Psychological Review
2016, Vol. 123, No. 2, 125–150

0033-295X/16/$12.00 http://dx.doi.org/10.1037/rev0000013

Perception of the Speech Code Revisited: Speech Is Alphabetic After All

Carol A. Fowler
University of Connecticut

Donald Shankweiler and Michael Studdert-Kennedy
University of Connecticut and Haskins Laboratories, New Haven, Connecticut

1932

We revisit an article, "Perception of the Speech Code" (PSC), published in this journal 50 years ago (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) and address one of its legacies concerning the status of phonetic segments, which persists in theories of speech today. In the perspective of PSC, segments both exist (in language as known) and do not exist (in articulation or the acoustic speech signal). Findings interpreted as showing that speech is not a sound alphabet, but, rather, phonemes are encoded in the signal, coupled with findings that listeners perceive articulation, led to the motor theory of speech perception, a highly controversial legacy of PSC. However, a second legacy, the paradoxical perspective on segments has been mostly unquestioned. We remove the paradox by offering an alternative supported by converging evidence that segments exist in language both as known and as used. We support the existence of segments in both language knowledge and in production by showing that phonetic segments are articulatory and dynamic and that coarticulation does not eliminate them. We show that segments leave an acoustic signature that listeners can track. This suggests that speech is well-adapted to public communication in facilitating, not creating a barrier to, exchange of language forms.

Keywords: speech perception, coarticulation, phonetic gestures, speech code, acoustic signal

Fifty years ago Liberman, Cooper, Shankweiler, & Studdert-Kennedy (1967) published an article in *Psychological Review*, "Perception of the Speech Code" (henceforth, PSC), that has received considerable attention both within the field of speech and outside of it and that is still cited frequently today.[1] The views on speech perception set out in PSC were motivated in part by the difficulties in finding a suitable acoustic output for a reading machine for the blind. (At that time, unless books were read aloud by a human, speech itself was not a possible output.) The starting point of that research was an assumption that speech sounds could be replaced by any system of discrete, sufficiently distinct signals. Findings discussed in PSC revealed that that assumption was false. This was apparent when subjects repeatedly failed to learn to perceive speech surrogates conveyed by any of a variety of sound alphabets (see also Shankweiler & Fowler, 2015). It was also inconsistent with the then recent discovery of the continuous nature of the speech signal itself based on spectrographic studies. In response to these findings and others suggesting that listeners' percepts conform more closely to speech articulations than to the acoustic patterning they cause (Liberman, Delattre, & Cooper, 1952; Liberman, Delattre, Cooper, & Gerstman, 1954) the authors of PSC proposed a motor theory of speech perception. This is the contribution for which the article is probably best known although only a few pages of PSC were devoted to it (see, Galantucci, Fowler, & Turvey, 2006, for a recent evaluation of the motor theory).

In this article, we focus instead on the main thesis of PSC, concerning the relation between elements of language form (phonetic segments or phonemes)[2] and the acoustic signal that provides listeners with information about them. This thesis is widely accepted among theorists whose perspectives on speech diverge widely both from each other and from that of the motor theorists of PSC; however, we will argue, it is wrong. Recognizing the respects in which it is wrong has important implications for understanding communication by means of speech.

Based on their spectrographic studies of speech and studies of the perception of speech synthesized from simplified spectrograms, the writers of PSC concluded that in the acoustic speech signal, ". . . the acoustic cues for successive phonemes are intermixed in the sound stream to such an extent that definable segments of sound do not correspond to segments at the phoneme level" (Liberman et al., 1967, p. 432). Rejecting that speech is a sound alphabet, they considered the acoustic speech signal to be a "code" on the consonants and vowels of the language that requires a special biological decoder. While we accept the findings discussed in PSC that the speech signal is not composed of static, discrete segments, we will propose that signatures of discrete, but temporally overlapping, segments *are* present in the signal (and can be identified if it is examined appropriately), and that segments can be perceived by listeners without the need for a special

This article was published Online First August 24, 2015.

Carol A. Fowler, Department of Psychology, University of Connecticut; Donald Shankweiler and Michael Studdert-Kennedy, Department of Psychology, University of Connecticut and Haskins Laboratories, New Haven, Connecticut.

Correspondence concerning this article should be addressed to Carol A. Fowler, Psychology Department, Box U-1020, University of Connecticut, Storrs, CT 06269. E-mail: carol.fowler@uconn.edu

[1] Scopus finds 314 citations from 2011 to the present (June, 2015).

