Received: 31 March 2024 | Revised: 24 June 2025 | Accepted: 16 July 2025

DOI: 10.1111/mila.70003

ORIGINAL ARTICLE

WILEY

Language comprehension and the rhythm of perception

Alfredo Vernazzani

Ruhr-Universität Bochum/Hanse-Wissenschaftskolleg, Delmenhorst, Germany

Correspondence
Alfredo Vernazzani, Hanse-Wissenschaftskolleg, Lehmkuhlenbusch 4, 27753 Delmenhorst, Germany.
Email: alfredo-vernazzani@protonmail.com

It is widely agreed that language understanding has a distinctive phenomenology, as illustrated by phenomenal contrast cases. Yet it remains unclear how to account for the perceptual phenomenology of language experience. I advance a rhythmic account, which explains this phenomenology in terms of changes in the rhythm of sensory capacities in both reading and speech perception. After presenting conceptual and empirical foundations for the account, I argue that it should be abductively preferred over competing views, especially the semantic perceptual view, which holds that we literally perceive linguistic meaning.

KEYWORDS

phenomenal contrast, phenomenology of perception, reading, rhythm, rich content view, speech perception

1 | INTRODUCTION

Consider the following phenomenal contrast case: Mary is being asked a question or reads a written sentence in Singhalese[1]:

> **Scenario 1**: Mary does *not* know Singhalese. She undergoes an experience *e* with phenomenal character P_1.

[1]The two scenarios may be framed in different terms, for example of possible worlds—these technicalities are irrelevant for my purpose. I assume an intra-subject comparison to avoid complications resulting from the controversy over the Frege-Schlick view (Shoemaker, 2006).

Scenario 2: Mary *does* know Singhalese. She undergoes an experience *e* with phenomenal character P_2.

Most philosophers today agree that this contrast highlights a genuine phenomenal difference between #1 and #2, both in the hearing (Block, 1995, p. 234; Prinz, 2006, p. 452; Siegel, 2006, p. 490; Strawson, 2010, pp. 5–6) and in the reading case (Peacocke, 1992, p. 89; Siegel, 2010, p. 100). What explains the phenomenal contrast? It seems plausible that Mary's language competence in #2 is the relevant difference-maker, but is the phenomenal difference best explained in perceptual or cognitive terms?

Recent philosophical work has increasingly favored a perceptual account of language comprehension (Fodor, 1983; McDowell, 1998; Millikan, 2016; Siegel, 2010).[2] Very roughly, the idea is that the perceptual experience of reading or hearing linguistic content—understood broadly to include words, propositions, utterances, and so on—in a language one understands differs phenomenally from the experience of languages one does not know. The claim is that perceptual experiences exhibit a distinctive what-is-it-likeness (Nagel, 1974), shaped by the subject's linguistic competence. I refer to this view as *linguistic perceptualism*.

Linguistic perceptualism comes in two versions. On the strong version, the phenomenal character of language perception is determined solely by perceptual factors; the phenomenology is *simple*. On the weak version, perceptual factors still shape phenomenal character, but cognitive or other non-perceptual factors may also contribute. In such cases, the phenomenology is *complex*. Linguistic perceptualism contrasts with *linguistic cognitivism*, which denies that anything distinctively perceptual contributes to the phenomenal character of *language comprehension*.[3] On this family of views, language comprehension and its phenomenology are entirely post-perceptual, resulting from inferences or cognitive processes based on or initiated by perceptual inputs (Jacob, 2015; O'Callaghan, 2011; Reiland, 2015; Stanley, 2005). Any phenomenal difference (if one exists) would thus be best explained by means of post-perceptual, cognitive phenomenology (for discussion, see Drożdżowicz, 2021).[4]

A further distinction is in order. It is widely accepted that the phenomenal character of a (phenomenally conscious) perceptual state is grounded, at least in part, in its content—that is, in what is perceived: properties, events, and objects (Chalmers, 2004; Kriegel, 2002). For present purposes, we can remain neutral about the nature of content, allowing it to be interpreted in representationalist or naïve realist terms. However, many philosophers also emphasize that *how* content is experienced affects phenomenal character. A clear example is the influence of attention on perceptual content (Block, 2010). This motivates a rough distinction. I refer to views that explain the phenomenal contrast solely in terms of perceptual content as

[2]Fodor (1983) noted that language comprehension and perceptual mechanisms are both triggered automatically by the right input like reflexes (pp. 52–53, 54–55), and process information very rapidly (p. 61) (Brogaard, 2020, p. 117). These features are characteristic of domain-specific systems and distinguish them from domain-general processes, which are typically slower and less automatic (Fodor, 1983, p. 55; see also Machery, 2009, p. 9).

[3]An anonymous reviewer notes that O'Callaghan (2011) allows for the detection of language-specific, non-semantic phonological features. Reiland (2015, p. 485), commenting on O'Callaghan, emphasizes the distinction between *phonological* and *semantic* competence: the former does not entail speech or text *comprehension*. This highlights an important point—linguistic cognitivism, as I define it here, concerns episodes of *language comprehension* and is therefore compatible with the detection of non-semantic, language-specific features.

[4]The nature of language perception bears also on epistemological questions such as the structure of perceptual justification (Brogaard, 2018; Gasparri & Murez, 2021; Jacob, 2015; Pettit, 2010). In this paper, I gloss over the epistemological aspects to focus on the phenomenological ones.

"what-strategies". In contrast, "how-strategies" focus on the way content is processed or experienced. An example of the former is the semantic perceptual view (SPV), which holds that we literally perceive linguistic meaning as high-level perceptual properties (or "meaning properties," Bayne, 2009; Brogaard, 2017, 2018, 2020; Nes, 2016; Pettit, 2010; Siegel, 2006, 2010; see Section 4). An example of the latter is Connolly's (2019) perceptual-learning account (Section 3.3).

I advance a novel perceptual account of the phenomenal contrast case: the *rhythmic account*. Roughly put, the idea is that possessing linguistic competencies alters the rhythm or speed at which sensory processes are mobilized during speech perception and reading. Since, as I will clarify, the rhythm or speed of the sensory processes at least partially shapes phenomenal character, the rhythmic account is compatible with both the weak and the strong versions of linguistic perceptualism. It therefore contrasts with linguistic cognitivism, which holds that the phenomenology of language comprehension is purely cognitive. Although I apply the rhythmic account to language perception here, I believe its core insights extend to other domains, deepening our understanding of perceptual experience more broadly.

This is how I proceed. In Section 2, I introduce the abductive structure of phenomenal contrast arguments and single out some criteria for assessing putative explanations of the phenomenal contrast. In Section 3, I introduce the rhythmic account. Then in Section 4, I turn to the what-strategies, with special focus on the SPV and argue that we should prefer the rhythmic account.

2 | PHENOMENAL CONTRAST ARGUMENTS

Phenomenal contrast arguments have an abductive structure (inference to the best explanation, IBE) that takes as its explanandum phenomenon the difference in phenomenal character between two experiences (Koksvik, 2015; Siegel, 2007). There are different ways to frame IBEs (Douven, 2025; Josephson & Josephson, 1994; Lipton, 2004; Walton, 2006, p. 167); for my purposes, the following simple schema will suffice:

(1) F is a finding or a fact.
(2) E_1, E_2, ... E_n may explain F.
(3) No other putative explanation explains F as well as E does.
(4) Therefore E is plausibly true.

(adapted from Josephson & Josephson, 1994, p. 14)

In our case, F is the phenomenological discrepancy highlighted by the phenomenal contrast case between #1 and #2. There are multiple candidate explanations E_1, E_2, ... E_n that may explain F, the goal is to identify the single *best* explanation among multiple candidate explanations.

Assuming linguistic perceptualism, the range of candidate explanations is constrained to options that have at least a perceptual component. This brings us back to the two strategies outlined earlier. The what-strategy seeks to identify the best explanation among content-based options, whereas the how-strategy seeks to identify the best how-based explanation. To adjudicate which explanation is best, we need criteria to determine when one "explanation is better than another" (Walton, 2013, p. 161). Two obvious criteria are explanatory scope and

parsimony. In our case, two further criteria may be plausibly added, a phenomenological one and an empirical one.

The *criterion of phenomenological plausibility* holds that any account purporting to explain or describe the phenomenal character of mental states should be phenomenologically plausible:

> **Phenomenological plausibility:** If E is phenomenologically more plausible than E', then we should prefer E over E'.

