

Phonetic Segments and the Organization of Speech

PREPRINT | To appear in *Philosophy of Science*

Abstract: According to mainstream linguistic phonetics, speech can be modeled as a string of discrete sound segments or “phones” drawn from a universal phonetic inventory. Recent work has argued that a mature phonetics should refrain from theorizing about speech and speech processing using sound segments, and that the phone concept should be eliminated from linguistic theory. The paper lays out the tenets of the phone methodology and evaluates its prospects in light of the eliminativist arguments. I claim that the eliminativist arguments fail to show that the phone concept should be eliminated from linguistic theory.

Author: Luca Gasparri, Institut Jean Nicod (ENS, EHESS, CNRS, PSL University), 29 rue d’Ulm, 75005 Paris, France. lcgasparri@gmail.com

Acknowledgments: I am indebted to Claire Beyssade, Paul Egré, Mikhail Kissine, Salvador Mascarenhas, and François Recanati for discussion and/or comments on ancestors of this paper, and to two anonymous reviewers for detailed and very helpful feedback on the submitted manuscript. All errors and misunderstandings are my own. The research that led to this paper was supported by an individual fellowship from the École des Hautes Études en Sciences Sociales, and by grants ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL*.

1. Introduction

According to the received way of dividing up the study of linguistic sounds, phonetics investigates the low-level properties of speech, whereas phonology is concerned with the way speech sounds are organized in a language. The division of labor is set roughly as follows. Phonetics produces an inventory of speech sound types or “phones” individuated according to their articulatory and acoustic characteristics. Phonology describes how the sound systems of individual languages group the elements of this inventory into patterns of functional sameness and difference (Davenport and Hannahs 2013). For example, phonetics categorizes the speech sound type instantiated by the burst of breath that accompanies the release or the closure of an obstruent consonant as an “aspiration”, and symbolizes it as [^h]. Phonology proceeds to clarify whether in a given language the presence of aspiration is contrastive, or aspirated consonants are allophones in complementary distribution with their unaspirated counterparts. In English, the speech sound types [t] and [t^h] are perceived as instances of the same phonological element /t/ and do not mark lexical distinctions: [teɪst] and [t^heɪst] are both acceptable pronunciations of ‘taste’ (though the latter is preferred by natives). In Hindi, by contrast, the speech sound types [p] and [p^h] are perceived as instances of two distinct phonemes /p/ and /p^h/, and their alternation is lexically contrastive: [pəl] and [p^həl] correspond to two different words, ‘pal’ and ‘phal’, meaning “moment” and “fruit”, respectively.

A key assumption underlying this approach to linguistic sounds, formalized by classical post-SPE phonology (SPE = *The Sound Pattern of English*, Chomsky and Halle 1968) and derived from earlier feature theories (Jakobson, Fant and Halle 1952; Hockett 1955) is what we can call the Phone Principle (PP). PP can be stated as follows.

- (PP) Speech can be modeled as a series of discrete sound units drawn from a universal phonetic inventory.

This assumption goes back to the earliest serious investigations of language. From Panini's grammar of Classical Sanskrit, which operated on discrete segments of speech, to Halle's (1954) "phonemic strategy", Hockett's (1960) "duality of patterning", and most contemporary theories of phonetics and phonology, PP has consistently represented one of the most basic methodological principles of our inquiries into the sound structure of human languages (see Raimy and Cairns 2015). The idea that connected speech is structured into letter-like segments (e.g., concatenations of consonants and vowels) is at the same time so intuitive and such a fundamental part of the way we theorize about language, that its assumption is often completely implicit, and its tenability seldom discussed in any critical fashion.

Yet, the evidence for PP is conflicting, and nowadays the principle is by no means uncontested. Morphophonemic and lexical phenomena speak strongly in its favor, but some psycholinguistic and acoustic data raise concerns about its soundness. Building on these opposing results, PP has been called into question by eliminativist views of the phone concept.¹ Taking issue with the received phonetic agreement, such approaches have claimed that we should refrain from characterizing spoken utterances as series of discrete sound units drawn from a universal phonetic inventory. A mature or complete linguistic phonetics should eliminate phones and model speech as an unsegmented continuum with properties to be expressed in purely quantitative-parametric terms. This raises an overarching question: what sort of claims

1 Nota bene: the labels 'eliminativism' and 'eliminativist' are nowhere to be found in the phonetic literature.

However, as will rapidly become clear, the anti-PP arguments we will consider can be safely filed under the philosophical rubric of eliminativism, as their main claim is that the phone concept is unsuited for sustained theorizing about the properties of connected speech and the dynamics of speech processing, and thus should be eliminated from linguistic theory (by analogy, think of the brand of eliminativism about the species concept discussed by Ereshefsky 1998).

can be made about the status of phones and about PP's ability to generate adequate theories of the organization of speech?

This paper has a broad rationale and a specific purpose. The broad rationale is to draw attention on a branch of the study of language which appears to raise major methodological questions (in this case: in what sense and to what extent is it possible to appeal to PP in the scientific description of speech and speech processing?), questions that are at the center of a lively debate among linguists, and nonetheless have been mostly neglected by philosophers of science (see, however, Appelbaum 1999; Bromberger and Halle 2000; Wetzel 2009; Carr 2012). There is relatively little existing work in the philosophy of linguistics (Scholz, Pelletier and Pullum 2016), and the attention that philosophers have dedicated to phonetics is even more marginal. I hope to be able to show that this discipline harbors a number of foundational questions that deserve deeper philosophical scrutiny.²

The specific purpose is to unpack the claims underlying PP, characterize the eliminativist arguments against them, and lay out a first informed assessment of the prospects of PP in light of the eliminativist arguments. The take-home message I will defend is that the eliminativist line of thinking does force us to reexamine our grounds for accepting the phone methodology, but fails to show that the phone concept should be eliminated from linguistic theory.

The discussion will proceed as follows. Section 2 will break PP down into three specific claims (Discreteness, Universality, and Computationalism) and introduce the eliminativist objection against each. Section 3 will respond to the eliminativist argument against Discreteness.

2 Along the way, the reader should also notice that the issue of the tenability of PP presents obvious points of contact with mainstream debates in philosophy of science, from classical work on the analog-digital divide in cognition (Dretske 1981), to more recent work on the nature of scientific models (e.g., Weisberg 2013), the semantics of scientific theories (e.g., Azzouni 2010), and the admissibility of idealization in science (e.g., Elgin 2017; Potochnik 2017).

Section 4 will respond to the eliminativist argument against Universality. Section 5 will respond to the eliminativist argument against Computationalism. Section 6 will conclude.

2. Digital Phonetics

We can get a clearer understanding of the specific claims underlying PP by means of a concrete example. Consider a normal utterance of the word ‘taste’. According to proponents of PP, we can claim what follows.

- a) Utterances of ‘taste’ can be described as concatenations of the sound segments [t^h], [e], [ɪ], [s], and [t].
- b) The segments [t^h], [e], [ɪ], [s], and [t] instantiate each an individual speech sound type or “phone” represented in a universally available phonetic inventory (IPA 1999).
- c) Utterances of ‘taste’ are processed in speech perception by association to discrete mental symbols corresponding to [t^h], [e], [ɪ], [s], and [t], and discreteness at the level of phonetic representation guarantees discreteness at the higher levels of grammatical processing.

