, for participants could also respond with 'temporal structure' or 'frequency structure'. The fact that subjects sometimes responded 'intensity' even though sound intensity is never perceived calls

---

[9] Another possible explanation is that the model is wrong with respect to encoded intensity's contribution to auditory salience. This seems unlikely. In order to be wrong in a way that delivers these results – that there is *no* difference in the influence of encoded intensity on salience between 'rely' and non-'rely' judgments – instead of expected results – that 'rely' judgments involve more influence of encoded intensity on salience than non-'rely' judgments – then the model would have to be dramatically wrong about how intensity-based-salience is determined. Given that their model on which intensity *alone* is able to predict which stimulus is judged as more salient nearly as well as the more complicated model, it would be surprising if they were deeply mistaken about determining salience from intensity.

out for explanation. The reason that subjects sometimes say intensity made the difference is that they are using loudness as an indicator of sound intensity. If loudness reflects a sound's salience, then it reflects not only salience accrued on the basis of intensity, but also the salience accrued from other qualities of the sound. When the salience of a sound is due in part to features other than intensity – temporal structure, frequency structure, or perhaps features that have not yet been identified as contributors to the auditory saliency map – then there will be a mismatch between a sound's intensity and its loudness. Subjects report (correctly) that the louder of two sounds is more salient, but err in thinking that loudness indicates the sound's intensity.

Similarly, if a subject is not sure which feature of a selected soundscape is principally contributing to the sound's salience – if the pitch or the temporal structure is not especially notable – and if she is not keen to the distinction between loudness and intensity, then she is likely to report that the sound's intensity drove her judgment since the prevailing feature of her sound experience is its loudness. When a subject reports that a sound's intensity drove a salience assessment, she reports, in effect, that the most salient aspect of the sound is its salience. It is no wonder, then, that these assessments fail to meaningfully reflect the contribution of intensity alone.

Here, then, is my proposal. Encoded intensity does not represent proximal (or distal) intensity. However, in absence of a state that more accurately reflects sound wave intensity, and in virtue of the fact that nearby sound events tend to be more intense and are more likely to be behaviorally relevant than distant sound events, encoded intensity is highly useful for determining how much attentional processing a sound representation receives. If auditory attention involves a saliency map, then there is a feature map corresponding to encoded intensity. Encoded intensity contrast exerts a dominating but not exhaustive influence on the auditory saliency map, which is why loudness, a reflection of the salience of a sound representation, tends to (only) approximately correlate with encoded intensity. Other factors that likely influence auditory salience are frequency contrast and temporal contrast, and the explanation for ILR phenomena is that the tone with its loudness reduced is less salient by virtue of insufficiently contrasting with the inducer tone.[10] The saliency map for auditory

---

[10] The ILR study is interesting for another reason. The component of the auditory evoked potential (AEP) that correlates with loudness is the N1-P2 deflection; reductions in loudness correspond to decreasing N1-P2 amplitude. The N1 component of the N1-P2 complex is associated with change detection, including the occurrence of deviant and oddball stimuli (Pratt, 2011). The N1 potential begins around 100 ms after stimulus onset and is believed to be generated by feature traces antecedent to integrated representations of auditory objects. It is hypothesized that feature integration of auditory objects occurs between 150-200 ms after stimulus onset, which overlaps with the end of the N1 component and the beginning of the P2 (Näätänen & Winkler, 1999). In other words, AEP activity that

attention encodes the degree to which various "locations" on the auditory landscape are salient, and the degree to which an auditory object is salient is the degree to which its "location" is encoded as salient.[11] In short, sounds are salient because they are intense (among other things), and loud because they are salient.

Summing up, the view that loudness is a quantitative character has the following advantages over the mainstream view, viz. that though loudness is a systematically *inaccurate* representation of intensity, it is nonetheless a feature representation – one that merely *contributes* to salience. First, the discrepancy between loudness and sound intensity is due not only to systematically inaccurate intensity encoding, but also to effects of downstream auditory processing; some sounds differ in loudness even though their intensities are equivalently encoded. In these cases, the best explanation for the processing discrepancy has to do with the relative salience of the sounds, i.e., that the target tone of the ILR phenomenon is softer because it is qualitatively the same as the inducer tone, and so the lack of novelty diminishes its salience and hence its loudness. Second, there is tremendous interpersonal variability with respect to loudness. This is unusual with respect to feature representations, but typical of experiences that are differently modulated by attention. Third, though intensity is a strong predictor of auditory salience, subjects are not able to tell when intensity is responsible for their salience judgments as they are when some other audible feature is responsible. If loudness was a feature representation, then subjects ought to be able to identify whether loudness was the principal contributor to their salience judgments, given that they can do so with respect other auditory features. If, on the other hand, loudness reflects the salience of sound representations, then using loudness as a barometer for encoded intensity is expected to be prone to error. To my knowledge,