It is unclear how to spell out more precisely the notion of phenomenological plausibility. Nonetheless, philosophers of perception frequently invoke phenomenological plausibility when evaluating the theoretical virtues of accounts of perceptual experience.[5]

The second is a *criterion of empirical plausibility*. It is reasonable to require that any proposed explanation not be ruled out or directly contradicted by current scientific understanding. Moreover, we should favor a candidate explanation that is not only logically compatible with the empirical evidence but, where possible, supported by it. This brings us to:

> **Empirical plausibility:** If E is empirically more plausible than E', we should prefer E over E'; this means:
> (i) E should not be logically incompatible with our best and most recent empirical data or theories T;
> (ii) There must be comparatively some empirical evidence that favors E over E'.

A few final clarifications. First, the two criteria are not excessively restrictive. It would arguably be too demanding to frame empirical plausibility in terms of empirical truth or maximal evidence. While (perhaps) in a few cases a philosophical account may be neatly and decisively confirmed empirically, more often than not empirical data and theories—if and when they are relevant—only provide scant support for philosophical accounts.

Second, it may be argued that measuring the success of philosophical speculations on empirical grounds is unfair. But notice that the criterion of empirical plausibility merely states that if E is empirically more plausible than E', then we should prefer E; it does not, by itself, establish the truth of E, nor does it imply that there always is empirical data or theories that are relevant to assess E. Finally, one may dispute the notion of plausibility used here. I follow Josephson and Josephson (1994, pp. 266–272) and assume an intuitive notion of plausibility. Further refinements of this notion (e.g., in terms of probability distributions) are not required for my argument.

[5]Consider an example. Proponents of gappy Russellian contents analyze hallucinations as ordered pairs of a gap and a property, $< \_\_, P >$ in contrast to genuine perceptions where the gap is filled by an object, $< o, P >$ (Tye, 2007). This implies that hallucinations involve the perception of uninstantiated universals (Johnston, 2004). But as Schellenberg (2010) notes, this is phenomenologically implausible since "it is not clear what it would be to be sensorily aware of an uninstantiated property" (pp. 33–34). Her point can be strengthened by two related concerns. First, uninstantiated universals lack spatiotemporal location, yet perceptual content—even in hallucination—typically presents things as located in space and time (e.g., Macbeth's "dagger of the mind" appears before him and he tries to grasp it; Smith, 2002, chap. 5). Second, it is unclear how uninstantiated universals could ground phenomenal character: the universal red is not itself red. I thank an anonymous reviewer for encouraging me to clarify this point.

3 | THE RHYTHMIC ACCOUNT

3.1 | The dependency argument

Let us begin with a simple observation: perceptual experiences have a diachronic dimension (James, 1890/2019; Noë, 2012; O'Shaughnessy, 2000; Prosser, 2016; Sattig, 2019; Soteriou, 2013). Reading a text or hearing someone speak are experiences that unfold over time—they have a beginning and an end. Within this temporal span, various sensory capacities or processes are automatically engaged. These include, among others, oculomotor movements, motion and color detection capacities, and sensorimotor contingencies. The rhythmic account builds on the widely accepted idea that our perceptual openness to the world depends on these processes and how they unfold over time. I take a liberal view regarding the nature of such processes and how to type them; what matters for my purposes is that they can temporally unfold in different patterns, as I argue below.

The argument is divided into two parts. In the first part, I connect the unfolding of sensory capacities to an experience's phenomenal character. I call this the *dependency argument*. In the second part (Section 3.2), I examine how sensory capacities unfold and substantiate with empirical studies my claim that linguistic competence modifies their rhythm and speed. Turning to the first argument, it runs as follows:

Dependency argument

I. *Dependency thesis*: The phenomenal character Φ of a perceptual experience e necessarily depends (at least in part) on the way W in which the relevant set of sensory processes unfold.
II. If the phenomenal character Φ of a perceptual experience e necessarily depends (at least in part) on the way W in which a set of relevant sensory processes unfold, then some changes in W elicit a change in Φ.
III. Some changes in W elicit a change in Φ.

I take the *dependency thesis* (I) to be a truism. Everyone agrees that the phenomenal character of experience necessarily depends, at least in part, on the unfolding of certain sensory processes. These include, but are not limited to, auditory processes and oculomotor movements. Disagreement arises not over this basic dependency, but over how to interpret the role of these processes in perceptual experience—whether they "reveal" a tract of the environment (e.g., Campbell, 2002), or represent it (e.g., Burge, 2010). The *dependency thesis* is compatible with various views about the individuation of experiences. Since it is widely accepted that an experience's phenomenal character partially depends on what is experienced (the content) and how it is presented, it follows that the structure and dynamics of sensory processing partly ground phenomenal character (Section 4.3). The *dependency thesis* only asserts a necessary, not a sufficient, condition for phenomenal character. Phenomenal internalists maintain that an experience's phenomenal character depends solely on internal sensory processes (and perhaps other internal factors) (Burge, 2010; Byrne, 2002; Price, 1932; Tye, 1995). Externalists, by contrast, hold that it also depends on worldly particulars (Campbell, 2002; Martin, 2002; McDowell, 1984; Noë, 2012; see Schellenberg, 2018 for a "mixed" view).

The *dependency thesis* establishes a connection between two levels of properties: a more fundamental level of properties of sensory processes, that is, the way sensory processes unfold

(more on this below and Section 3.2), and the higher-level property of a perceptual experience, that is, its phenomenal character.[6] It is not the mere presence of sensory processes that engenders (at least in part) the phenomenal character of one's experience, but the way they unfold. Arguably, not all sensory processes will affect one's phenomenal character; the subset of *relevant* sensory processes is that whose modulation governs our access to the world and processing of related sensory information. When framed in these terms, the second premise (II) represents an instance of a more general principle:

> **Basis**: If a property P of α necessarily depends on a (set of) property(-ies) P' of β, then some changes in P' will affect P.

Basis bears a close resemblance to Shoemaker's microrealization, which is a "many-one relation between properties instantiated by different individuals" (Bennett, 2011, p. 83). Think of the mass of a table as microbased on the masses of its constituent parts, or the wetness of a quantity of water as microbased in the properties of hydrogen and oxygen atoms. Of course, this leaves open a number of metaphysical questions, such as whether the higher-level property is "over and above" the low-level ones, and so on. I gloss over these issues here as they are not relevant for the present project.[7]

The argument rests on the fact that sensory processes may unfold in different ways W. It remains to clarify in what ways they unfold. Processes that unfold in time will do so with distinctive rhythms and speeds. Determining how sensory processes unfold, however, must be uncovered empirically. I turn to this issue in the next section.

3.2 | The rhythm of perception

I have argued that sensory processes may unfold in time in different ways. In order to illustrate this point, I will use music notation as a model. For simplicity's sake, I shall assume that the two points in time of onset (t_1) and end (t_2) delimit a single meter, that the time is a four-four (common time), and a single voice standing for a single sensory process, SP. Not all features of music notation are relevant in the present context (e.g., we can ignore pitch), just as not every feature of a model affords epistemic access to the target phenomenon (Elgin, 2017).

Sensory processes may unfold in a homogeneous way throughout the entire experience. Call this the *simple model (Figure 1)*.

[6]These two levels do *not* correspond to low- and high-level properties of the rich content view (Siegel, 2010). The relevant levels here are the physiological and the experiential ones.

[7]Broadly speaking, the metaphysical relation I am concerned with is a "building relation" (Bennett, 2011). Crucially, the fact that sensory processes have a distinctive rhythm and speed does not entail that we are phenomenally aware *of* rhythm and speed (Hoerl, 2013). As Drożdżowicz (2021) notes, phonetic and phonological computations occur beneath the level of conscious awareness. I do not claim that we are aware of changes in perceptual rhythm; it is enough that such changes affect the overall phenomenal character of experience. As I clarify in the argument, not all base-level changes impact higher-level phenomenology. My claim is that, all else being equal, changes in rhythm and speed (W) produce changes in phenomenology. In Section 3.2, I present extensive empirical evidence for a strong correlation between rhythm of perceptual capacities and the experience of language comprehension. I thank an anonymous reviewer for prompting this clarification.

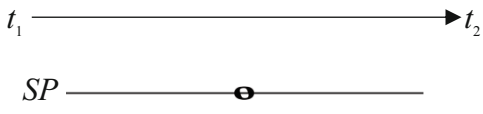

FIGURE 1 The simple model.