Abstracting from the ‘taste’ example, PP can thus be decomposed into the following three claims (cf. Kenstowicz 1994; Laver 1994; Ladd 2011).

- | | |
|--------------------|--|
| (Discreteness) | Speech can be described as the concatenation of discrete sound segments. |
| (Universality) | Speech can be described as the concatenation of a finite set of speech sound types picked from a universally available phonetic inventory. |
| (Computationalism) | Real-time speech processing can be described as the manipulation of discrete phone-sized mental symbols. |

Discreteness and Universality argue for the viability of phone-based descriptions of the physical makeup of connected speech, and therefore make a claim about the organization of speech specifically. Computationalism argues for the viability of phone-based descriptions of the information manipulated by listeners in speech processing tasks, and therefore makes a claim about the nature of the mental representation of speech.

For illustration, here is how speech processing is understood to operate for those subscribing to Discreteness, Universality, and Computationalism (cf. Hale and Reiss 2008). Speech is emitted by a speaker, it spreads through an elastic medium (air), and reaches a listener. Here it generates a raw auditory percept, which stores information about the speech signal and the other sounds present in the environment. The raw auditory percept is then broken down into its sub-components. The operation involves first isolating the aspects of the raw auditory percept which are due specifically to the speech signal (as opposed to, say, the sound of the washing machine in the background), and then withdrawing from this body of speech-specific information data about voice quality, the speaker's emotional state, the linguistic content of the utterance, and the like. Subsequently, the information about the linguistic content of the utterance is "printed" into a speech score. This is transmitted to a parser, which breaks the speech score into sound segments, analyzes their physical characteristics, pairs them to a mental symbol, and outputs a phonetic score (e.g., [t^h]-[e]-[ɪ]-[s]-[t]). The phonetic score is a concatenation of digital mental symbols which is readable at the interface with phonology and feeds the additional processes required to perform higher-order tasks such as word recognition and semantic interpretation.

This is the orthodox story. Even Optimality Theory (Prince and Smolensky 2004), a now mainstream approach to phonology whose view of phonological grammar diverges significantly from the classical SPE codification, continues to assume that the phonological component of the

grammar computes digital representations captured by the phone concept, and that the linear string is cut into a number of chunks or segments that delineate the application of phonological processes (Lombardi 2001). However, in the last couple of decades, in good part following the emergence of laboratory phonology (Kingston and Beckman 1990), mainstream assumptions about the descriptive legitimacy of PP-style phonetics have come under increasing scrutiny, and have been called into question.

To be sure, contemporary phonetics and phonology abound with controversies about central conceptual tenets of these disciplines (Dunbar and Idsardi 2010), and stances opposing PP or casting doubts on its viability hardly make up a uniform front. Reactions have ranged from Klatt's (1979) early work on power spectra, to the exemplar theory of Goldinger (1996), to articulatory phonology (Browman and Goldstein 1992), to the full-blooded rejection of formal phonology championed by Port and Leary (2005) and Port (2007a). However, it seems fair to say that these different approaches share, in different ways and degrees, an overarching conceptual thrust: PP-style phonetics is not an accurate way of looking at the acoustic realization of language and at the mental representation of speech. The models of speech and of the mental representation of speech produced on the basis of the phone idealization should be eliminated and replaced by a descriptive framework making no appeal to phones.

Call this Phone Eliminativism (PE). Parallel to PP, PE can be formulated as follows.

(PE) Speech cannot be modeled as a series of discrete sound units drawn from a universal phonetic inventory.

Again parallel to PP, PE can be decomposed into the following three claims.

- (Anti-Discreteness) Speech cannot be described as the concatenation of discrete sound segments.
- (Anti-Universality) Speech cannot be described as the concatenation of a finite set of speech sound types picked from a universal phonetic inventory.
- (Anti-Computationalism) Real-time speech processing cannot be described as the manipulation of discrete phone-sized mental symbols.

Anti-Discreteness and Anti-Universality insist on the absence of a well-behaved correspondence between PP-style phonetic transcriptions and concrete speech. They maintain that a careful observation of the low-level properties of connected speech reveals that it is unsuited to be characterized as a concatenation of segments instantiating universally available phonetic types. Support to this conclusion includes the notorious imprecision of phonetic transcriptions, effects of coarticulation (absence of clear boundaries between segments), the noisy nature of speech, evidence that individual phones are implemented by different languages through acoustically dissimilar sound configurations, the displacement of acoustic cues for individual segments across the entire signal, and the presence in the speech signal of phonologically relevant timing elements that resist translation into discrete symbols (Fodor, Bever and Garrett 1972; Liberman 1996; Gafos 2002).

Anti-Computationalism argues that there is no stage in speech processing properly describable as the translation of speech into chains of phone-like mental symbols, and draws the epistemological moral that the phone concept is unsuited to psychological theorizing (Port 2010). Instead of viewing speech perception as a routine based on the manipulation of symbolic phonetic scores, real-time speech analysis should be modeled within a dynamic-embodied framework of the sort introduced by van Gelder (1995), Kelso (1995), and Clark (1997).

Can PP respond to these challenges? And if so, how so? The answer plausibly hinges on the possibility of producing a version of Discreteness, Universality, and Computationalism that reasserts the phone methodology while taking into account the best instances of the eliminativist line of thinking. The key, I will submit, lies in committing to PP as a hypothesis about the level up to which the phone methodology is conducive to sound generalizations about the physical makeup and the cognitive organization of speech, and in construing PE as a pointer to the upper boundary of PP's explanatory power.

3. Discreteness

Let us start with Anti-Discreteness: speech cannot be described as the concatenation of discrete sound segments. There are two basic ways of construing Anti-Discreteness. One is to read it as the claim that the phone idealization cannot generate reliable descriptions of the physical makeup of connected speech, which is essentially continuous, and thereby should be eliminated on the basis of its base-level ontological inadequacy. The other is to read it as the claim that PP-style descriptions of connected speech are too coarse-grained and arbitrary to lead to dependable models of the physical organization of utterances, and thereby should be eliminated because the way they idealize speech cannot constitute the premise of adequate (non-ontological) theorizing.