[2] Our choice of *phonetic segment* or *phoneme* is not meant to be a theoretical claim, say, that the segments are not appropriately characterized in some other ways. We have just selected two useful terms. *Phoneme* refers to more abstract segments than *phonetic segment*, encompassing in English, for example, the more and less aspirated forms of voiceless stops.

decoder. As we will show, this discovery reinstates the long-abandoned view that speech is alphabetic and it leads to simplification of theory with potential practical benefits.

The conclusion of Liberman et al. (1967) that segments corresponding to phonemes cannot be found in the sound stream presents a paradox (Schane, 1973; Studdert-Kennedy, 1987) that persists in the field of speech research today. It is that, while phonemes are taken for granted as linguistic units in discussions of speech perception and reading by PSC and many others, researchers generally agree with Liberman, et al. (Diehl, Lotto, & Holt, 2004; Kleinschmidt & Jaeger, 2015; Magnuson & Nusbaum, 2007, among others) that segments do not exist in public manifestations of speech. Rather, the acoustic speech signal presents both a lack of invariance between phonetic segments and their acoustic correlates and an absence of acoustic segments corresponding to phonetic segments (respectively, the classic invariance and segmentation problems; Fant & Lindblom, 1961; Kleinschmidt & Jaeger, 2015; Perkell & Klatt, 1986; Pisoni, 1985). That segments exist in the mind, but not in public is paradoxical in itself, but it should be unexpected also because language forms, including phonetic segments, constitute the means within language for making intended communications public, and they constitute something that language learners need to learn about. Why would languages converge on forms that, thanks to the need to coarticulate, are not realized as such by speaking?

The paradox has created difficulties for explanations of language learning. Without resorting to the notion of innate ideas, it is difficult to explain, how segments can be components of language competence. If there is no detectable acoustic information that language has phonetic-segmental structure, where do the segments of language knowledge come from?[3] The question is often ignored by psycholinguistic researchers, including those who investigate children's acquisition of language (but see Braine, 1994; Lindblom, 2000; Studdert-Kennedy, 2000) even though, for most theorists, a belief in innate ideas of language forms is probably distasteful. Many researchers accept the linguist's basic segments of phonology, phonemes, and their features, in specifying the targets for language acquisition while at the same time accepting that these linguistic entities are not physically identifiable. We will propose evidence that speakers produce phonetic segments as individual or as coupled gestures of the vocal tract, that the gestures cause information in acoustic speech signals for the segmental structure of utterances, and that experienced listeners are sensitive to that information. If segments are, indeed, "out there" to be discovered, it would eliminate the paradox and clarify the process of perceiving language and learning it.

Besides creating difficulties for explaining language learning, the paradox muddies the explanation of why reading alphabetic writing is possible. The existence of alphabetic writing and the obvious fact that it is capable of being learned by most members of a language group is evidence for the reality of phonemic segments in language (see "The alphabet" below; but also see Port, 2010a, 2010b; Linell, 2005). However, as one of us noted, "Any discussion of the relation between speech and writing faces a paradox: the most widespread and efficient system of writing, the alphabet, exploits a unit of speech, the phoneme, for the physical reality of which we have no evidence" (Studdert-Kennedy, 1987, p. 68). This perspective poses a puzzle for a theory of reading. Arguably, there have been practical, educational consequences: The paradox has been exploited by some influential educators who oppose any teaching based on analysis of the internal structure of words (Goodman, 1986), a position that has worked to the detriment of children's learning to read (I. Y. Liberman & Liberman, 1990; National Reading Panel, 2000).

A final negative impact of the paradox is that its wide acceptance perpetuates the chasm between linguistics and speech science and, within linguistics, between phonology and phonetics, divisions that have long stymied cooperation and crosstalk between the several disciplines concerned with the sounds of language and their functions (Beckman & Kingston, 1990). Eliminating the paradox and restoring the alphabetic model for speech would remove a persisting reason why, despite the recognized importance of its subject matter, linguistics has remained an outlier among the sciences.

Perspectives on the Status of Phonetic Segments

In the following, we identify three perspectives on the status of phonetic segments that have proponents in the literature and then provide support for the view that we judge most plausible and most consistent with the evidence. It is that phonetic segments are real components of the language that underlie generativity in the lexicon; they are preserved in articulation, leave a perceivable signature in acoustic speech signals, and thereby can be perceived by listeners.