On the simple model, *SP* has a single value—represented by a whole note—meaning it is exercised evenly throughout the entire episode. (Here, I use "value" in the musical sense, i.e., referring to note duration.) This model reflects a baseline or grade zero of *rhythm*. The concept of rhythm admits various definitions (Cowell, 1930/1996, p. 45); I adopt a broad notion of rhythm as a pattern of durations unfolding at a certain *tempo* (London, 2012).[8] Roughly, tempo refers to the pace or speed at which a piece is performed or a process unfolds (Cowell, 1930/1996, pp. 90–98).[9] One might add further information about the overall speed and duration of the process—it may last 2 s or 10 min. More importantly, however, sensory processes can unfold in more complex patterns, with alternating peaks and rests of varying durations. Call this the rhythmic model (Figures 2 and 3). As before, the entire episode may unfold at different tempos. But unlike the simple model, the rhythmic model features a more complex internal rhythmic structure. According to this model, the sensory process unfolds in sequences of varying durations (and, again, tempos). Note that the model admits multiple possible articulations—just as notes in a four-four time signature can have different durations. I refer to these different variations as *rhythmic profiles*. Here is an example of an alternative rhythmic profile. Let me stress again an important point. The two models that I have just sketched out are meant to capture the rhythmic unfolding of the sensory processes themselves, *not* of the phenomena they are about or exercised on. The models are simplified idealizations, since obviously in each experience multiple processes will unfold. This may be represented as multiple voices, each with its own rhythmic structure. To stay with the musical metaphor, the deployment of sensory processes looks more like a polyphony than a monophony.

If sensory processes unfold more like the simple model—rather than the rhythmic model—we would have a fairly rigid margin of flexibility in sensory processes. The reason is that according to the simple model sensory processes are always exercised in the same uniform way, although the model allows for changes in speed. The rhythmic model allows for a greater diversity in terms of different rhythmic profiles.

I will now argue for the following claims:

(i) The rhythmic model exemplifies the way sensory processes unfold.
(ii) Understanding the target language modifies the rhythmic profile of those processes.

[8]This definition of rhythm differs from Simons's "repeatable patterns in sound and silence" (2019, p. 70), which ties the notion of rhythm to sounds only. Rhythm is a widespread phenomenon that occurs wherever there is a sequence of events or processes of different durations (Bolton, 1894).

[9]There is more to tempo in music than mere speed, as it usually encodes information about expressivity as well. I will ignore this feature here.

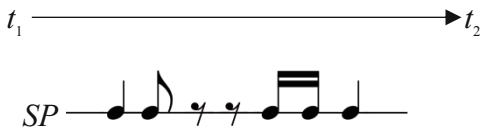

FIGURE 2 The rhythmic model, rhythmic profile 1.

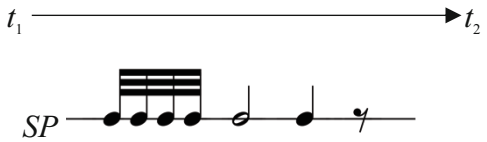

FIGURE 3 The rhythmic model, rhythmic profile 2.

If understanding a language modifies the rhythmic profile of sensory processes (our *W* in the *dependency argument*), then it elicits a change in phenomenal character *Φ*. Yet, it is an empirical matter to establish (i–ii) in cases of language perception. I focus on studies on oculomotor movements in reading experiences (Section 3.2.1), and of rhythmic entrainment in speech perception (Section 3.2.2).

3.2.1 | Oculomotor movements in reading

In ordinary cases of reading, our eyes do not scan a text in a linear fashion moving from left to right (or in other directions, depending on the language system).[10] It is well-known that our eyes follow a fairly general pattern of oculomotor movements. This pattern consists of forward movements, word fixations and refixations on some words (within-word refixation), and movements back to previous words (inter-word regression). Nearly two-thirds of the words are fixated, and only a few of the skipped words are re-fixated after inter-word regression. According to Vitu (2011, p. 736), inter-word regression amounts to circa 21% of eye movements during reading in the case of adult readers, whereas in first grade children this amounts to 34% (McConkie et al., 1991). According to a more recent estimate by Schotter and Rayner (2015), the regressive saccade amounts to circa 10%–15% of the total saccades. We thus get an average of 4–5 saccades per second and 1 regressive saccade every 2 s. The fixations consist of sequences of 150–500 ms, followed by forward saccades. The forward movements (for English readers) are on average 7–9 letters, whereas the average saccade duration is between 20 and 35 ms. This process is regulated by different factors, among them, word length and fixation location. This is illustrated by the fact that the likelihood of

[10]By ordinary cases of reading I mean episodes where subjects are genuinely engaged with the text, not merely skimming or "mindless reading." I also assume that ideal subjects do not have reading impairments (e.g., dyslexia).

word skipping decreases with word length. For 1–2-letter words, the chance of skipping it is close to 0.76, whereas for 4-letter words it approximates to 0.42, and sinks to 0.05 for 9–10 letter words (Vitu et al., 1995).[11]

These studies show that oculomotor movements unfold in a rhythmic pattern closer to the rhythmic model than the simple model (i). According to the simple model, the eyes would scan text in a continuous, uniform flow from beginning to end, without significant variation in duration or pauses. But this is not how our eye movements actually behave in the reading experience. Instead, we observe here a complex pattern of durations where the saccades function as rests and fixations vary in duration ("values," in the rhythmic model). Having established (i), I now turn to (ii).

Word length and font are objective factors influencing the rhythm and speed of eye movements, but there is more. As we have seen, young children with limited understanding of their native tongue make significantly more inter-word regression saccades than adults. Furthermore, research over the past 30 years has unequivocally shown that the fixation time on a word and saccadic length also depend on the reader's language competence (Rayner & Liversedge, 2011). More specifically, word *frequency* (Rayner & Duffy, 1986) and *predictability* (Kuperberg & Jaeger, 2016; Rayner & Well, 1996) play a major role in determining saccadic length and fixation times:

> [A]s the *text becomes more difficult* ... *saccade length decreases* ... how long readers look at a word is strongly influenced by factors like *word frequency* and *word predictability* ... how easy a word is to process has a large impact on how long the *eyes remain* on that word (Schotter & Rayner, 2015, my emphases).

Word frequency depends on the target language, and their predictability is a function of a subject's competence in the target language and text understanding. The better someone knows the target language, the better she will be able to anticipate word patterns depending on the text's meaning in context. As Schotter and Rayner emphasize, the ease of processability of a word also has an impact on fixation length. Thus, words that are difficult to process increase the overall duration of the fixation, which, as we have seen, may last between 150 and 500 ms. (The average reading speed is 200–300 words per minute for competent English native speakers, depending, among other things, on their level of education.) Note that processing difficulty is, again, partly a function of the subject's lexical competence.[12]

The relative difficulty of a text is also a determining factor for the timing of eye movements and the speed of the overall reading experience (Rayner & Liversedge, 2011). I say "relative difficulty" because the difficulty of a text may subjectively vary depending on multiple dimensions:

[11]Fonts also influence reading, as shown by Song and Schwarz (2008): Texts in an easy-to-read Arial font are read more quickly and feel easier to process than identical texts in a harder-to-read Brush font.

[12]Processing difficulty may also stem from conditions like dyslexia, which is known to affect reading speed (Ashby et al., 2005). While my focus is on oculomotor movements—often used as proxies for reading comprehension—reading involves a hierarchical process spanning multiple levels, from letter recognition to orthographic processing and lexical access (Kemmerer, 2023, chap. 15). A key component is the visual word form area (VWFA), which responds primarily (though not exclusively) to written words. In a fMRI study conducted of 63 Portuguese and Brazilian adults—some literate, some ex-illiterate (i.e., they learned to read in adulthood), and some illiterate—Dehaene et al. (2010) found that VWFA activation increased with reading performance, measured in words per minute. This suggests that the rhythmic profile of reading may recruit broader sensory processes. Still, the precise neural and computational architecture of reading remains an area for further research. I thank an anonymous reviewer for encouraging me to consider additional processing levels.

the domain-specificity of the text and the reader's competencies in that domain, general textual complexity (arguably, reading Joyce's *Finnegans Wake* will be a challenge for all readers, even though skilled readers will be better able to navigate it), and the degree of language proficiency.[13]

Understanding the target text thus affects the reading speed and rhythm of oculomotor movements, altering the duration of fixations and the saccadic movements (Rayner et al., 2006). In short, understanding a given language alters the rhythmic profile of oculomotor movements when reading a text in that language (ii). (This partly vindicates Wittgenstein's suggestion of a link between the speed of reading and understanding, 1977/2006, MS 134, 76, p. 65e.)

3.2.2 | Rhythmic entrainment in speech perception

It is well-known that the flow of speech is rhythmic (Bolton, 1894; Haegens & Golumbic, 2017; Langus et al., 2017; Poeppel & Assaneo, 2020). Rhythms in speech have a hierarchical structure. The lowest level is the segmental one, which is marked by the alternations between consonants and vowels; at the supra-segmental level rhythm is marked by the alternations of stressed elements (Langus et al., 2017). The waveform of speech signals—commonly called the "speech envelope"—reveals a sequence of increments and decrements that is thought to play a role in language comprehension (Poeppel & Assaneo, 2020, p. 323). Before examining the role of sensory processes in decoding the speech signal, however, we must briefly turn to the diversity of languages' rhythmic structures.