---

corresponds to loudness also corresponds to the processing immediately before and throughout the generation of a feature-integrated percept. This is precisely where we should expect to see loudness-related activity on the view that loudness reflects salience. After all, the saliency map is the guide by which bottom-up attention directs its resources, and, on FIT, it is this directing of attention that binds features at the attended location. If there is a saliency map for auditory bottom-up attention, we should expect it to arise at the interface of feature traces and feature-integrated auditory objects, and it would seem that the N1-P2 complex is where we should expect to locate that sort of processing. This does not, of course, constitute evidence that loudness is a presentation of salience; there is much more to say about the N1-P2 complex than I have discussed here. I raise it only to suggest that current understanding of the N1-P2 complex is consistent with loudness reflecting salience.

[11] Näätänen and Winkler (1999) point out that one way in which feature-integrated auditory objects differ from feature-integrated visual objects is that the "medium" of object formation is space while the "medium" for auditory object formation is time. Hence, the analog of a saliency map for auditory attention is unlikely to be a topographic spatial representation, and so the notion of a "location" is here an analogy for whatever it is that "temporal maps" encode.

the view that loudness is reflects salience is consistent with the current state of empirical work on loudness and auditory processing. So, I contend that the most parsimonious explanation for the foregoing considerations is that loudness is not a representation of a feature of sounds, but rather a reflection of the salience of sound representations.

3. Quantitative Character

Though the terms ‘qualitative character’ and ‘phenomenal character’ are typically used interchangeably, the phenomenal is not exhausted by the qualitative; there is also *quantitative* character. Quantitative character is, roughly, the apparent intensity of an experience. I do not claim it follows from the mere fact that we might describe some aspects of phenomenal character as quantitative that the phenomenal is not exhausted by the qualitative. For, even if some aspects of experience feature intensity, it does not follow that these aspects are not subsumed by the experience’s qualitative character as it is typically understood. By way of analogy, it would be obviously mistaken to argue that the phenomenal is not exhausted by the qualitative on the grounds that some experiences have a *reddish* character, and therefore phenomenal characters are either qualitative or reddish. Being reddish is a way of being a qualitative character, and so the presence of reddishness does not preclude (and in fact entails) the presence of qualitative character. In principle, the apparently quantitative might be explainable in terms of the qualitative; experience alone does not reveal whether the intensity of a pain may be understood as an aspect of the experience’s qualitative character.

Neither do I claim that all manner of phenomenal intensities are instances of quantitative character. One color swatch might look *more blue* than another, and so an experience of one features a higher “degree” of blueness than an experience of the other swatch. Nonetheless, apparent hue is a qualitative character; apparent degree of blueness does not feature the right kind of content, nor is it underscored by the right kind of functional state, to be a quantitative character. In §3.2 I describe how quantitative characters may be distinguished from qualitative characters both phenomenally and functionally.

Quantitative characters do not reduce to representational states. Instead, they reflect the salience of the representation to which they are attached. In the foregoing sections I argued that loudness is a quantitative character. In this section, I clarify what quantitative character is. I begin with a discussion of prothetic and metathetic sensory dimensions since quantitative characters bear significant similarities to prothetic dimensions of experience. Despite their apparent similarity, there are subtle but important ways in which quantitative character is distinct from prothetic qualitative character, both phenomenally and functionally. I describe these differences and, in so doing, explain the core features of quantitative character.

3.1 What Quantitative Character Is Not

The distinction between qualitative and quantitative character resembles that of metathetic and prothetic sensory dimensions in psychophysics. According to Stevens (1957), metathetic sensory dimensions report on "what kind and where" and are "mediated by a physiological process that is substitutive." On the other hand, discrimination across a prothetic sensory dimension reports on "how much" and is mediated by "an additive or prothetic process at the physiological level," (p. 154). For example, hue is a metathetic sensory dimension, while brightness is a prothetic sensory dimension. We perceive progression along the dimension of hue as involving qualitative changes from blues through yellows, then greens, then reds. Progression along the dimension of brightness, however, feels like a series of degrees, not qualities; stimuli look *more* luminous as the continuum progresses. Note that it does not make sense to think of red as being *more hue-ful*, as it were, than blue.