The ontology-driven interpretation of Anti-Discreteness can be neutralized quite easily. Whether or not strings of phones allow for an exhaustive description of the physical makeup of speech is irrelevant to the explanatory interests at issue within linguistic phonetics, and proponents of Discreteness are certainly not committed to it as a posit of fine-grained descriptive ontology (i.e., to the notion that the phonetic vocabulary tracks measurable or observable entities in the external world). What matters, much more practically, is whether modeling speech as a sequence of phones allows us to render it tractable relative to the explanatory goals of linguistic

theory. For example, Port and Leary (2005) are certainly right in stressing that atemporal phones are constitutively unsuited to be found in a temporal signal, no matter how meticulously analyzed. But from this simple premise, the case for Anti-Discreteness does not take off. The commitment to an analog ontology for speech (i.e., utterances are continua which are not materially divided into discrete phonetic units) is fully compatible with the claim that utterances are continua with gradient internal discontinuities (e.g., alternating degrees of energy and sonority) which become scientifically tractable only once modeled as concatenations of discrete phones (cf. Lieberman and Blumstein 1988).³

As for the second interpretation of Anti-Discreteness, it is useful to illustrate the point with an example (from Lodge 2009). Consider a standard utterance of the word ‘bed’. It is supposed to contain three segments: [b], [e], and [d]. Yet, advocates of Anti-Discreteness insist, this characterization fails to account for some crucial asymmetries between the physical parameters operating within the purported segments, which are themselves temporally complex. Voicing starts after the bilabial closure and ceases before the final release of the alveolar closure. Furthermore, the tongue is already transitioning to the position required to articulate [e] before the completion of bilabial closure. Thus, the initial stages of a typical utterance of ‘bed’ encompass at least the following events:⁴

- 3 Incidentally, recall that even the discourse domains of highly reductive disciplines such as macro-physics include objects that do not reflect or only “obliquely reflect” the ontological commitments of their practitioners (Azzouni 2010). We treat Mars as a point-mass and as a homogeneous spheroid because this allows us to formulate efficient predictions about Mars’s gravitational effects on neighboring planets, even if the notions of point-mass and homogeneous spheroid hardly pick out anything we would define as “existing” or “real” in our metaphysical moments.
- 4 I am glossing over the fact that for many speakers of English utterance-initial voiced stops like /b/ are actually voiceless throughout the period of bilabial closure and voicing does not begin until 10-15 ms after the release of the closure. In such cases, the contrast between utterance-initial /b/ and utterance-initial /p/ is manifested in the fact that in /p/ voicing does not begin until after an aspiration (usually 50 ms or more after the release of the

- i) the tongue transitions to the position needed for the vocalization of [e];
- ii) bilabial closure occurs;
- iii) voicing starts;
- iv) after the build-up of pressure above the glottis, bilabial closure is released;
- v) [e] is vocalized.

The temporal distribution of the parameters involved can be represented with a simplified variant of the notation used in articulatory phonology, as in figure 1.

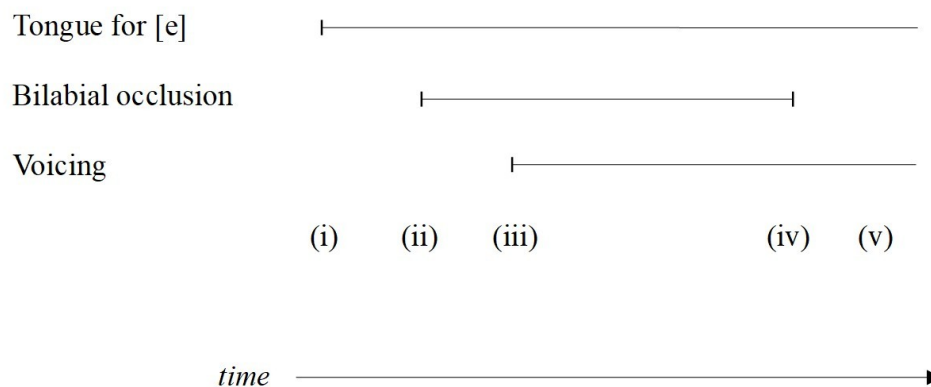


Figure 1

Instead of being arranged into a linear sequence, the parameters evolve independently and overlap in a complex way here, one that casts doubt on the supposed naturalness of cross-parametric phones. The segments [b], [e] and [d] do not emerge “transparently” from the chart (Fowler 2015), and the theorist is left free to choose their exact positioning. But if the exact

closure). The situation, hence, is a bit more complex than what I am explicitly recognizing here, but this does not affect the argument. Thanks to an anonymous reviewer for pressing me on this point.

localization of the phones involved cannot be but the result of an active deliberation on the theorist's part, how can we trust PP in its ability to inform reliable theorizing?

Let us grant that a 1:1-scale description of the structural information represented in the chart would not appeal to [b], [e] and [d]. Still, it appears that turning the example into a case against Discreteness would require more effort. To start, even if the notation used to represent the temporal evolution of the parameters does not resort to cross-parametric phones, it nonetheless cuts up the speech continuum in a segmental way, as the representation is still committed to the existence of discontinuities at the positions (i)-(iv). For example, the temporal interval between (ii) and (iii) delineates a chunk eligible to be picked out by a segmental vocabulary because it is preceded and followed by two intervals in which bilabial closure is absent and voicing is present, respectively.

Second, the impossibility of identifying without a deliberate theoretical decision where, e.g., the segment [b] is positioned does not seem to imply that any attempt to describe utterances of 'bed' as containing a [b] unit is hopeless, precisely because the discontinuities at (i)-(iv) make the decision possible. Once we posit that [b] is instantiated whenever the articulators and the speech signal exhibit the familiar set of signature features (manner of articulation: occlusive; place of articulation: bilabial; phonation: voiced), we can coherently deliberate that [b] is instantiated between (iii) and (iv), which happens to be the only interval of time where both [Bilabial occlusion] and [Voicing] are present. Accordingly, we can coherently deliberate that our utterance of 'bed' starts at (iii), at that the events occurring before the discontinuity at (iii) simply prepare the actual utterance.

A potentially more serious issue is that the exact temporal location of the discontinuities (i)-(iv) themselves is fuzzy. For example, the rearrangement of the lips from the occlusion of [b] to the position at the center of [e] occupies a non-null interval of time (typically between 60 and 100 ms): (iv) can be arbitrarily placed at any location within this delta. However, the

classificatory elasticity licensed by the vagueness of these boundaries does not imply that the distinction between [b] and [e] is entirely artifactual, since it is constrained by the macroscopic variation in parameter values occurring with the time flow. To express the point with an analogy familiar to philosophers, using the temporal fuzziness of the discontinuity at (iv) as an argument against describing the signal as segmented into [b] and [e] would be close to maintaining that the inability to exactly locate the beginning of the Seine is proof that the Seine is not an objective whole (cf. Varzi 2011). The evidence does imply that the segments in question can enjoy an individuality only as a result of a fiat: their “objectivity” is independently enforced by the broad transitions in parameter values highlighted by the chart, whereas their “individuality” depends on an active theoretical decision on our part. However, this appears sufficient to warrant the operational division into phones called for by Discreteness. Speech signals are continua structured into objective, gradiently alternating regions that can be represented by convention as concatenations of phones for the purposes of efficient theoretical treatment.

4. Universality

Let us now turn to Anti-Universality. This horn of PE builds on the observation that phones like [b], [e], and [d] are implemented in different languages and phonetic contexts through speech sounds having highly, sometimes radically different physical and articulatory characteristics. This seems to jeopardize Universality’s project of capturing the almost unlimited diversity of speech production by relying on the fixed catalog of phones familiar to users of the IPA. I shall discuss two main variants or aspects of the issue: the Variability Problem and the Locality Problem.