It is of course a familiar experience that different languages sound in different ways. This is due to a language's phonemes, its phonotactics (the phonological segments and sequences that are legal or illegal in a given language; Vitevitch & Aljasser, 2021), and its prosody. There are different ways of classifying the rhythm of languages (Haegens & Golumbic, 2017). A classical approach is to identify three rhythmic classes: stress-timed (e.g., English, Dutch), syllable-timed (e.g., Italian, Spanish), and mora-timed (e.g., Japanese). Languages belonging to the same class will "sound" more similar to each other. Ramus et al. (1999) have proposed a way of measuring languages' rhythm at the most basic segmental level in terms of vocalic space $\%V$ in the speech stream, and the deviation of consonantal intervals ΔC. This enables us to plot languages' different rhythmic structures—for instance, Japanese has a very small syllabic inventory with only three syllable types, which means it is a language with high $\%V$ and low ΔC; while languages like English have a very rich syllabic repertoire (Langus et al., 2017). Having clarified that spoken languages have different rhythms, we must now return to sensory processes in the percipient.[14]

Sensory areas involved in acoustic perception are subject to intrinsic brain oscillations that reflect rhythmic fluctuations in neural ensembles excitability—alternating between high and low states (Haegens & Golumbic, 2017). EEG studies show that oscillation frequencies vary by sense modality, but in the auditory domain, they typically fluctuate between delta (<4 Hz) and theta (4–7 Hz) bands. These fluctuations are known to play a key role in the rhythmic sampling

[13]Another interesting case are garden-path sentences (Bever, 1970/2013): grammatically correct sentences but prone to misinterpretation on first reading. For example, "The old man the boat," initially suggests "old" is an adjective modifying "man," though the correct reading takes "old" as a noun and "man" as a verb (MacDonald & Hsiao, 2018; Section 4.2). As Rayner et al. (2006) note, such sentences elicit longer fixation times due to their odd structure.

[14] There are other ways of measuring a language's rhythmic structure at the segmental level (for further references, see Langus et al., 2017). Although two languages may *sound* very similar, I am not aware of any two languages that have exactly the same rhythm. Note that rhythm can also vary within a single language, for example across dialects.

of stimuli, facilitating sensory processing, target discrimination, segmentation, and tracking (Haegens & Golumbic, 2017; see also Kemmerer, 2023, chap 5; Meyer, 2018; Rimmele et al., 2018). The fact that auditory sensory processes unfold within the delta and theta bands reinforces the claim that the rhythmic model best captures the dynamics of sensory processing (i): neural oscillations are neither a-rhythmic nor constant, as the simple model would suggest. Languages exhibit distinctive rhythms, and sensory processes unfold rhythmically. What remains to be seen is whether, and how, knowing the target language *L* in a given perceptual episode *e* modifies the rhythmic profile of auditory sensory processing.

The key to explain what occurs in episodes of language comprehension lies in the phenomenon of *rhythmic entrainment*—the synchronization of the neural oscillations with the rhythm of the input (Kemmerer, 2023, pp. 126–128; London, 2012). This mechanism is active in both music and language perception. A growing body of empirical research highlights the importance of rhythmic entrainment across various sensory domains and its central role in sensory processing. In language perception specifically, the entrainment of low-frequency auditory brain activity in the temporal lobe to the rhythmic structure of the speech envelope is known to be essential for speech segmentation (Poeppel & Assaneo, 2020). More precisely, rhythmic entrainment operates across different levels of phonological processing. Studies have shown that subphonemic features and often entire phonemes are associated with gamma-band oscillations (>30 Hz); syllables align with theta-band (4–8 Hz) activity; and intonation phrases with delta-band (0.5–4 Hz) oscillations (Kemmerer, 2023, p. 126; on the role of phonological processes, see also Drożdżowicz, 2021). As Kemmerer notes, rhythmic entrainment or synchronization of neural oscillations to a speech stream "significantly enhances speech perception," and the extent to which a listener's brain oscillations are aligned with the speech stream "predicts the degree to which that listener perceives that speech as being intelligible" (2023, p. 128).[15]

A substantial body of empirical research demonstrates the top-down modulation of rhythmic entrainment in language perception and underscores its role in speech segmentation (Giraud & Poeppel, 2012). Park et al. (2015) showed that the entrainment of low-frequency oscillations in auditory areas to the rhythmic components of speech increases with intelligibility (see also Gross et al., 2013; Peelle et al., 2013). In other words, speech that is less intelligible—or entirely unintelligible—exhibits weaker synchronization with neural oscillations. This reduced synchronization is observed, for example, during backward speech perception and in the perception of unfamiliar foreign languages (Poeppel & Assaneo, 2020, p. 326). Further evidence comes from a study by Riecke et al. (2018), who devised a cocktail-party-like scenario to explore the role of rhythmic entrainment in speech comprehension. The findings suggest that rhythmic entrainment is not merely an epiphenomenon of intelligibility but reflects the role of language knowledge in modulating entrainment to the speech signal.

Knowing a language thus affects the rhythmic profile of sensory processes, significantly altering their rhythmic unfolding toward a better entrainment with the speech signal (ii). The fact that this mechanism is very close to the mechanism of rhythmic entrainment in episodes of music perception lends itself to the suggestion of a very close relation between music perception and speech perception mechanisms. This partially vindicates Wittgenstein's conjecture that understanding a sentence in a language is much closer to understanding a music theme than one may suspect (1958/2009, §527).

[15]The different hierarchical levels of phonological processing must then be coordinated through a process called "nesting" (Kemmerer, 2023, p. 128; also Arnal et al., 2016). I thank an anonymous reviewer for pointing out to me further literature on phonological processing.

3.3 | Assembling the rhythmic account

I can now assemble the rhythmic account:

Rhythmic account

I. If a subject knows the target language *L* that is the object of an experience *e*, then that knowledge changes the rhythmic profile of the relevant sensory processes upon which *e*'s phenomenal character *Φ* necessarily depends.
II. If the rhythmic profile changes, then there is a change in *e*'s phenomenal character *Φ*.
III. If a subject knows the target language *L* that is object of an experience *e*, then there is a change in *e*'s phenomenal character *Φ*.

The rhythmic profile refers to the rhythm and speed of sensory processes—this is an instance of the ways *W* in which sensory processes unfold in the *dependency argument* (Section 4.1). Premise (I) is justified in light of the empirical evidence reviewed in the foregoing section (Section 4.2). The second premise (II) has been argued for in the *dependency argument*. Since knowing the language that is presently perceptually experienced modifies the rhythmic profile of sensory processes, it follows that knowing the language that is the object of a perceptual episode changes the experience's phenomenal character.[16] This explains the phenomenal contrast case in both the reading experience and the speech perception. In the reading case, if Mary knows Singhalese (#2), her oculomotor movements will unfold faster and with a rhythm different from #1. In the speech perception case, if Mary knows Singhalese (#2), then the neural oscillations that undergird her acoustic experience will be better entrained to the rhythm of the acoustic signal. These effects reflect at the perceptual and sensory levels the higher degree of fluency that Mary displays when perceiving a language she is familiar with.

Some concluding remarks. First, the rhythmic account is fully compatible with *both* weak and strong versions of linguistic perceptualism. On the strong view, the rhythmic account offers the sole perceptual explanation for the phenomenal contrast case. On the weak view, it supports the intuition that something perceptual (at least partially) contributes to phenomenal character, while allowing room for additional cognitive factors—such as epistemic feelings or cognitive phenomenology—to shape the overall phenomenal character.[17] By contrast, the rhythmic account is incompatible with linguistic cognitivism, which denies any perceptual or sensory role in the phenomenal character of language perception.

[16]To the extent that knowing a language is a high-level cognitive achievement, one might interpret the rhythmic account as an instance of cognitive penetration (Vetter & Newen, 2014; Zeimbekis & Raftopoulos, 2015) and perhaps also a case of perceptual learning (Connolly, 2019; on the relation between cognitive penetration and perceptual learning, Burnston, 2021). If this is correct, one may reasonably expect the rhythmic account to have significant epistemic import in our linguistic experiences (Jenkin, 2022; Siegel, 2013). I leave this as an open avenue for further research.