Stevens held that at the physiological level, the process underlying prothetic dimensions of sensation is *additive*, which is to say that increases in the experienced dimension are evoked by increases in the total activity of the relevant sensory receptors. In contrast, he claimed that the physiological process underlying change in a metathetic dimension is substitutive, relying on a change not in total, but *type* of receptor activity – *which* receptors are activated, not how many.

Empirical investigation has revealed this claim to be overly simplistic in the more than sixty years since Stevens introduced the distinction between metathetic and prothetic sensory dimensions. For example, stimulation of photoreceptors in the retina results in their being hyperpolarized rather than depolarized, and resultantly causes a *decrease* in their firing rate. Hence, it is inhibition rather than activation of photoreceptors that carries information about increases in light intensity (Kandel, 2013). Stevens was sensitive to the possibility that further study might undermine this claim, noting that

> [w]hether all perceptual continua that behave in the prothetic manner are mediated by additive physiological processes is not certain, of

> course, but in at least some instances it seems evident that the existence of two basic kinds of physiological mechanisms is reflected in the behavior of the psychological scales and functions which we construct from subjective measurements in the sensory domain. (1960, p. 234)

The how much/what kind distinction and the additive/substitutive physiological process distinction were intended by Stevens as mere heuristics that only roughly identify prothetic and metathetic sensory dimensions, respectively.

The constitutive difference between prothetic and metathetic dimensions of experience, he maintained, is that apparent magnitudes of metathetic dimensions bear linear relations to magnitudes of the external continua they represent, whereas the magnitudes of prothetic continua are related to the magnitudes of their external continua by power functions.[12] For example, pitch is a metathetic dimension – differences in apparent pitch vary linearly with differences in the frequency of a sound wave. Contrastingly, the sensation of electric shock is a prothetic dimension – a doubling of electric current through the fingers results in a sensation roughly ten times as strong (Stevens, 1960).

As we will see, quantitative characters are akin to prothetic sensory dimensions in each of the three ways described above. Both are defined in part by their coming in degrees, both (typically) involve additive physiological processes, and neither is linearly related to the magnitude of any external content. In at least these respects, quantitative characters differ both phenomenally and functionally from metathetic sensory dimensions like hue and pitch.

Nonetheless, further differences indicate that quantitative characters are not prothetic dimensions of experience. In §3.2 I explain how quantitative dimensions of experience differ phenomenally from prothetic continua. In §3.3 I explain how they differ functionally.

In each section I appeal to loudness as an exemplar of quantitative character. In holding that loudness is a quantitative character, I defy Stevens, who appeals to loudness as a paradigmatically prothetic continuum. However, Stevens did not have in mind the further distinction between the prothetic and the quantitative for which I

---

[12] Hence, it is in principle discoverable that an apparently metathetic sensory dimension – one that bears the heuristic hallmarks described above – is actually prothetic, and vice versa. For example, Stevens was surprised to discover that apparent saturation does not bear a linear relationship to objective saturation, and thereby counts as a prothetic sensory dimension by his lights (Panek & Stevens, 1966).

will argue. Once this distinction is drawn, it will be clear that loudness is a quantitative character, not a prothetic dimension of experience.

3.2 The Phenomenal Distinction

One way of characterizing the phenomenal difference between metathetic continua like pitch and hue and prothetic or quantitative dimensions of experience is that metathetic sensations from distinct perceptual modalities appear irreconcilably different from one-another. For example, a middle-C experience bears no intrinsic similarity to a red experience. In this way, metathetic continua appear to feature a *proprietary phenomenology* – a way of appearing that is wholly unique to and constitutive of the modality in which it arises. Red experiences are uniquely and constitutively visual, middle-C experiences are uniquely and constitutively auditory. By this I mean something more than what may be trivially ascertained from the fact that we *call* the colors 'visual' and the pitches 'auditory'; the point is that it is inconceivable that any quality of a metathetic dimension in one modality could arise in any other modality, e.g. that a red experience could be seamlessly incorporated into the auditory landscape amid various pitches. More generally, each determinate value of a metathetic determinable is, by definition, distinct from each other determinate of that determinable, but also akin to each of those other determinates in a way that is not true of the values of the metathetic determinables in any other modality. Though red is distinct from green, both red and green belong together on a spectrum that cannot accommodate qualities from other metathetic dimensions. Even if phenomenal characters were nothing more than mental paint, it is evident that not all of those paints may be mixed.