The Variability Problem arises from the difference in physical characteristics between the acoustic configurations supposedly implementing the same phonetic elements in different languages. For example, the VOT (voice onset time) values for [p] and [b] vary noticeably

across languages, even across dialectal variants of the same language, such as standard French and Quebec French (Caramazza and Yeni-Komshian 1974). Similarly, the phones grouped under English /i/ are usually higher and more front than those falling under Spanish /i/, whereas the sounds corresponding to Spanish /u/ are closer than English /u/ to cardinal /u/ (Bradlow 1995). The same goes for the durational reflexes of post-vocalic /s/-/z/ in French and English (Flege and Hillenbrand 1986).

The Locality Problem lies in the fact that distinctive features often “bleed” outside their target position and affect the acoustic implementation of speech sounds in their neighborhood (Beckman and Edwards 2010). For example, in words like ‘butting’ and ‘budding’ the phonemes /t/ and /d/ should both be “neutralized” into a flap. This is reflected in the standard practice of notating the speech sounds produced in the middle of both words as [r]. However, the acoustic configurations actually generated in the production of the two segments are not physically symmetric, because the /t/ of ‘butting’ is systematically implemented by native speakers of American English as a phone oddly in between an alveolar flap, the segment’s expected acoustic implementation in a neutralizing context, and an alveolar stop, the segment’s canonical realization in non-neutralizing contexts. Yet, subjects presented with minimal pairs like ‘budding’ vs. ‘butting’ appear able to set them apart better than chance. This indicates, plausibly, that the features identifying the neutralized segment are in fact outside the position it is taken to occupy in standard phonetic transcriptions, and reside in the preceding vowel (Fox and Terbeek 1977; cf. Kleber, John and Harrington 2010 for data about German).⁵

5 Likewise, Kelly and Local (1986) observe that in Standard Southern British English syllable-onset /l/ (realized as a clear palatal) and syllable-onset /r/ (realized as a dark palatal) yield resonance effects that color the entire syllable of which they are part (e.g., the /i/ of ‘Henry’ is darker than the /i/ of ‘Henley’) and may modify the acoustic characteristics of vowels even at 2-3 syllables of distance. Importantly, these non-local cues can be used by listeners to identify lexical /l/ and /r/. This suggests that long-domain co-articulatory information about the /l/-/r/ distinction is available to listeners outside the particulate position purportedly occupied by the

Let us see how Universality can react. The Variability Problem can be mitigated in two immediate ways. A first strategy would be to increase the acoustic granularity of the inventory of phones available to transcriptions. On this solution, for example, the 3-4 VOT value types originally envisioned by Chomsky and Halle (1968) would be raised to the number of values required to account for the entire scope of the VOT targets exploited for lexical contrast in known languages, the speech sounds implementing Spanish /i/ and English /i/ would be paired to different phones, and the issue of cross-linguistic variability would dissolve. One might worry that an excessively fine-grained and rich inventory of phones would run the risk of rendering speech as hard to describe as it would be if it were assumed to be continuous, thus undermining the purported explanatory advantage of adopting the segmental framework. But to the best of my knowledge, there is no principled reason to suspect that the expansion at stake would necessarily have the proportions required to jeopardize the base-level assumption of a fixed inventory of phones.⁶

A second, possibly more sophisticated strategy would be to keep the phonetic catalog at its current size (160+ items) and assign each different language or dialect a specific set of fine-grained acoustic realizers for the basic phonetic arsenal. The move could be implemented by proposing that individual languages and dialects are, among other things, functions mapping the segment (West 1999), and hence that the alleged phone is in fact “distributed” in a region extending far beyond the boundaries of a single sound segment.

6 An instructive parallel can perhaps be made with the domain of phoneme inventories. According to the most conservative estimates of the size of the vowel inventory of Modern German, it comprises less than ten phonemes. This assumption is challenged by more liberal estimates, which allow up to almost twenty phonemes (Wiese 2000). Theories of the sound structure of German positing a rich vowel inventory, however, converge with their conservative counterparts on the methodology of minimal pairs as a criterion for determining phonemic status and on the utility of phonemes in the study of sound structure. Likewise, it should be possible to debate on the appropriate size of the phonetic inventory (and, e.g., argue for a richer catalog of phones) while remaining within the conceptual framework of PP.

phonetic alphabet to loci in Speech Space, conceived of as the collection of all sound configurations capable of being generated by an idealized human vocal apparatus. On this basis, one could further argue that each language comes with a proprietary vocabulary of speech sound types, that each specific speech sound type corresponds to a distinctive position in Speech Space, and that the same phone may be realized in two languages through acoustic configurations corresponding to two different positions in Speech Space. Of course, the issue would remain of giving a precise account of the conditions to be met in order for two speech sound types (or loci in Speech Space) to be viewed as instances of the same phone, given that the classification no longer runs on sameness in low-level properties. A weaker, counterpart-inspired notion of similarity could do the job. In any case, a response to the Variability Problem running along these lines is certainly open and should be evaluated with due attention before jumping to the conclusion that Universality is bound to collapse.

The Locality Problem can be addressed by insisting on the specific explanatory work that the phone methodology is designed to perform, which is essentially grouping under general type-categories chunks of speech that share a sufficient amount of signature functional-constitutive properties while differing, even systematically, at the level of acoustic detail. We mentioned that coarticulation and coloring effects seem in principle inconsistent with the idea that speech may be usefully described as the concatenation of discrete elements having a stable set of low-level properties. However, coarticulation and coloring (and the evidence of their role in establishing hypotheses about the lexical form conveyed by a chunk of speech) do not obliterate the basic acoustic blueprint of the segments involved. Stevens (1989) discusses a relevant example. Place the tongue against your upper teeth and produce [s]. Keep the shape of the tongue constant and slowly move it towards the back of palate. You should notice (irrespective of your sound system of reference) that after a relatively long interval where the movement of the tongue has no categorial effect, there is an abrupt transition to [ʃ], followed by

another steady state in which the movement of the tongue does not alter phonetic category. Thus, in the experiment we can distinguish three phases, labeled for convenience A, B, and C, and evolving as intuitively illustrated in figure 2.

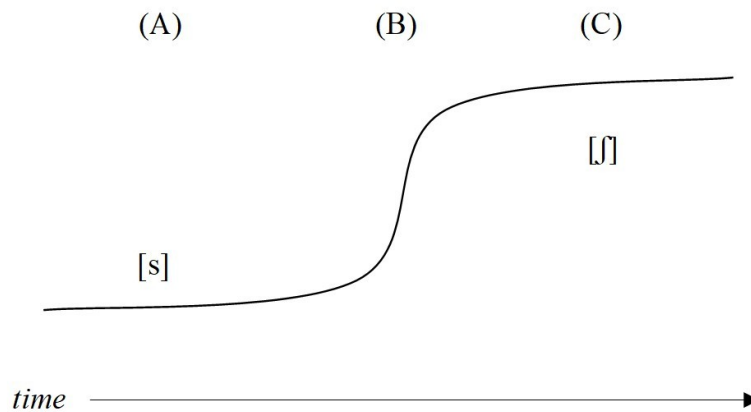


Figure 2

Phases A and C mark intervals where change in articulation and acoustic features correlate with a stable perceptual state. Phase B individuates a quantal region where small changes in the state of the articulators correspond to an abrupt change in phonetic category. The important point is that while in B small variations in the state of the articulators are relevant for determining whether the acoustic output belongs to [s] or [ʃ], throughout A and C phonetic typing remains constant across a wide range of articulatory configurations. For example, the acoustic output is stably classifiable as [s] for the whole temporal delta in which the tongue hits positions ranging from the front teeth to, approximately, the middle of the alveolar ridge. In such an interval, there are both change in observable acoustic properties (the counterpart of coloring and coarticulation in our example) and steadiness at the level of phonetic class.