[17]It seems unlikely that the overall phenomenal character of language experience is determined exclusively by perceptual factors. Reading, after all, is not a purely perceptual task. As a prima facie case against strong linguistic perceptualism, consider the following. In his *The great cat massacre* (1984), historian Robert Darnton argues that reading experience has changed significantly across centuries and cultures. In "Readers respond to Rousseau" (pp. 215–256), he presents two examples of reading in what he calls "alien *mentalités*". The first is a Balinese practice in which reading serves to protect the soul of the deceased from demonic possession. The second concerns the practice of extensive, as opposed to intensive, reading in 18th century France. Both cases plausibly involve distinct phenomenal characters, shaped by cultural inflections in the act of reading.

Second, I believe the rhythmic account offers significant advantages over other how-accounts. O'Callaghan (2010, 2011, 2015) argues that language knowledge enables subjects to segment the speech into words—an account updated by Connolly (2019, pp. 164–170) and framed in terms of perceptual learning. The idea is that knowing a language allows one to automatically parse speech by deploying capacities for stimulus differentiation and unitization. In response, there are three points in favor of the rhythmic account. First, O'Callaghan's and Connolly's view has a narrower explanatory scope. It targets only certain cases of speech perception and fails to account for many reading cases. This leads to a second, twofold point. Parsing speech is not always difficult even when we do not understand the language. Anglophone philosophers frequently cite the example of hearing someone speaking in French when one does not know it (Block, 1995, p. 234). English and French are phonologically and prosodically very different languages. But consider Italian and Spanish, which are much closer in rhythm than French and English. An Italian unfamiliar with Spanish might still perceive the speech as segmented into meaningful units, despite not understanding it.[18] Similarly, in reading one might struggle to detect semantic units in Chinese if one does not know the language, yet easily identify single words in alphabetic languages like Dutch or Hungarian, without understanding them. Third, the rhythmic account subsumes such cases of differentiation and unitization, offering a more fundamental explanation in terms of changes of rhythmic profiles.

A final proposal worth mentioning is Drożdżowicz's (2021) account in terms of epistemic feeling. Her accounts is specifically meant to capture the "feeling of understanding" that accompanies experiences of language perception. Such epistemic feelings, in her view, are metacognitive states generated by monitoring lower-level processes related to speech-perception. She maintains that the phenomenology of language comprehension is both genuinely perceptual *and* genuinely cognitive. This places her account squarely into what I have called weak linguistic perceptualism. I have no qualms against epistemic feelings. The rhythmic account is silent regarding any further cognitive inflections on the experience's phenomenal character. In this regard, our accounts are complementary.[19]

4 | THE WHAT STRATEGY

I have shown that the rhythmic account should be preferred over other how-account alternatives. But abductive arguments are inherently comparative, and it remains to be seen whether the rhythmic account should be preferred over the what-strategy. These accounts of the phenomenal contrast focus on content as the key explanatory ingredient. This brings us to the question of the reach of phenomenally conscious content.

4.1 | Thin perceptual content

Two options are available in the literature. The first option is the "thin" account of perceptual content (Tye, 1995). On this view, only low-level properties—such as colors and forms in visual

[18]This point is somewhat similar to Reiland's (2015, p. 485) distinction between phonological and semantic competence (see footnote 3).

[19]A referee has pointed out that cognitive processes, too, may unfold with distinctive rhythms. I find this suggestion intriguing, though it lies outside the scope of the present study.

experience, or pitch in acoustic experience—feature in perceptual content. Virtually all researchers agree that we perceive low-level properties; this claim is both phenomenologically and empirically accepted.

The problem with low-level accounts is that they lack the resources to explain our phenomenal contrast case. The reason is straightforward. In both scenarios (#1 and #2), Mary perceives the very same low-level acoustic or visual properties. Hence, it is not clear why the phenomenal character of #2 should differ from that of #1. If one accepts that there is indeed a phenomenal contrast that needs to be explained, one way out of this predicament for friends of the low-level account who embrace the what strategy is to abandon linguistic perceptualism. In order to do that, one may embrace cognitive phenomenology (Chudnoff, 2015) or theories of epistemic feelings (Drożdżowicz, 2021; Section 3.3). On these options, while perceptual content remains the same, it is some cognitive going-on that explains the phenomenal contrast case.

I do not have any in principle objection against cognitive phenomenology or the epistemic feeling account. Indeed, this option is clearly contemplated by weak linguistic perceptualism. The problem is that a what-strategy based on a thin account of perceptual content per se does not provide any perceptual explanation of the phenomenal contrast case, and is thus outside the scope of linguistic perceptualism. At this juncture, there are two ways out. One may either combine a thin account with my rhythmic account—thus obtaining a neat perceptual explanation of the phenomenal contrast case, either alone (strong linguistic perceptualism) or in combination with some further cognitive effects (weak linguistic perceptualism)—or resort to some version of the rich account of perceptual experience. It is to this alternative that I now turn.

4.2 | The SPV

The second option is to defend a rich account of perceptual content. On this view, we do not merely perceive low-level properties, but also high-level properties, such as natural kind properties (Siegel, 2010) or emotion and expressive properties (Newen, 2016). Friends of the rich content view usually set up an abductive argument to show that phenomenal contrast cases are best explained via high-level properties, precisely because low-level accounts lack the perceptual resources to explain such cases.

Over the last years, an interesting version of the rich content view has gained some traction, the semantic perceptual view (SPV) (Bayne, 2009; Brogaard, 2017, 2018, 2020; Nes, 2016; Pettit, 2010; Siegel, 2006, 2010). This view has it that our language competences literally enable us to perceive (linguistic) meaning. Here are Brogaard, Pettit, and Siegel respectively on the SPV:

> Fluent speakers of a language have a non-inferential capacity to auditorily (or otherwise) perceive not just the sounds of speech but also what was said or conveyed by the speaker (Brogaard, 2017, p. 144).

> According to the perceptual view, normal auditory perception of speech normally comes to us interpreted for content of the speech. We have the capacity to hear the content of speech—to hear it *as* having a particular content (Pettit, 2010, p. 22).

> once you can easily read it [a Cyrillic text], it takes a special effort to attend to the shapes of the script separately from its semantic properties. You become disposed

> to attend to the semantic properties of the words in the text, and less disposed to attend to the orthographic ones (Siegel, 2010, p. 100).

A few clarifications are in order. First, the SPV is modality neutral. Although much of the literature focuses on speech perception—and hence hearing—the SPV is meant to apply to reading cases as well, hence also visual (or perhaps tactile) experiences. For this reason, I will drop any modal qualification, and simply phrase the SPV in terms of meaning perception.

Second, it is not clear what is exactly meant with "meaning". This raises two sets of issues. The first set concerns the very nature and structure of meaning. There is no consensus about the nature of meaning. It is clear that the problem of the nature of meaning bears on the intelligibility of the SPV and whether meaning is the kind of entity that *can* be perceived. I turn to this issue below (Section 4.3). The second set of issues is the relevant level of linguistic meaning. Is the SPV a claim about detecting the meaning of words, sentences, utterances, propositions, discourses, or what else? Comparing Pettit's and Brogaard's quotes with Siegel's, the former two focus on speech—such as discourses, utterances, or sentences—while Siegel focuses on words. Moreover, as we will see in Section 4.3, Brogaard cites purported evidence supporting the SPV at both the word and the sentence levels. For present purposes, I will remain neutral on the issue of which level of meaning is primary, and assume that the SPV is intended to apply across all levels of linguistic meaning.

The SPV *does* have the perceptual resources to account for phenomenal contrast cases. It is Mary's linguistic competences that enable her to perceive high-level meaning properties. This generates a difference in what Mary experiences in #1 and #2. The SPV is fully compatible with both weak and strong versions of linguistic perceptualism. The SPV thus seems a real contender against the rhythmic account. Relying on the criteria outlined above (Section 2), I will now argue why one should prefer the rhythmic account over the SPV.[20]

4.3 | Troubles with the SPV

I now examine the SPV in light of the criteria of phenomenological plausibility (Section 4.3.1) and of empirical plausibility (Section 4.3.2), and argue that it has significant shortcomings.