On the other hand, prothetic (and quantitative) continua are distinct from one-another only insofar as the qualities they comment on are distinct; loud middle-C and saturated red appear quite different because pitch and color are quite different, but the bare intensities, the "how much" of the experiences are otherwise phenomenally comparable. We may describe both as "very intense" without the need to explain further. So, unlike metathetic dimensions of experience, prothetic and quantitative dimensions do *not* feature proprietary phenomenal character.[13] In general, the values

---

[13] Indeed, it is an open question in neuroscience whether our brains feature a multisensory magnitude estimator; Baliki *et al.* (2009) suggest that the insula may be a hub for "how much" representation. It would be beyond the pale to consider such a possibility if magnitudes across modalities were not phenomenally comparable.

of both prothetic and quantitative dimensions fall on a one-dimensional quality space arranged from "none" to "maximum."

If neither prothetic nor quantitative dimensions have proprietary phenomenal character and present as simply a range of intensities, then it stands to reason that they cannot be phenomenally distinguished in themselves. Indeed, determinate values of prothetic and quantitative character both *appear* as a degree to which a represented item bears a certain feature. However, there is one way of phenomenally distinguishing the prothetic from the quantitative: the phenomenal consequence of their null values. For a prothetic dimension of experience, we can imagine being aware of an apparent object featuring a null value on that dimension. For example, consider looking at an item coated in ultrablack carbon nanotube coating, a material that absorbs 99.995% of incoming light. This coating is the blackest material on Earth. It is as close to 0% on the scale of luminance as any visible surface may achieve. Nonetheless, an item coated in ultrablack carbon nanotube coating is perfectly visible; it can hardly be missed against a contrasting background. Even if the item in fact absorbed 100% of light (and was thereby 0% luminous), the item would be clearly visible in most conditions. So, an item's being 0% luminous does not entail that we cannot be visibly aware of it.

This is not so for quantitative continua. If there is no degree to which a sound is loud, then the sound is *inaudible*, and we cannot be aware of inaudible sounds. Similarly, we cannot be aware of a pain that is painful to no degree, or of perfectly camouflaged items that do not "pop out" from their surroundings.[14] Reducing a quantitative dimension to nothing is tantamount to expelling its object from the phenomenal landscape of the relevant modality. Quantitative characters are a scale of the phenomenal prominence of the *whole represented item*, not just one of its qualities.

3.3 The Functional Distinction

The phenomenal difference described above arises in virtue of a difference in the sort of thing each kind of intensive continua quantifies. Determinate values of prothetic dimensions of experience represent magnitudes of some-or-other property. Quantitative characters do not; they are, in a sense to be elaborated, magnitudes of representations *themselves*. I will argue that quantitative character is determined by the degree to which a qualitative representation is salient.

---

[14] I hold that pain intensity and the phenomenology of visual pop-out are quantitative characters, but it is beyond the scope of this paper to argue that this is so.

Before saying more, I must explain what I mean by 'salient'. Wu (2014) distinguishes between "phenomenal salience" and the sense of 'salience' employed in psychology. Phenomenal salience is the claim that "the phenomenology of attention is the rendering of the attended object as phenomenally salient." The psychological sense of salience "refers to a property of a stimulus that draws attention to it," (p. 127). This distinction arises in the context of a discussion about whether there is a unique phenomenology of attention: if attention does not have a unique phenomenology, then the apparent phenomenal difference in, say, covert shifts in visual attention between one object and another must be explained in terms of something other than attentional differences. The term 'phenomenal salience' is just a placeholder for *that kind* of phenomenal difference. It is an open question whether phenomenal salience is a primitive phenomenal feature or is reducible to something else, e.g. a content difference.

Psychological salience, on the other hand, plays a vital role in empirical investigations of attention. Experiments on attention presuppose that when items grab attention, they do so by virtue of their properties, and hence that attention-grabbing stimuli have salient properties. That some properties are psychologically salient is a basic assumption that underlies the corpus of empirical work on attention. Psychological theories of attention make no demands on the nature of attentional phenomenology; it may be that attention has a proprietary phenomenology, but it may also be that attention affects the overall phenomenology of an experience only by altering experience content. Indeed, the psychological conception of salience is compatible with there being no attentional phenomenology whatsoever. Hence, psychological salience is not beholden to any specific view about phenomenal salience.