Another example. We know that the primary acoustic cue for [d] in the syllable /di/ is a rising second-formant transition, while the [d] in the syllable /du/ is signaled by a falling second-formant transition (Lieberman et al. 1967). We also know that a single burst of noise at a frequency of 1440 Hz is heard as a [p] if followed by [i] and as a [k] if followed by [a] (Lieberman, Delattre and Cooper 1952). In the [d] case, there is an acoustic incongruity that is disregarded in phonetic interpretation.⁷ In the second example, there is an acoustic congruity (the burst at 1440 Hz) that somehow gives rise to a difference in phonetic categorization. But our definition of the membership criteria for [d] and, consequently, our notion of [d]-ness can be suitably complex. For example, the [d] case can be reconciled with Universality if the criteria for [d]-ness are tentatively defined as follows.

([d]-ness) $\forall x$, if $\{(rising\ second\ formant\ transition)x\}$ and $\{(followed\ by[i])x\}$ or $\{(falling\ second\ formant\ transition)x\}$ and $\{(followed\ by[u])x\}$, then [d]x.

In short, the fact that real-time phonetic categorization computes non-local parameters is far from obviously inconsistent with PP. The phenomenon does prove that the conditions that a micro-interval of speech must meet to be classified as an instance of a phone are sensitive to broader characteristics of the speech environment and respond to complex constraints (that we should try to uncover and formalize). But not that there are no possible versions of Universality clever enough to maintain the phone idealization while acknowledging the facts inspiring the formulation of the Locality Problem.

⁷ To be sure, the acoustic incongruity at work here is systematic, since the two second-formant transitions point to a common frequency locus. However, the point remains: the two consonants themselves have little in common in terms of observable low-level features. For more on the notion of “frequency locus”, see, e.g., Harrington (2010).

We can thus reiterate our initial methodological pledge. Variation in acoustic detail and non-locality effects do not seem to invalidate the basic tenet that speech sounds can be viewed as organized into a system of types reproduced by an appropriately large and appropriately sensitive inventory of phones. Speech consists of a continuum whose internal structure can be usefully modeled as a concatenation of segments, and such segments, in turn, can be usefully viewed as instances of elements from an appropriately fine-grained, cross-linguistically available inventory of phones.

5. Computationalism

We now turn to Anti-Computationalism. As we have seen, this horn of PE argues that there is no stage in real-time speech processing describable as the recovery of strings of discrete phones. Instead of viewing speech perception as a routine based on the manipulation of symbolic phonetic scores, real-time speech analysis should be modeled within a dynamic-embodied framework of the sort introduced by van Gelder (1995), Kelso (1995), and Clark (1997). For proponents of Anti-Computationalism, the key motivation for this claim is that speech contains gradient information that resists symbolic transduction while yet being causally relevant to the recovery of linguistic form. Contra PP, Anti-Computationalists maintain that the attractiveness and apparent naturalness of phone-based accounts of the cognitive representation of speech are a figment of the combined pressure of three factors: a) alphabetic literacy; b) the so-called “particulate principle”; c) phonological theory.

As for (a), the argument is that our putatively natural ability to segment speech into phones and the theoretical program of PP itself are byproducts of alphabetic training. In other words, literacy in alphabetic writing biases us into reasoning about speech processing in digital terms, and we simply reiterate the mistake in our efforts to produce scientific phonetics (Read et al. 1986; Faber 1993; Silverman 2006; Port 2007b; Cowley 2011).

As for (b), the argument is that virtually all work in mainstream formal linguistics runs on the implicit assumption that natural languages are complex systems arranging primitive particles on multiple levels of organization. It is no coincidence that the strongest rationale for the psychological reality of phones comes from the levels of linguistic representation posited to explain phenomena such as word-formation processes. Words are built from morpheme particles, morphemes are built from phoneme particles, so phonemes should be built from sound particles (e.g., Studdert-Kennedy 1998). Yet, according to advocates of Anti-Computationalism, this is just a systematic desideratum for which we have no real independent argument.

As for (c), the argument is that the issues surrounding the computational-symbolic nature of speech processing have always been framed in line with PP because phonetic representations are supposed to provide the input to phonological processes. Insofar as one accepts that phonemes are mentally real categories which pattern the phonetic space in classes of “lexically equivalent” phones, the psychological reality of phones seems to follow by formal necessity (Ohala 1992). But once again, belief in Computationalism is just a misleading consequence of the methodological desiderata (and of the ill-placed psychological ambitions) of formal phonology, not an attitude warranted by objective features of the language-speech interface itself (Archangeli and Pulleyblank 1994; Shockey 2003). The theoretical vocabulary of PP-style phonetics may well suit the descriptive constraints of mainstream phonology (and semantics, and syntax), but it is remote from the vocabulary one would need to talk about the actual properties of the mental representation of speech (van der Hulst 2003).

Let us try to respond. To start, Computationalism need not assume that the analog-to-digital conversion involved in the preparation of phonetic scores is a process which samples chunks of auditory representation, analyzes their content, and serially pairs them to a symbol in the phonetic string. The working assumption is that the reanalysis of the acoustic landscape *eventually* inputted to the phonological machinery can be characterized via the phone concept.

But nothing in the assumption is committed to conceiving the preparation of phonetic scores as a one-to-one mechanical process serially converting bits of auditory representations into phonetic symbols, with no sensitivity to suprasegmental or idiosyncratic features involved. As it turns out, most contemporary accounts reject the view that the recovery of phonetic representations is an either/or phenomenon where symbolic computation and non-symbolic processing are mutually exclusive (Luce and McLennan 2005). For example, it has been proposed that the properties of the speech signal that resist symbolic transduction might operate as attractors or constraints on the concatenation of phone-sized symbols eventually inputted to phonological calculus or, alternatively, be processed by separate non-symbolic units of processing running in parallel with the symbolic core of phonetic and phonological computation (Pisoni and Levi 2007; Hall 2015). In short, the evidence that non-symbolic factors play a role in early speech perception can be accounted for within frameworks according to which the backbone of early speech perception is symbolic and can be modeled on the basis of the phone concept.⁸

So far for the claim that early speech perception cannot be entirely reduced to the serial transduction of chunks of speech into phone-like mental symbols. As for the claim that early speech processing is completely unfit to be described as a phone-manipulating machinery, we have seen that according to prominent advocates of Anti-Computationalism real-time speech