4.3.1 | The phenomenological plausibility criterion

The SPV should not be conflated with the claim that we perceive language signs as meaningful. It is clear that one may see a string of Cyrillic characters or hear a question in Singhalese and judge—or perhaps even perceive—them as meaningful, despite having no understanding of what they mean. I cannot speak Chinese, but if I accidentally listen to someone speaking in Chinese, I would plausibly judge (or, perhaps, perceive) their speech as meaningful in spite of

[20]It may be rightly noted that the SPV and the rhythmic account are logically compatible. In principle, there is no reason we could not perceive high-level properties with distinctive rhythm and tempo. I address this objection below (Section 4.4). While proponents of the SPV tend to explain the phenomenal contrast case in purely perceptual terms, the SPV is compatible with weak linguistic perceptualism.

my inability to tell what they are actually saying.[21] The SPV however makes a stronger claim: that particular content or meaning is perceptually presented to us (Pettit, 2010, p. 22). But the very idea that meaning—whether of a word, an utterance, and so on—can be perceived strikes me as deeply unclear. The problem is compounded by the fact that proponents of the SPV offer little clarification about how their main claim should be understood. Yet making sense of the SPV's central claim is, I believe, essential for evaluating the view's phenomenological plausibility—and its overall tenability. I thus strongly disagree with O'Callaghan's suggestion that it "perhaps sets the bar too high to ask for an informative descriptive answer" to the question of what it would mean to perceive meaning (2011, p. 796).

A plausible way to determine whether meaning is the kind of thing that can be perceptually presented is to inquire further into the nature of meaning (Glock, 2019; Soames, 2010; Speaks, 2024).[22] For example, meaning could consist in some broadly understood notion of use (Brandom, 2008; Dummett, 1993). If meaning simply is some form of use, it is unclear how it could be perceptually presented.[23] Glock (2019) emphasizes that while use may play a fundamental role in any account of meaning, it is doubtful that meaning can be identified with use (see also Child, 2019).[24] A plausible view is that meaning—whatever its ultimate nature—is a kind of *abstract* entity, as opposed to a spatiotemporal entity.

However, it is unclear whether we perceive abstract entities. This would seem to significantly constrain the range of theories of perception compatible with the SPV, as many philosophers consider perception to be of spatiotemporal entities only (Campbell, 2002). But even if *some* abstract entities were perceptible, we would still need to specify what abstract entities can be perceived. As Jacob notes:

> [A]n addressee may understand that the speaker intends to verbally convey to him her belief that there is no greatest integer ... But it does not make much sense to assume that either the speaker or her addressee could perceive what the speaker's utterance is about (Jacob, 2015, p. 9).

Jacob's argument is that at least some abstract entities lie beyond the reach of our perceptual capacities. Entities such as democracy, sets, and numbers do not exert any effect on our

[21]Two objections. First, Wittgenstein writes: "[i]f we hear a Chinese we tend to take his speech for inarticulate gurgling. Someone who understands Chinese will recognize *language* in what he hears" (1977/2006, p. 3e, MS 101 7c). I believe Wittgenstein is mistaken. There is a difference between recognizing something (written signs, speech) as language and understanding *what* it conveys. We do not need to understand Chinese to recognize it as language. The second objection runs as follows. One might argue that someone who does not know Chinese could fake speaking it—by producing phonotactically legal sounds in Chinese—while talking complete nonsense. But this is easily addressed by recalling the fallibility of perception: I may simply fail to judge or perceive the utterance as meaningful, depending on one's preferred account of perceptual error.

[22]It has been suggested that the nature of meaning has no bearing on the intelligibility of the SPV. I strongly disagree. If we take phenomenological plausibility as a criterion for evaluating accounts of perceptual experience, then one way to assess the plausibility of the SPV is to examine theories of meaning and ask whether it makes sense to speak of meaning as perceptible in the first place.

[23]It may be argued that meaning is the kind of entity that is "given" to us. In this case, proponents of the SPV would still need to show that meaning is presented *perceptually* rather than *intellectually* (Bengson, 2015).

[24]Glock (2019, p. 193) rightly notes that Wittgenstein held an apparently ambiguous position on meaning and use—at times approaching their identification (Wittgenstein, 1958/2009, §30, §138), and at other times avoiding it (§43, §139).

senses.[25] Likewise, one could say that meaning does not seem to be the sort of entity that can be perceived, even if one allows some abstract entities in the ontology of perceptual content.

These considerations compel a closer examination of the SPV's positive claim, but they also show that, absent a satisfactory "informative answer," the claim that we perceive meaning remains phenomenologically obscure. Drawing on a suggestion from Drożdżowicz (2019), one might reframe the SPV not in terms of meaning perception, but in terms of the perception of semantic or meaning-related properties. However, this move does little to rescue the SPV from the charge of phenomenological obscurity. First, it remains unclear what such meaning-related properties are—or what distinguishes them from low-level phonological or visual properties. Second, in order to make this a live option, one would also need to clarify how these properties might mediate between low-level properties and meaning. Without such an account, the suggestion remains inert.

In sum, the SPV remains phenomenologically obscure. Compared to the SPV, the rhythmic account offers a stronger claim to phenomenological plausibility. Accordingly, in light of the criterion of phenomenological plausibility, the rhythmic account should be preferred.

4.3.2 | The empirical plausibility criterion

Despite the foregoing phenomenological considerations, one might argue that the SPV is empirically more plausible than other options. I thus turn to the criterion of empirical plausibility.

Brogaard (2017, 2018, 2020) rightly emphasizes that the SPV requires empirical foundations. I will not examine all the empirical evidence Brogaard cites. Two case studies, however, are particularly noteworthy because Brogaard shows how the SPV can respond to the contextualist challenge (Gasparri & Murez, 2021; see also Evans, 2009; Recanati, 2004).[26] The challenge is as follows: if meaning is highly context-dependent, then it seems implausible that meaning can be directly perceived without the aid of high-level inferences (Stanley, 2005). Brogaard argues, however, that a closer look at chunking and garden-path sentences allows us to sidestep the contextualist challenge and provide empirical support for the SPV. Let us first examine chunking.

The brain represents known words as meaningful chunks, rather than by assembling multiple letters together (Glezer et al., 2015). Typically, word recognition occurs when the eyes fixate

[25]In her reply against Jacob, Millikan says that:

> "Democracy," "religion," "electron," and "intelligence" do not reflect light or ruffle the air. But they can be perceived, like everything else perceptible, through infosigns, in this case through linguistic infosigns (Millikan, 2016, p. 200).

Millikan's view is closely aligned with the SPV, though her motivation is not primarily phenomenological. I will therefore set it aside, as it deserves separate discussion. However, my arguments against the phenomenological plausibility of the SPV apply, *mutatis mutandis*, to her account as well.

[26]A few words on the choice of case studies. First, although I focus on Brogaard's reply to the contextualist challenge, my aim is not to evaluate whether her reply succeeds, but to assess whether the empirical evidence she discusses genuinely supports the SPV or provides reasons to prefer it over the rhythmic account. Second, the case studies examined here are not addressed in Calzavarini's recent paper (2022) on the empirical status of the SPV. Calzavarini systematically reviews the empirical evidence for the SPV and reaches a pessimistic conclusion about its empirical viability—conclusions largely compatible with my account.

near the center of the word, rather than decoding each letter sequentially from left to right (at least in languages that proceed from left to right) (Vitu, 2011, p. 735). A striking illustration of this phenomenon is the familiar example of texts with "jumbled" internal letters, where only the first and last letters remain in place:

> [T]he huamn mnid deos not raed ervey lteter by istelf, but the wrod as a wlohe (Millikan, 2016, p. 198).

These words can be read effortlessly by most competent English speakers. From this, Brogaard infers that knowing a word grants us direct perceptual access to its meaning. Rather than processing the individual letters and their order, our brains—thanks to acquired lexical competence—allegedly "jump" to the detection of meaning, bypassing the visual properties of the written sentence (Connolly, 2019, pp. 172–173 for a critique of Brogaard's interpretation of Glezer et al., 2015).

To assess whether word chunking supports the SPV over the rhythmic account, we can apply the criterion of empirical plausibility (Section 2). This criterion first suggests (i) that the proposed explanation not be contradicted by current empirical evidence on chunking. On this front, there is no reason to think that the SPV is empirically refuted: recognizing a word without perceiving each letter individually does not contradict it. However, this fact alone does not straightforwardly support the SPV either. Turning to the second feature (ii) of our criterion, we should expect the SPV to align more closely with the empirical evidence on chunking than rival explanations of the phenomenal contrast.

Chunking does not *necessitate* the SPV. It can just as plausibly be explained by the fact that the cognitive system requires only a few letters in the right positions to trigger a recognitional capacity. This alternative interpretation avoids the need to posit meaning as a perceptual entity and remains compatible with the rhythmic account. Let us now turn to garden-path sentences before concluding the assessment.

A second way in which top-down processes may shape our perception of speech or text is through anticipatory top-down processes. Background expectations and contextual information appear to guide both auditory and visual processing (Section 3.2). A well-known illustration of this is Matlin and Farmer's ambiguous newspaper headlines, and Bever's "horse" sentence:

- Kids make nutritious snacks (Matlin & Farmer, 2016, p. 333).
- The horse raced past the barn fell (Bever, 1970/2013, p. 40).