Perspective 1: Phonetic Segments or Phonemes Are Components of Language User's Knowledge of Language, But Not of Their Physical Implementations in Articulation Or the Acoustic Signal

For the authors of PSC, "No theory [of speech perception] can safely ignore the fact that phonemes are psychologically real" (Liberman et al., 1967, p. 452). Indeed, in most linguistic and psychological accounts, fundamental language forms are phonemes or phonetic segments. These are meaningless particles that compose word forms of the language. Following linguistic theories of the time (and, for the most part, of the present), Liberman, and colleagues (1967) viewed phonetic segments as discrete, static, context-free units that are specified by their featural attributes (e.g., consonantal place of articulation or voicing). In contrast, they had come to the view that articulation consists of dynamic actions in which there are no discrete, static units. Because articulations cause acoustic speech signals, speech signals likewise necessarily lack phonetic segmental structure.

MacNeilage and Ladefoged (1976) also represented this perspective explicitly. As noted, it is probably still the most commonly held one. After discussing the considerable context-

[3] It is notable that the computational models of phonetic-segmental category learning of which we are aware all begin with uncommitted segment categories in memory some of which acquire content in learning (Feldman, Griffiths, Goldwater, & Morgan, 2013; McMurray, Aslin, & Toscano, 2009; Vallabha, McClelland, Pons, Werker, & Amano, 2007). That is the models "know" that there *are* segments and categories of them in the language; their task is to identify them. However, unless segments and categories of them are innate ideas, infants have to *discover* that there are segments, have to find evidence of them in acoustic speech signals, and have to determine their groupings into categories (cf. Lindblom, 2000).

sensitivity of articulatory actions, MacNeilage and Ladefoged (1976, p. 90) remarked that:

> . . . this has led in turn to an increasing realization of the inappropriateness of conceptualizing the dynamic processes of articulation itself in terms of discrete, static, context-free categories, such as "phoneme" and "distinctive feature." This development does not mean that these linguistic categories should be abandoned. . . . Instead, it seems to require that they be recognized . . . as too abstract to characterize the actual behavior of articulators themselves.

MacNeilage and Ladefoged (1976) restricted their comments to speech production. However, an implication from their characterization is that, if phonemes are absent in articulation, specification of them must be absent in the acoustic speech signals they cause, and so, segments cannot be perceived directly from acoustic signals.

The "discovery" by Haskins researchers, alluded to above, that the speech signal is not alphabetic, led to an intensive search for context-sensitive acoustic "cues" for segments (beginning, e.g., with Liberman et al., 1952, 1954) that listeners might use to reconstruct segments in perception. Compatibly, on the basis of findings (Savin & Bever, 1970; Warren, 1971; but see McNeill & Lindig, 1973) that listeners identify phonemes more slowly than the spoken words or syllables in which they are embedded, Warren (1976) concluded that phonemes are not perceived directly, but, rather, can be derived by listeners if needed for a task such as phoneme identification. Also compatibly, Klatt (1979) proposed a model of lexical identification from acoustic speech input that mapped from spectra to words without first identifying phonetic segments.

A different, but, for present purposes, equivalent view, is that of Goldinger and Azuma (2003), who invoke Grossberg's ART (Adaptive Resonance Theory) model (Grossberg, 2003). They suggest that perceptual units emerge online in individual acts of perceiving. What units emerge in different perceptual settings may include conventional phonemes, but also and instead may include diphones, triphones, syllables, and more. Phonetic segments do not have primacy among emergent units.

Finally, this perspective on the status of phonetic segments is evident in a recent article by Kleinschmidt and Jaeger (2015), who address the issue of how listeners' flexibly adapt to the multiplicity of sometimes unfamiliar cues for phonetic categories that speakers having different accents, or speaking in different registers, rates, and so forth, produce. Category names (e.g., voiceless, bilabial, or stop) or identities (e.g., /p/) in memory remain stable as listeners learn new mappings of acoustic cues to them.

This perspective accepts the paradox of PSC as real. We present two alternatives next, and then devote the remainder of the article to a defense of the final one.

Perspective 2: Phonetic Segments or Phonemes Are Not Part of Language As Known or Used

Like the authors of PSC and other proponents of Perspective 1, and like many theorists, Faber (1992) rejects the existence of phonetic segments in articulation or the acoustic signal. Her grounds are those of MacNeilage and Ladefoged (1976): there are no steady-state, discrete segments in either domain. However, in contrast to the theorists who espouse Perspective 1, she argues that developments of phonological theory in linguistics (Clements, 1985; Sagey, 1986) have made segments less evident components of phonological competence as well.