8 The claim that symbolic and non-symbolic processing may coexist and cooperate in speech perception is not unheard of even within formal phonology. Approaches such as Bybee's (2001) frequency phonology, according to which the acoustic details of specific instances of speech are retained in phonological representations, appear able to incorporate the data emphasized by fans of Anti-Computationalism while remaining committed to the intervention of symbolic representations in speech processing. Pierrehumbert (2003) adds that although the conception of phonology as a machinery operating on phone-sized variables is often assumed to stand in opposition to the idea that the phonological grammar runs on statistical knowledge, this opposition is a spurious one, since probability theory assigns probability distributions to variables, and without variables (i.e., PP-style segments), there would be no way for a statistical learning model to tabulate any statistics about anything.

analysis should be modeled following a dynamically-oriented approach to cognition (van Gelder and Port 1995; Port and Leary 2005). There are two immediate issues here. The first is that the criticisms to symbol-processing models advanced by van Gelder (1995) and affiliates, however sound, do not generalize easily to the informationally complex domains of speech perception and production (see Markman and Dietrich 2000). The second issue is that a number of phenomena in speech perception and production appear extremely difficult to account for without some synthesis of feature properties at the phone level, and seem to require a mental machinery operating on phone-sized representations. This includes phoneme-restoration effects (Warren 1970), phoneme detection in non-words (Foss and Gernsbacher 1983), data on misperceptions of fluent speech involving single segments (Bond 2005), single-segment errors in speech production (Fromkin 1980), the perception of illusory segments inside illegal consonant clusters (Dupoux et al. 2011), and the existence of language games based on the insertion, deletion, or movement of a sound sequence (Vaux 2011).

Finally, the argument from alphabetic literacy: we find it so natural to conceive of speech processing as the recovery of a concatenation of phones only because we have been trained to read and write in a system involving the manipulation of alphabetic symbols, and our phonetic systems reproduce this pervasive illusion (Port 2007b). Now, it is true that our ability to report on the segmental structure of speech is greatly facilitated by literacy in alphabetic writing (Gillon 2004), and that subjects who do not master alphabetic orthography tend to be explicitly aware of- and able to manipulate only larger chunks of phonetic structure, such as its division in syllables (Morais, Alegria and Content 1987). The problem with the argument is that the evidence it relies on can be understood to support the opposite conclusion. For one thing, it seems that alphabetic systems could not have developed in the first place unless listeners, at some level, organized speech into minimal sound units. Otherwise, it would be natural to wonder why alphabetic writing is so widespread and has enjoyed such a far-reaching success in

the recent history of our species (Fowler 2010). Furthermore, available evidence about the difference in segmentation abilities between literates and non-literates in alphabetic systems seems to show only that alphabetism enhances preexisting segmentation abilities, not that it creates them *ex nihilo* (Brady and Shankweiler 2001). For example, Rozin and Gleitman (1977) present evidence that the development of the skills that allow children to analyze speech into minimal sound units is a prerequisite, rather than an effect, of the ability to read in alphabetic systems. Wan and Jaeger (1998) show that Mandarin speakers with little exposure to alphabetic writing produce speech errors with whole-segment deletions and metatheses. Similarly, Qu, Damian and Kazanina (2012) provide ERP data on Mandarin production giving support to the claim that letter-sized segments constitute fundamental units of phonological encoding even for speakers of languages that do not encode such units orthographically. It appears difficult to explain patterns of such a kind without assuming that some segmental organization of speech is cognitively available prior to alphabetic learning.

Once again: none of this proves once and for all that a completely reductive picture of the transition from physical speech to grammatical computation should be realist about an internalized phonetic transcriber which feeds strings of phones into phonology. We have no proof that our speech-grammar interface “actually runs on phones”. But what matters to Computationalism as a program in psychology is that PP generates empirically dependable and formally tractable approximations of the information flowing in the speech-grammar interface, not that concatenations of phones track the micro-ontology of the content feeding the transition between auditory and phonological processing. In other words, what matters is that we agree on the premise that by adopting the stance of PP we can generate viable characterizations of the content manipulated by the neurally implemented systems in charge of speech processing and production, even if we prefer to remain agnostic on the precise nature of the operations occurring at the underlying level, or, e.g., we believe that such operations are distributed patterns

of activation in a PDP substrate which has nothing to do with phones (Smolensky, Goldrick and Mathis 2014). The issue is whether conceiving of the speech-grammar interface as computing strings of phones is conducive to sound generalizations about the makeup of our linguistic machinery given its observable outputs (cf. Carr 2000 on “weak realism”), not whether concatenations of phones give us the “raw”, metaphysically unfiltered picture of how our mind translates physical speech into a format amenable to higher-level grammatical processing. As far as I can tell, this refined formulation of the psychological ambitions of PP remains motivated by a number of observable phenomena in speech processing, and none of the arrows in the anti-computationalist quiver threatens it. Hence, Computationalism stands.

6. Conclusion

The paper has proceeded as follows. Section 1 introduced the division of labor between phonetics and phonology, formulated the Phone Principle (PP), and stated the aim of the discussion: assessing the prospects of PP in light of eliminativist views of the phone concept. Section 2 outlined the three main horns of PP (Discreteness, Universality, and Computationalism), summarized the motivations appearing to call for Phone Eliminativism (PE), and formulated the eliminativist objection against each of the three horns of PP (Anti-Discreteness, Anti-Universality, and Anti-Computationalism). Section 3 tried to respond to the eliminativist argument against Discreteness. Section 4 tried to respond to the eliminativist argument against Universality. Section 5 tried to respond to the eliminativist argument against Computationalism.

The moral of the discussion is twofold. On the one hand, we should acknowledge the constructive role played by PE in reminding us that speech and speech processing are considerably more complex than one might think when looking at them through the lenses of a naive understanding of PP. Works interrogating the basic assumptions of linguistic phonetics are

rare, and the proponents of PE have certainly the merit of stirring up a fair debate on the conceptual foundations of field. Questioning the legitimacy of PP gives us a welcome opportunity to reflect on its ability to provide serviceable characterizations of the physical organization of speech, on the functions it serves in the broader landscape of linguistic theory, and on its relationship to the psychology of speech processing. On the other hand, I believe that the empirical evidence and the theoretical considerations gathered so far by the supporters of PE fail to show that the phone concept should be eliminated from linguistic theory. In other words, I believe we should be careful inferring, from the simple premise that a problematization of the phone concept is needed, the conclusion that linguistic phonetics should stop reasoning within the conceptual paradigm set by PP.

I hope this paper will contribute to raising awareness of the many foundational issues surfacing from contemporary phonetics, and will help stimulate critical inquiry into these areas among philosophers of science and philosophers interested in language more generally.