These are paradigmatic examples of *garden path sentences* (see footnote 13). Despite the apparent ambiguity, one interpretation typically strikes us as most plausible upon first reading. On this basis, Brogaard concludes that:

> [O]ur expectations at a higher level of processing automatically influence lower-level processing, quickly generating an *appearance of the intended meaning* (Brogaard, 2018, pp. 2978–2979; my emphasis).

On Brogaard's interpretation, garden-path effects are best explained by the idea that top-down processes modify the (perceptual) appearance of meaning.

The question, then, is whether these effects provide empirical grounds for favoring the SPV over the rhythmic account. As before, I do not claim that the SPV is refuted by the empirical

evidence—but neither does the evidence offer any reason to prefer it over the rhythmic account. The fact that readers may initially adopt an incorrect interpretation of a garden-path sentence does not entail that a "meaning appearance" is perceptually generated.

In conclusion, applying the criterion of empirical plausibility, it seems reasonable to favor the rhythmic account over the SPV. The former is supported by substantial empirical evidence (Section 3.2), whereas the latter has received little empirical backing (Calzavarini, 2022).

4.4 | Final remarks

One might object that the rhythmic account and the SPV are compatible. In principle, nothing rules out the possibility that linguistic competence alters the rhythmic profile of sensory processes while also enabling the perception of meaning properties. However, the foregoing considerations weigh against this conclusion. If the rhythmic account offers a better overall perceptual explanation of the phenomenal contrast case, there is no need to posit the SPV. If the SPV is introduced to account for the perceptual dimension of the phenomenal contrast, but the rhythmic account already does so in a way that is empirically and phenomenologically more plausible, then the SPV becomes explanatorily redundant.

We are now in a position to see that the "what-strategy" faces a dilemma in this context. Either one adopts a thin account of content, in which case content alone is insufficient to explain the phenomenal contrast—unless supplemented by either a cognitive explanation or the rhythmic account; or one embraces the SPV, which, as we have seen, faces significant phenomenological and empirical difficulties.

A final point: The unfolding of sensory processes can affect perceptual content. This is analogous to the widely discussed influence of attention on perceptual content (Nanay, 2010). Two possibilities follow. First, if changes in rhythmic profile alter content—for example, via segmentation or Gestalt-like grouping—then the phenomenal contrast between cases #1 and #2 can be fully accounted for by differences in low-level content shaped by the dynamic development of sensory processes. Alternatively, if the rhythmic profile does not alter content, we can still allow that #1 and #2 share the same content while still differing in phenomenal character due to differences in how sensory processes unfold.

ACKNOWLEDGEMENTS

Thanks to audiences at the universities of Bochum, Antwerp, and Rutgers. For helpful comments, I am grateful to Julian Bacharach, David Barack, Brian McLaughlin, Kyle Landrum, Kael McCormick, Bence Nanay, Albert Newen, James Openshaw, Andrew Rubner, Francesca Secco, Barbara Tillmann, Ivano Zanzarella, and an anonymous reviewer. While revising this manuscript, I was grappling with the aftermath of trauma caused by abuse. My heartfelt thanks go—among others—to Silvestro Aurilio, Federico Boem, Giulia Candeloro, Angelica Kauffmann, Magdalini Koukou, Anna Luongo, Federica Isabella Malfatti, Joshua Myers, Chiara Rita Napolitano, Nicola and Ottavio Natale, Giusy Pastorelli, Sofia Pedrini, Gabriele Rasi, Zuzia Rucińska, Christian Scholz, Andrea Velardi, and Marco Viola for their support. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

There is no data available.

ORCID

Alfredo Vernazzani https://orcid.org/0000-0002-6458-8478

REFERENCES

Arnal, L. H., Poeppel, D., & Giraud, A. L. (2016). A neurophysiological perspective on speech processing in "the neurobiology of language". In G. Hickok & S. Small (Eds.), *Neurobiology of language* (pp. 463–478). Elsevier.

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology*, *58A*, 1065–1086.

Bayne, T. (2009). Perception and the reach of phenomenal content. *The Philosophical Quarterly*, *59*(236), 385–404.

Bengson, J. (2015). The intellectual given. *Mind*, *124*(495), 707–760.

Bennett, K. (2011). Construction area (no hard hat required). *Philosophical Studies*, *154*, 79–104.

Bever, T. G. (1970/2013). The cognitive basis for linguistic structures. In M. Sanz, I. Laka, & M. K. Tanenhaus (Eds.), *Language down the garden path* (pp. 1–80). Oxford University Press.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*, 227–247.

Block, N. (2010). Attention and mental paint. *Philosophical Issues*, *20*, 23–63.

Bolton, T. (1894). Rhythm. *The American Journal of Psychology*, *VI*(2), 145–238.

Brandom, R. (2008). *Between saying and doing*. Oxford University Press.

Brogaard, B. (2017). The publicity of meaning and the perceptual approach to speech comprehension. *ProtoSociology*, *34*, 144–162.

Brogaard, B. (2018). In defense of hearing meanings. *Synthese*, *195*, 2967–2983. https://doi.org/10.1007/s11229-016-1178-x

Brogaard, B. (2020). Seeing and hearing meanings: A non-inferential approach to speech-comprehension. In A. Nes & T. Chan (Eds.), *Inference and consciousness* (pp. 99–124). Routledge.

Burge, T. (2010). *Origins of objectivity*. Oxford University Press.

Burnston, D. (2021). Perceptual learning, categorical perception, and cognitive permeation. *Dialectica*, *75*(1), 25–58.

Byrne, A. (2002). Don't PANIC: Tye's intentionalist theory of consciousness. In A field guide to the philosophy of mind*, symposium on Tye's consciousness, color, and content*. https://philpapers.org/rec/BYRDPT

Calzavarini, F. (2022). The empirical status of semantic perceptualism. *Mind & Language*, *38*, 1000–1020. https://doi.org/10.1111/mila.12444

Campbell, J. (2002). *Reference and consciousness*. Oxford University Press.

Chalmers, D. (2004). The representational character of experience. In B. Leiter (Ed.), *The future of philosophy* (pp. 153–181). Oxford University Press.

Child, W. (2019). Meaning, use, and supervenience. In J. Conant & S. Sunday (Eds.), *Wittgenstein on philosophy, objectivity and meaning* (pp. 211–230). Cambridge University Press.

Chudnoff, E. (2015). *Cognitive phenomenology*. Routledge.

Connolly, K. (2019). *Perceptual learning*. Oxford University Press.

Cowell, H. (1930/1996). *New musical resources*. Cambridge University Press.

Darnton, R. (1984). *The great cat massacre*. Basic Books.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science*, *330*, 1359–1364.

Douven, I. (2025). Abduction. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/sum2021/entries/abduction/

Drożdżowicz, A. (2019). Do we hear meanings? Between perception and cognition. *Inquiry*, *66*, 196–228. https://doi.org/10.1080/0020174X.2019.1612774

Drożdżowicz, A. (2021). Experiences of linguistic understand as epistemic feelings. *Mind & Language*, *38*(1), 274–295.

Dummett, M. (1993). *The seas of language*. Oxford University Press.

Elgin, C. Z. (2017). *True enough*. MIT Press.

Evans, V. (2009). *How words mean*. Oxford University Press.