Ultimately, quantitative character is an account of phenomenal salience. However, my approach to phenomenal salience is unusual. My goal is not to explain phenomenal salience, but rather to explain a certain class of phenomenal characters – quantitative characters – that are typically presumed to be perceptual. It is a consequence of my analysis that quantitative characters are instances of phenomenal salience; I do not assume from the outset that there is such a thing. To that point, I will never use the term 'salience' to refer to the phenomenal character of a state, unless explicitly noted. The sense of 'salience' that I employ is closely related to the psychological one, though differences between my use of 'salience' and its typical application in psychology will emerge as I go. For now it is enough to say that salience, as I intend the term, is the factor by which representations are granted access to attentional processing.

Returning now to the functional difference between prothetic and quantitative continua, a distinction emerges with respect to the explanation for the nonlinearity of the relationships between each kind of intensive character and correspondent physical magnitudes. Stevens claims that there is a nonlinear relationship between the magnitudes of physical continua and the prothetic continua to which they are related. This is also true of quantitative continua, but for a different reason than Stevens posits for prothetic continua. According to Stevens (1960), the power function relating a prothetic continuum to its domain reflects a transformation performed by sensory transducers. The value of the exponent describing the relationship between the intensity of an external feature and its correspondent apparent intensity varies from feature to feature. Some functions increase by a power less than one – they are compressive – while others increase by a power greater than one – expansive. This, Stevens hypothesized, is due to sensory transducers being designed in such a way that the experiences they give rise to may adequately cover a wide range of physical intensities (in the case of compressive functions) or present small but biologically significant changes in a narrower domain as more pronounced (in the case of expansive functions).

Contrastingly, the reason that quantitative characters are not linearly related to correspondent physical magnitudes is that quantitative characters do not represent external features at all; the phenomenal feel of quantitative character reflects the magnitude of an internal, functional state of the attention system. So, it is unsurprising that the determinate values of quantitative characters do not bear linear relations to physical magnitudes. The question relevant to quantitative characters and physical magnitudes is why they bear *any* functional relation to one-another at all.

The reason for this relation is that salience is mediated by sensory representations. When a distal stimulus is represented as salient, it is not because the stimulus instantiates *salience*, for salience is not a property of distal stimuli. Consider Susan, who overhears her name at a party. When Susan hears her name, her attention is automatically pulled to the speech stream in which it occurs; for her, the 'Susan' sound is a highly salient stimulus. For most others, the 'Susan' sound is not so salient since most people are not called 'Susan', but they are sensitive to mentions of their non-'Susan' names in a way that Susan is not. The sound of one's name is not highly salient by virtue of its physical properties alone, but rather because one bears a particular relation to that sound. Representations of some properties of distal stimuli enjoy widespread salience in human attentional economies, but this does not entail that these properties are mind-independently salient. Humans with normal hearing find high intensity sounds to be highly salient. But, as with the 'Susan'-sound, there is nothing *objectively* salient about any given sound intensity level; a high-to-us sound intensity level

– 90 decibels, say – is salient *to each of us*, but maybe not to a creature whose hearing is differently attuned to sound intensities.

Since salience is not a distal property, information about salience is not detectable by sense organs. Why, then, is there a correspondence between quantitative and distal continua? Why does loudness covary with sound intensity if loudness reflects a mind-dependent feature? This is because salience is accrued by representations of distal properties. Because representations are informational states of sensory systems, and because salience is accrued by a representation on the basis of the distal features it represents, quantitative character ends up correlated with those objective magnitudes that are factored into determinations of salience.

4. Salience and the Attentional Economy

Salience has traditionally been understood as pertaining solely to bottom-up attention. Within the domain of bottom-up attention, salience is afforded variously to distal stimuli, stimulus features, representations of stimulus features, locations, and actions: a stimulus is salient when its features attract attention to it; a stimulus feature is salient when it attracts attention; representations of a stimulus feature are salient when they enter into computational processes that allocate attention (e.g. feature maps that feed a saliency map); locations are salient when the saliency map directs attention to them; actions are salient when they are driven by the saliency map.[15] In typical use, 'salience' broadly describes causal relations between attention and its inputs, so salience need not be ascribed to one of these things at the exclusion of the others. Representations of salient features (which are typically caused by salient distal features, which are borne by salient distal objects) are causally responsible for the distribution of salient locations on the saliency map. The map, in turn, causes salient actions like eye saccades, which support access to attention for salient stimuli. Since each of these plays a causal role in a stimulus's getting noticed by the faculty of attention, they each count as salient on the typical meaning.