References

- Appelbaum, Irene. 1999. "The Dogma of Isomorphism: A Case Study from Speech Perception." *Philosophy of Science* 66: S250-S259.
- Archangeli, Diana, and Douglas Pulleyblank. 1994. *Grounded Phonology*. Cambridge, MA: MIT Press.
- Azzouni, Jody. 2010. *Talking About Nothing: Numbers, Hallucinations, and Fictions*. New York, NY: Oxford University Press.
- Beckman, Mary E., and Jan Edwards. 2010. "Generalizing Over Lexicons to Predict Consonant Mastery." *Laboratory Phonology* 1: 319-343.
- Bond, Zinny S. 2005. "Slips of the ear." In *The Handbook of Speech Perception*, edited by David B. Pisoni and Robert E. Remez, 290-310. Oxford: Blackwell.
- Bradlow, Ann R. 1995. "A Comparative Acoustic Study of English and Spanish Vowels." *Journal of the Acoustical Society of America* 97: 1916-1924.
- Brady, Susan A., and Donald P. Shankweiler, eds. 2001. *Phonological Processes in Literacy: A Tribute to Isabelle Y. Liberman*. Hillsdale, NJ: Erlbaum.
- Bromberger, Sylvain, and Morris Halle. 2000. "The Ontology of Phonology (Revised)." In *Phonological Knowledge: Conceptual and Empirical Issues*, edited by Noel Burton-Roberts, Philip Carr, and Gerard Docherty, 19-38. Oxford: Oxford University Press.
- Browman, Catherine P., and Louis Goldstein. 1992. "Articulatory Phonology: An Overview." *Phonetica* 49: 155-180.
- Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Caramazza, Alfonso, and Grace H. Yeni-Komshian. 1974. "Voice Onset Time in two French Dialects." *Journal of Phonetics* 2: 239-245.
- Carr, Philip. 2000. "Scientific Realism, Sociophonetic Variation and Innate Endowments in Phonology." In *Phonological Knowledge: Conceptual and Empirical Issues*, edited by Noel Burton-Roberts, Philip Carr, and Gerard Docherty, 67-104. Oxford: Oxford University Press.
- . "The Philosophy of Phonology". In *Philosophy of Linguistics*, edited by Ruth Kempson, Tim Fernando, and Nicholas Asher, 403-444. Amsterdam: Elsevier.
- Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. New York, NY: Harper and Row.
- Clark, Andy. 1997. *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.
- Cowley, Stephen J. 2011. "Taking a Language Stance." *Ecological Psychology* 23: 185-209.

- Davenport, Mike, and Stephen J. Hannahs. 2013. *Introducing Phonetics and Phonology*, 3rd edn. London: Routledge.
- Dretske, Fred. 1981. *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
- Dunbar, Ewan, and William J. Idsardi. 2010. "Review of D. Silverman, *Critical Introduction to Phonology: Of Sound, Mind, and Body*." *Phonology* 27: 325-331.
- Dupoux, Emmanuel, Erika Parlato, Sonia Frota, Yuki Hirose, and Sharon Peperkamp. 2011. "Where Do Illusory Vowels Come From?" *Journal of Memory and Language* 64: 199-210.
- Elgin, Catherine Z. 2017. *True Enough*. Cambridge, MA: MIT Press.
- Ereshefsky, Marc. 1998. "Species Pluralism and Anti-Realism." *Philosophy of Science* 65: 103-120.
- Faber, Alice. 1993. "Phonemic Segmentation as Epiphenomenon: Evidence from the History of Alphabetic Writing." In *The Linguistics of Literacy*, edited by Pamela A. Downing, Susan D. Lima, and Michael Noonan, 111-134. Amsterdam: Benjamins.
- Flege, James E., and James E. Hillenbrand. 1986. "Differential Use of Temporal Cues to the /s/-/z/ Contrast by Native and Non-Native Speakers of English." *Journal of the Acoustical Society of America* 79: 508-517.
- Fodor, Jerry A., Thomas G. Bever, and Merrill F. Garrett. 1972. *The Psychology of Language*. New York, NY: McGraw Hill.
- Foss, Donald J., and Morton A. Gernsbacher. 1983. "Cracking the Dual Code: Towards a Unitary Model of Phoneme Identification." *Journal of Verbal Learning and Verbal Behavior* 22: 609-632.
- Fox, Robert A., and Dale Terbeek. 1977. "Dental Flaps, Vowel Duration, and Rule Ordering in American English." *Journal of Phonetics* 5: 27-34.
- Fowler, Carol A. 2010. "The Reality of Phonological Forms: A reply to Port." *Language Sciences* 32: 56-59.
- . 2015. "The Segment in Articulatory Phonology." In *The Segment in Phonetics and Phonology*, edited by Eric Raimy and Charles E. Cairns, 23-43. London: Wiley.
- Fromkin, Victoria A, ed. 1980. *Errors in Linguistic Performance*. New York, NY: Academic Press.
- Gafos, Adamantios I. 2002. "A Grammar of Gestural Coordination." *Natural Language and Linguistic Theory* 20: 269-337.
- Gillon, Gail T. 2004. *Phonological Awareness: From Research to Practice*. New York: Guilford Press.
- Goldinger, Stephen D. 1996. "Words and Voices: Episodic Traces in Spoken Word Identification and Recognition Memory." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22: 1166-1183.
- Hale, Mark, and Charles Reiss. 2008. *The Phonological Enterprise*. Oxford: Oxford University Press.

- Hall, Kathleen C. 2015. "Categorical Segments, Probabilistic Models." In *The Segment in Phonetics and Phonology*, edited by Eric Raimy and Charles E. Cairns, 129-145. London: Wiley.
- Halle, Morris. 1954. "The Strategy of Phonemics." *Word* 10: 197-209.
- Harrington, Jonathan. 2010. "Acoustic Phonetics." In *The Handbook of the Phonetic Sciences*, edited by William J. Hardcastle, John Laver, and Fiona E. Gibbon, 2nd edn., 81-129. Malden, MA: Blackwell.
- Hockett, Charles F. 1955. *A Manual of Phonology*. Baltimore, IN: Indiana University Publications in Anthropology and Linguistics.
- . 1960. "The Origin of Speech." *Scientific American* 203: 89-97.
- IPA. 1999. *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.
- Jakobson, Roman, Gunnar M. Fant, and Morris Halle. 1952. *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates*. Cambridge, MA: MIT Press.
- Kelly, John, and John K. Local. 1986. "Long-Domain Resonance Patterns in English." In *Proceedings of the IEE Conference on Speech Input/Output*, 304-308. London: Institution of Electrical Engineers.
- Kelso, J. A. Scott. 1995. *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press.
- Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Oxford: Blackwell.
- Kingston, John, and Mary E. Beckman, eds. 1990. *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press.
- Klatt, Dennis H. 1979. "Speech Perception: A Model of Acoustic-Phonetic Analysis and Lexical Access." *Journal of Phonetics* 7: 279-312.
- Kleber, Felicitas, Tina John, and Jonathan Harrington. 2010. "The Implications for Speech Perception of Incomplete Neutralization of Final Devoicing in German." *Journal of Phonetics* 38: 185-196.
- Ladd, D. Robert. 2011. "Phonetics in Phonology." In *The Handbook of Phonological Theory*, edited by John Goldsmith, Jason Riggle, and Alan C. L. Yu, 2nd edn., 348-373. London: Wiley-Blackwell.
- Laver, John. 1994. *Principles of Phonetics*. Cambridge: Cambridge University Press.
- Liberman, Alvin M. 1996. *Speech: A Special Code*. Cambridge, MA: MIT Press.
- , Pierre C. Delattre, and Franklin S. Cooper. 1952. "The Role of Selected Stimulus-Variables in the Perception of the Unvoiced Stop Consonants." *The American Journal of Psychology* 65: 497-516.