Fodor, J. (1983). *The modularity of mind*. MIT Press.
Gasparri, L., & Murez, M. (2021). Hearing meanings: The revenge of the context. *Synthese*, *198*, 5229–5252.
Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*, 511–517. https://doi.org/10.1038/nn.3063
Glezer, L., Kim, J., Rule, J., Jiang, X., & Riesenhuber, M. (2015). Adding words to the brain's visual dictionary: Novel word learning selectively sharpens orthographic representations in the VWFA. *The Journal of Neuroscience*, *35*(12), 4965–4972.
Glock, H.-J. (2019). What is meaning? A Wittgensteinian answer to an un-Wittgensteinian question. In J. Conant & S. Sunday (Eds.), *Wittgenstein on philosophy, objectivity and meaning* (pp. 185–210). Cambridge University Press.
Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*(12), e1001752. https://doi.org/10.1371/journal.pbio.1001752
Haegens, S., & Golumbic, E. Z. (2017). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience and Biobehavioral Reviews*, *86*, 150–165. https://doi.org/10.1016/j.neubiorev.2017.12.002
Hoerl, C. (2013). A succession of feelings, in and of itself, is not a feeling of succession. *Mind*, *122*(486), 373–417.
Jacob, P. (2015). Millikan's teleosemantics and communicative agency. In T. Metzinger & J. Winds (Eds.), *Open MIND 20(T)*. MIND Group.
James, W. (1890/2019). *The principles of psychology* (Vol. 1). Dover.
Jenkin, Z. (2022). Perceptual learning and reason-responsiveness. *Noûs*, *57*, 481–508. https://doi.org/10.1111/nous.12425
Johnston, M. (2004). The obscure object of hallucination. *Philosophical Studies*, *120*, 113–183.
Josephson, J. R., & Josephson, S. G. (1994). *Abductive inference: Computation, philosophy, technology*. Cambridge University Press.
Kemmerer, D. (2023). *Cognitive neuroscience of language*. Routledge.
Koksvik, O. (2015). Phenomenal contrast: a critique. *American Philosophical Quarterly*, *52*(4), 321–334.
Kriegel, U. (2002). Phenomenal content. *Erkenntnis*, *57*, 175–198.
Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition, and Neuroscience*, *31*(1), 32–59.
Langus, A., Mehler, J., & Nespor, M. (2017). Rhythm in language acquisition. *Neuroscience and Biobehavioral Reviews*, *81*, 158–166.
Lipton, P. (2004). *Inference to the best explanation*. Routledge.
London, J. (2012). *Hearing in time: Psychological aspects of musical meter*. Oxford University Press.
MacDonald, M. C., & Hsiao, Y. (2018). Sentence comprehension. In S.-A. Rueschenmeyer & G. Gaskell (Eds.), *The Oxford handbook of psycholinguistics* (pp. 171–196). Oxford University Press.
Machery, E. (2009). *Doing without concepts*. Oxford University Press.
Martin, M. G. F. (2002). The transparency of experience. *Mind & Language*, *17*, 376–425.
Matlin, M., & Farmer, T. (2016). *Cognition* (9th ed.). Wiley.
McConkie, G. W., Zola, D., Grimes, J., Kerr, P. W., Bryant, N. R., & Wolf, P. M. (1991). Children's eye movement during reading. In J. F. Stein (Ed.), *Vision and visual dyslexia* (pp. 251–262). Macmillan.
McDowell, J. (1984). De re senses. *The Philosophical Quarterly*, *34*, 283–294.
McDowell, J. (1998). Meaning, communication, and knowledge. In *His meaning, knowledge and reality* (pp. 29–50). Harvard University Press.
Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, *48*, 2609–2621.
Millikan, R. G. (2016). *Beyond concepts*. Oxford University Press.
Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, *83*(4), 435–450.
Nanay, B. (2010). Attention and perceptual content. *Analysis*, *70*(2), 263–270.
Nes, A. (2016). On what we experience when we hear people speak. *Phenomenology and Mind*, *10*, 58–85.
Newen, A. (2016). Defending the liberal-content view of perceptual experience: Direct social perception of emotions and person impressions. *Synthese*, *194*, 761–785. https://doi.org/10.1007/s11229-016-1030-3
Noë, A. (2012). *Varieties of presence*. Harvard University Press.
O'Callaghan, C. (2010). Experiencing speech. *Philosophical Issues*, *20*(1), 305–332.

O'Callaghan, C. (2011). Against hearing meanings. *The Philosophical Quarterly*, *61*(245), 783–807.
O'Callaghan, C. (2015). Speech perception. In M. Matthen (Ed.), *The Oxford handbook of the philosophy of perception* (pp. 475–494). Oxford University Press.
O'Shaughnessy, B. (2000). *Consciousness and the world*. Oxford University Press.
Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, *25*, 1649–1653.
Peacocke, C. (1992). *A study of concepts*. MIT Press.
Peelle, J., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*, 1378–1387.
Pettit, D. (2010). On the epistemology and psychology of speech comprehension. *The Baltic International Yearbook of Cognition, Logic and Communication*, *5*, 1–43.
Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Neuroscience*, *21*, 322–334. https://doi.org/10.1038/s41583-020-0304-4
Price, H. H. (1932). *Perception*. Methuen.
Prinz, J. (2006). Beyond appearances: The content of sensation and perception. In T. Szabó-Gendler & J. Hawthorne (Eds.), *Perceptual experience* (pp. 434–459). Oxford University Press.
Prosser, S. (2016). *Experiencing time*. Oxford University Press.
Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292.
Rayner, K., Chace, K. H., Slattery, T., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*(3), 241–255.
Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*, 191–201.
Rayner, K., & Liversedge, S. P. (2011). Linguistic and cognitive influences on eye movements during reading. In S. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 751–766). Oxford University Press.
Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*, 504–509.
Recanati, F. (2004). *Literal meaning*. Cambridge University Press.
Reiland, I. (2015). On experiencing meanings. *The Southern Journal of Philosophy*, *53*(4), 481–492.
Riecke, L., Formisano, E., Sorger, B., Başkent, D., & Gaudrain, E. (2018). Neural entrainment to speech modulates speech intelligibility. *Current Biology*, *28*, 161–169.
Rimmele, J., Gross, J., Molholm, S., & Keitel, A. (Eds.). (2018). *Brain oscillations in human communication*. Frontiers in Human Neuroscience.
Sattig, T. (2019). The sense of temporal flow: A higher-order account. *Philosophical Studies*, *176*, 3041–3059.
Schellenberg, S. (2010). The particularity and phenomenology of perceptual experiences. *Philosophical Studies*, *149*, 19–48.
Schellenberg, S. (2018). *The unity of perception*. Oxford University Press.
Schotter, E. R., & Rayner, K. (2015). The work of the eyes during reading. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 44–59). Oxford University Press.
Shoemaker, S. (2006). The Frege-Schlick view. In J. J. Thomson (Ed.), *Content and modality: Themes from the philosophy of Robert Stalnaker* (pp. 18–33). Clarendon Press.
Siegel, S. (2006). Which properties are represented in perception? In T. Szabó Gendler & J. Hawthorne (Eds.), *Perceptual experience* (pp. 481–503). Oxford University Press.
Siegel, S. (2007). How can we discover the contents of experience? *Southern Journal of Philosophy*, *XLV*, 127–142.
Siegel, S. (2010). *The contents of visual experience*. Oxford University Press.
Siegel, S. (2013). The epistemic impact of the etiology of experience. *Philosophical Studies*, *162*(3), 697–722.
Simons, P. (2019). The ontology of rhythm. In P. Cheyne, A. Hamilton, & M. Paddison (Eds.), *The philosophy of rhythm* (pp. 62–75). Oxford University Press.
Smith, A. D. (2002). *The problem of perception*. Harvard University Press.
Soames, S. (2010). *Philosophy of language*. Princeton University Press.
Song, H., & Schwarz, N. (2008). If it's hard to read it's hard to do: Processing fluency affects effort prediction and motivation. *Psychological Science*, *19*(10), 986–988.

Soteriou, M. (2013). *The mind's construction*. Oxford University Press.
Speaks, J. (2024). Theories of meaning. In E. Zalta & U. Nodelman (Eds), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/win2024/entries/meaning/
Stanley, J. (2005). Hornsby on the phenomenology of speech. *The Aristotelian Society Supplementary Volume*, *79*(1), 131–145.
Strawson, G. (2010). *Mental reality*. MIT Press.
Tye, M. (1995). *Ten problems of consciousness*. MIT Press.
Tye, M. (2007). Intentionalism and the argument from no common content. *Philosophical Perspectives*, *21*, 589–613.
Vetter, P., & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and Cognition*, *27*, 62–75.
Vitevitch, M. S., & Aljasser, F. S. (2021). Phonotactics in spoken-word recognition. In J. S. Pardo, L. C. Nygaard, R. E. Remez, & D. B. Pisoni (Eds.), *The handbook of speech perception* (pp. 286–307). Wiley.
Vitu, F. (2011). On the role of visual and oculomotor processes in reading. In S. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 731–749). Oxford University Press.
Vitu, F., O'Regan, K., Inhoff, A. W., & Topolski, R. (1995). Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. *Perception & Psychophysics*, *57*(3), 352–364.
Walton, D. (2006). *Character evidence: An abductive theory*. Springer.
Walton, D. (2013). *Methods of argumentation*. Cambridge University Press.
Wittgenstein, L. (1958/2009). In P. M. S. Hacker & J. Schulte (Eds.), *Philosophical investigations*. Wiley-Blackwell.
Wittgenstein, L. (1977/2006). *Culture and value*. Blackwell.
Zeimbekis, J., & Raftopoulos, A. (2015). *The cognitive penetrability of perception*. Oxford University Press.

How to cite this article: Vernazzani, A. (2025). Language comprehension and the rhythm of perception. *Mind & Language*, 1–23. https://doi.org/10.1111/mila.70003