---

[15] Ascription of salience to stimuli and their features is pervasive in the attention literature. Ascription of salience to feature representations and locations arises within the literature on computational models of bottom-up attention, e.g. Koch and Ullman (1987). Ascription of salience to actions arises within literature that conceives of attention as selection for action, e.g. Kerzel and Schönhammer (2013), including views on which bottom-up attention subserves a "belief optimizing" function of perception by determining the best locations to conduct "experiments" aimed at reduction of uncertainty, i.e. by determining locations of eye saccades (Parr & Friston, 2019)

Hence, the typical meaning of 'salience' is not very theoretically interesting. It picks out something like *being a causal antecedent of a bottom-up attentional process*. It does not describe any feature of attention's operation, only its inputs and their causes, and even there is limited to stimulus-driven inputs since top-down causal influences on attention are, by definition, excluded from the class of salient items.

The sense of salience to which I appeal applies to some of these causal antecedents. 'Salience', as I use it, applies only to representations. Feature representations and regions of saliency maps are representations, and they are salient when they have an impact on attention. Ultimately, however, it is exclusively the salience of feature-integrated object representations (henceforth "object-representations") that immediately explains quantitative character. An object-representation is a representation of an object bearing features that arises from feature binding, the binding of distinct feature representations into a single unit. Representations of an object's features, as on feature maps, explain *why* an object-representation is salient; an object-representation is endowed with salience accrued by these feature representations. However, it is object-representations that enter the attentional system with that accrued salience and command attentional processing. Salience, then, is the currency of the attentional economy. Early perceptual processes *earn* the salience that object-representations *spend* on attentional processing. Some object-representations are brought about by highly salient feature representations, and so are afforded a large allowance in the attentional marketplace, all else being equal. Others are brought about by less salient feature representations and are thereby less salient object-representations.

Quantitative character is a phenomenal reflection of an object-representation's salience. By this I mean that the salience accrued by an object-representation is traded for a correspondent amount of attentional processing; the amount of attentional processing an object-representation receives is fully determined by the object-representation's salience. Since there is no danger of the value of the goods coming apart from the currency, so to speak, there is no harm in using the term 'salience' to describe both the amount of attentional processing and the amount of currency used to afford that attentional processing; they are the same amount. To be clear, though, it is the attentional processing itself, and hence a functional state of the attention system, that realizes quantitative character.

There is a high degree of interpersonal and intrapersonal variability with respect to the salience afforded to exactly similar object-representations; two identical stimuli need not command the same amount of attentional processing across individuals or even within the same individual. The amount of salience afforded to an object-representation on the basis of its causal antecedents is determined by subjective and dynamic "market forces," as it were. For example, the salience afforded to the 'Susan' sound is subjective; it plausibly affords more attentional processing in the attentional economies of people named 'Susan' than it does in those with other names. Transient intrapersonal differences also affect the attentional economy, such as a subject's mood; the reason auditory objects sound louder to anxious subjects is that anxiety positively influences the rate-of-exchange for some auditory feature representations and the salience thereby afforded to the object-representation to which that feature representation contributes. Put another way, the same auditory stimulus sounds louder to an anxious subject because anxiety allows auditory objects to command more attentional processing.

To that point, it is worth noting that top-down processes may influence salience, and thereby influence quantitative character. Top-down factors may modulate the salience of object-representations by, e.g., altering salience-appraisal functions. For example, consider the cocktail party effect, in which one attends to a stream of speech amid competing speech streams. The attended stream receives preferential auditory processing; other streams are "tuned out" – not to the degree that we are completely unaware of them, of course, but at least to the degree that we are broadly unaware of their contents. As I am using 'salience', this can be described as a top-down influence on the relative salience of the attended stream compared to the unattended streams; selectively attending to the one stream increases the salience of representations of features of the selected stream and/or reduces the salience of features of other streams. Some distracting events – hearing one's name in a competing stream, for example – might register as sufficiently salient to briefly pull one's focus from the selected stream, for selection of the one stream does not make it so salient that nothing else can get through. Nonetheless, the modulation of salience introduces *enough* of a processing difference to accommodate comprehension of the selected stream.