- , Franklin S. Cooper, Donald P. Shankweiler, and Michael Studdert-Kennedy. 1967. “Perception of the Speech Code.” *Psychological Review* 74: 431-461.
- Lieberman, Philip, and Sheila E. Blumstein. 1988. *Speech Physiology, Speech Perception, and Acoustic Phonetics*. Cambridge, MA: Harvard University Press.
- Lodge, Ken. 2009. *Fundamental Concepts in Phonology: Sameness and Difference*. Edinburgh: Edinburgh University Press.
- Lombardi, Linda, ed. 2001. *Segmental Phonology in Optimality Theory: Constraints and Representations*. Cambridge: Cambridge University Press
- Luce, Paul A., and Conor T. McLennan. 2005. “Spoken Word Recognition: The Challenge of Variation.” In *The Handbook of Speech Perception*, edited by David B. Pisoni and Robert E. Remez, 591-609. Oxford: Blackwell.
- Markman, Arthur B., and Eric Dietrich. 2000. “In Defense of Representation.” *Cognitive Psychology* 40: 138-71.
- Morais, José, Jesus Alegria, and Alain Content. 1987. “The Relationships Between Segmental Analysis and Alphabetic Literacy: An Interactive View.” *Cahiers de Psychologie Cognitive* 7: 415-438.
- Ohala, John J. 1992. “The Segment: Primitive or Derived?” In *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, edited by Gerard J. Docherty and D. Robert Ladd, 166-183. Cambridge: Cambridge University Press.
- Pierrehumbert, Janet B. 2003. “Probabilistic Phonology: Discrimination and Robustness.” In *Probabilistic Linguistics*, edited by Rens Bod, Jennifer Hay, and Stefanie Jannedy, 177-228. Cambridge, MA: MIT Press.
- Pisoni, David B., and Susannah V. Levi. 2007. “Some Observations on Representations and Representational Specificity in Speech Perception and Spoken Word Recognition.” In *The Oxford Handbook of Psycholinguistics*, edited by Gareth Gaskell, 3-18. Oxford: Oxford University Press.
- Port, Robert F. 2007a. “How are Words Stored in Memory? Beyond Phones and Phonemes.” *New Ideas in Psychology* 25: 143-170.
- . 2007b. “The Graphical Basis of Phones and Phonemes.” In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by Ocke-Schwen Bohn and Murray J. Munro, 349-365. Amsterdam: Benjamins.
- . 2010. “Rich Memory and Distributed Phonology.” *Language Sciences* 32: 43-55.
- , and Adam P. Leary. 2005. “Against Formal Phonology.” *Language* 85: 927-964.
- Potochnik, Angela. 2017. *Idealization and the Aims of Science*. Chicago, IL: University of Chicago Press.

- Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. London: Blackwell.
- Qu, Qingqing, Markus Damian, and Nina Kazanina. 2012. "Sound-Sized Segments are Significant for Mandarin Speakers." *Proceedings of the National Academy of Sciences of the United States of America* 109: 14265-14270.
- Raimy, Eric, and Charles E. Cairns. 2015. "Introduction." In *The Segment in Phonetics and Phonology*, edited by Eric Raimy and Charles E. Cairns, 1-21. London: Wiley.
- Read, Charles, Zhang Yun-Fei, Nie Hong-Yin, and Ding Bao-Qing. 1986. "The Ability to Manipulate Speech Sounds Depends on Knowing Alphabetic Writing." *Cognition* 24: 31-44.
- Rozin, Paul, and Lila R. Gleitman. 1977. "The Structure and Acquisition of Reading II: The Reading Process and the Acquisition of the Alphabetic Principle." In *Toward a Psychology of Reading*, edited by Arthur S. Reber and Don L. Scarborough, 55-141. Hillsdale, NJ: Erlbaum.
- Scholz, Barbara C., Francis J. Pelletier, and Geoffrey K. Pullum. 2016. "Philosophy of Linguistics." In *The Stanford Encyclopedia of Philosophy (Winter 2016 Edition)*, edited by Edward N. Zalta. <https://plato.stanford.edu/archives/win2016/entries/linguistics/>
- Shockey, Linda. 2003. *Sound Patterns of Spoken English*. Oxford: Blackwell.
- Silverman, Daniel. 2006. *A Critical Introduction to Phonology: Of Sound, Mind, and Body*. London: Continuum.
- Smolensky, Paul, Matthew Goldrick, and Donald Mathis. 2014. "Optimization and Quantization in Gradient Symbol Systems: A Framework for Integrating the Continuous and the Discrete in Cognition." *Cognitive Science* 38: 1102-1138.
- Stevens, Kenneth N. 1989. "On the Quantal Nature of Speech." *Journal of Phonetics* 17: 3-46.
- Studdert-Kennedy, Michael. 1998. "The Particulate Origins of Language Generativity: From Syllable to Gesture." In *Approaches to the Evolution of Language*, edited by James R. Hurford, Michael Studdert-Kennedy, and Chris Knight, 202-221. Cambridge: Cambridge University Press.
- van der Hulst, Harry. 2003. "Cognitive Phonology." In *Germania et Alia. A Linguistic Webschrift for Hans den Besten on the Occasion of His 55th Birthday*, edited by Jan Koster and Henk Van Riemsdijk. <http://www.let.rug.nl/~koster/DenBesten/contents.htm>
- van Gelder, Timothy. 1995. "What Might Cognition Be, if not Computation?" *Journal of Philosophy* 92: 345-381.

- , and Robert F. Port. 1995. “It’s About Time: An Overview of the Dynamical Approach to Cognition.” In *Mind as Motion: Explorations in the Dynamics of Cognition*, edited by Robert F. Port and Timothy van Gelder, 1-44. Cambridge, MA: MIT Press.
- Varzi, Achille C. 2011. “Boundaries, Conventions, and Realism.” In *Carving Nature at Its Joints: Natural Kinds in Metaphysics and Science*, edited by Joseph K. Campbell, Michael O’Rourke, and Matthew H. Slater, 129-153. Cambridge, MA: MIT Press.
- Vaux, Bert. 2011. “Language Games.” In *The Handbook of Phonological Theory*, edited by John Goldsmith, Jason Riggle, and Alan C. L. Yu, 2nd edn., 722-750. London: Wiley-Blackwell.
- Wan, I-Ping, and Jeri Jaeger. 1998. “Speech Errors and the Representation of Tone in Mandarin Chinese.” *Phonology* 15: 417-461.
- Warren, Richard M. 1970. “Perceptual Restoration of Missing Speech Sounds.” *Science* 176: 392-393.
- Weisberg, Michael. 2013. *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.
- West, Paula. 1999. “Perception of Distributed Coarticulatory Properties of English /l/ and /r/.” *Journal of Phonetics* 27: 405-425.
- Wetzel, Linda. 2009. *Types and Tokens: On Abstract Objects*. Cambridge, MA: MIT Press.
- Wiese, Richard. 2000. *The Phonology of German*. Oxford: Oxford University Press.