The degree to which an object-representation is salient is reflected in the phenomenal representation of that object as quantitative character. The stream one focuses on has a dominating influence on one's phenomenal landscape; it is more phenomenally prominent than the other streams that muddle together in the background of the soundscape. It might seem odd to call this phenomenal difference a difference in loudness, but this is only because we so often use the term 'loudness' to mean 'distal intensity' and shifting focus to one person's speech stream does not make it seem like the person has begun to speak louder. If loudness is instead understood as a subjective feature of experience, what else could it describe other than the phenomenal prominence of an auditory object? If we impose a spatial metaphor on the auditory landscape, we might describe phenomenal prominence as the arrangement of auditory objects such that the most dominant one is "in front" of all the others, that the most prominent sound "sticks out." But the auditory landscape is not spatial, and as such phenomenal prominence must be due to some other phenomenally available aspect of auditory experience. I can think of no other aspect of the auditory experience that could underscore phenomenal prominence other than loudness; sounds certainly do not seem to increase in pitch or change in temporal structure as they increase in phenomenal prominence. If we genuinely *feel* phenomenal prominence in audition, then it must be via loudness. This point is perhaps clearer with respect to distracting stimuli. If she didn't know any better, Susan might think that the person she overhears saying 'Susan' at a party has intentionally said her name louder than all other words in her speech stream; it sounds so loud to her that she does not know how she managed to ignore the intruding speech stream in the first place.

5. Conclusion

The sort of salience that underlies quantitative character is compatible with any plausible empirical or computational model of attention. After all, quantitative character as I conceive of it is just a measure of the magnitude of the attentional processing response to an object-representation. The existence of such a construct requires only that attention processes object-representations, a minimal commitment that any model of attention is likely to satisfy.

In principle it could be empirically demonstrated that quantitative character fails to covary with attentional processing. However, given the relative infancy of research on attention and its neural implementation, the ability to determine how much attentional processing an object-representation receives with any degree of precision is far-off. Correspondingly, I can only gesture broadly at attentional processing as the functional correlate of quantitative character, for I do not know how to best type and quantify the relevant attentional processing. There are many open questions with respect to this account of quantitative character.

Nonetheless, it is difficult to imagine an analysis that makes better sense of the peculiar phenomenal characters that I call quantitative that does not appeal to salience. Quantitative characters are much like prothetic dimensions of perceptual experience, but differ in two significant ways. First, they modulate phenomenal intensity not of any particular feature, but of whole object representations, as evinced by the fact that their null values necessarily eliminate the objects to which they apply from the phenomenal landscape. Second, they seem intrinsically bound up with attention. Object-representations with relatively high values of quantitative character are presented in a manner that makes them easy to notice and difficult to ignore; engaging with them is irresistible. This is not the case for qualitative characters; even at objectively high values, no quality of a metathetic or prothetic dimension essentially drives attention toward it. A straightforward explanation for this pair of features is that quantitative character is an output not of perception directly, but of attentional processing of perceptual representations.

References

Arieh, Y., & Marks, L. E. (2011). Measurement of loudness, part II: Context effects. In M. Florentine, A. N. Popper, & R. R. Fay (Eds.), *Loudness* (pp. 57–87). Springer New York.

Baliki, M. N., Geha, P. Y., & Apkarian, A. V. (2009). Parsing pain perception between nociceptive representation and magnitude estimation. *Journal of Neurophysiology*, *101*(2), 875–887. https://doi.org/10.1152/jn.91100.2008

Burrows, B. E., & Moore, T. (2009). Influence and limitations of popout in the selection of salient visual stimuli by area V4 neurons. *Journal of Neuroscience*, *29*(48), 15169–15177. https://doi.org/10.1523/jneurosci.3710-09.2009

Glasberg, B. R., & Moore, B. C. (2002). A model of loudness applicable to time-varying sounds. *J Audio Eng. Soc.*, *50*(5), 331–342.

Huang, N., & Elhilali, M. (2017). Auditory salience using natural soundscapes. *The Journal of the Acoustical Society of America*, *141*(3), 2163–2176. https://doi.org/10.1121/1.4979055

Itti, L. (2005). Models of bottom-up attention and saliency. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 576–582). Elsevier Academic Press.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10-12), 1489–1506. https://doi.org/10.1016/s0042-6989(99)00163-7

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (Eds.). (2013). Low-level visual processing: The retina. In *Principles of Neural Science* (5th ed., pp. 1339–1388). McGraw-Hill Medical.

Kaya, E. M., & Elhilali, M. (2012). A temporal saliency map for modeling auditory attention. *2012 46th Annual Conference on Information Sciences and Systems (CISS)*. https://doi.org/10.1109/ciss.2012.6310945

Kayser, C., Petkov, C. I., Lippert, M., & Logothetis, N. K. (2005). Mechanisms for allocating auditory attention: An auditory saliency map. *Current Biology*, *15*(21), 1943–1947. https://doi.org/10.1016/j.cub.2005.09.040

Kerzel, D., & Schönhammer, J. (2013). Salient stimuli capture attention and action. *Attention, Perception, & Psychophysics*, *75*(8), 1633–1643. https://doi.org/10.3758/s13414-013-0512-3

Koch, C., & Ullman, S. (1987). Shifts in selective visual attention: Towards the underlying neural circuitry. *Matters of Intelligence*, 115–141. https://doi.org/10.1007/978-94-009-3833-5_5

Liao, H.-I., Kidani, S., Yoneya, M., Kashino, M., & Furukawa, S. (2015). Correspondences among pupillary dilation response, subjective salience of sounds, and loudness. *Psychonomic Bulletin & Review*, *23*(2), 412–425. https://doi.org/10.3758/s13423-015-0898-0

Liao, H.-I., Yoneya, M., Kidani, S., Kashino, M., & Furukawa, S. (2016). Human pupillary dilation response to deviant auditory stimuli: Effects of stimulus properties and voluntary attention. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00043

McDermott, J. H. (2013). Audition. In K. Ochnser & S. M. Kosslyn (Eds.), *The Oxford Handbook of Cognitive Neuroscience Volume 1: Core Topics*. Oxford University Press.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*(6), 826–859. https://doi.org/10.1037/0033-2909.125.6.826

O'Callaghan, C. (2007). *Sounds: A philosophical theory*. Oxford University Press.

Panek, W., & Stevens, S. S. (1966). Saturation of red: A prothetic continuum. *Perception & Psychophysics*, *1*(1), 59–66. https://doi.org/10.3758/bf03207823

Parr, T., & Friston, K. J. (2019). Attention or salience? *Current Opinion in Psychology*, *29*, 1–5. https://doi.org/10.1016/j.copsyc.2018.10.006

Pautz, A. (2015). The real trouble with phenomenal externalism: New empirical evidence for a brain-based theory of consciousness. In R. Brown (Ed.), *Consciousness inside and out: Phenomenology, neuroscience, and the nature of experience* (pp. 237–298). Springer.

Pratt, H. (2011). Sensory ERP Components. In E. S. Kappenman & S. J. Luck (Eds.), *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press.

Röhl, M., & Uppenkamp, S. (2012). Neural coding of sound intensity and loudness in the human auditory system. *Journal of the Association for Research in Otolaryngology*, *13*(3), 369–379. https://doi.org/10.1007/s10162-012-0315-6

Schmidt, F. H., Mauermann, M., & Kollmeier, B. (2020). Neural representation of loudness: Cortical evoked potentials in an induced loudness reduction experiment. *Trends in Hearing*, *24*, 1–13. https://doi.org/10.1177/2331216519900595

Siegel, E. H., & Stefanucci, J. K. (2011). A little bit louder now: Negative affect increases perceived loudness. *Emotion*, *11*(4), 1006–1011. https://doi.org/10.1037/a0024590

Stevens, S. S. (1957). On the Psychophysical Law. *Psychological Review*, *64*(3), 153–181. https://doi.org/10.1037/h0046162

Stevens, S. S. (1960). The psychophysics of sensory function. *American Scientist*, *48*(2), 226–253.

Suzuki, Y., & Takeshima, H. (2004). Equal-loudness-level contours for pure tones. *The Journal of the Acoustical Society of America*, *116*(2), 918–933. https://doi.org/10.1121/1.1763601

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 459–478. https://doi.org/10.1037/0096-1523.16.3.459

Wang, C.-A., & Munoz, D. P. (2014). Modulation of stimulus contrast on the human pupil orienting response. *European Journal of Neuroscience*, *40*(5), 2822–2832. https://doi.org/10.1111/ejn.12641

Wang, N., Kreft, H. A., & Oxenham, A. J. (2015). Loudness context effects in normal-hearing listeners and cochlear-implant users. *Journal of the Association for Research in Otolaryngology*, *16*(4), 535–545. https://doi.org/10.1007/s10162-015-0523-y

Wu, W. (2014). *Attention*. Routledge.

Zahorik, P., & Wightman, F. L. (2001). Loudness constancy with varying sound source distance. *Nature Neuroscience*, *4*(1), 78–83. https://doi.org/10.1038/82931