In classical representations of segments (Chomsky & Halle, 1968; Gleason, 1955), they are distinct columns of feature values. For example /p/ is the name for a column of feature values that might include: +bilabial, −voice, +obstruent. Such a representation works well if segments are discrete, one from the other, static (in having the same feature values throughout) and context-free. One challenge to this idea comes from segments having a feature that starts out with one value and then switches to another while values of other features do not change. In prenasalized stops, for example, a [+nasal] feature becomes [−nasal] while place and manner features remain unchanged. Diphthongal vowels (e.g., the vowels in *boy* or *cow*) that change in quality throughout their extent also pose a challenge for feature-column representations of segments.

Many such challenges to the representation of segments as discrete feature columns led to development of alternative phonological approaches, for example, autosegmental phonology (Goldsmith, 1976) that focused on cases in which the individual features of a segment were found to have some measure of independence from one another and so sometimes to have different domains (like the nasality feature of prenasalized stops that occupies less than a whole feature column). van der Hulst and Smith (1982) provide examples in which some feature values span more than one segment. For example many languages exhibit vowel harmony, in which multiple vowels in a word share a feature value. In Hungarian vowel harmony, vowels in stems of words agree in the feature ±back, so that, in autosegmental accounts a single feature value, for example, ±back, spans two or more vowels.

In theories of autosegmental phonology, instead of occupying a place in a discrete feature column, features are represented on their own (autosegmental) tiers that link to other tiers by means of association lines. In such representations, segmental discreteness becomes less evident than in feature-column representations. On grounds such as these, Faber (1992) resolves the paradox of Liberman et al. (1967) by denying the existence of segments in language both as used and as known.

Port (2007, 2010a, 2010b) also denies the existence of phonemes in language generally but for different reasons than those of Faber (1992). As for language in public use, for example, he (Port, 2010a) presents spectrographic displays of the syllables /di/ and /du/, which reveal not only the lack of segmentation into vowel and consonant regions, but also the remarkable acoustic "restructuring" to which coarticulation gives rise so that the /d/s in the two syllables appear to have nothing acoustically in common.

Regarding language as known, he argues that (Port, 2010a, p. 44): "the segment-based, 'economical,' common-sense, low-bitrate view of linguistic memory is largely illusory. It is not the kind of memory people have." And (p. 45): "speakers do not know their language using a low-dimensional phonological code." Instead, their memories are high dimensional, richly detailed, episodic representations that include nonlinguistic as well as linguistic information.

Many research findings (Goldinger, 1998; Palmeri, Goldinger, & Pisoni, 1993) confirm that listeners have "episodic" or "exemplar" memories for speech events that are richly detailed, in including multiple kinds of information about a speech event, such

as voice characteristics of the speaker, the speaker's emotional tone and more. Port (2010a) interprets these findings as evidence that the memories for speech events are spectrotemporal, and suggests, referring to the spectrographic display of the syllables /di/ and /du/ just described that these syllables "probably do not share anything in actual memory" (p. 49).[4]

Port (2010a) also remarks that "phonological awareness, " the metalinguistic ability to manipulate phonological units of the language including phonemes, typically follows rather than precedes literacy in a spoken language (but see "Phonemic awareness," below). In short, in his reading, evidence suggests that discrete, abstract phoneme-sized segments are not components of language-users' memories for language.

Port (2010a) does not deny that some language users, most likely literate ones, may represent linguistic abstractions in memory, but he claims that these are not the memory systems that underlie their use of language as speakers or listeners. Instead (Port, 2010a, p. 45): "The low dimensional description that linguists call 'phonological structure' actually exists only as a set of statistical generalizations across the speech corpus of some community." Impressions that words are composed of discrete ordered segments are driven by literacy in alphabetic writing systems (Port, 2010b; see also Linell, 2005).

Perspective 3: Phonetic Segments Are Parts of Language As Known and Used

From this perspective, the one that we support henceforth, there are segments in language both as part of humans' capacity to use language and as speech is implemented in articulation. Acoustic signals provide sufficient information about phonetic segments for listeners to perceive them.

In the knowledge domain, we are largely in agreement[5] with PSC and MacNeilage and Ladefoged (1976), but largely in disagreement with Port (eg, 2010a), and Faber (1992). In the implementation domain, we are in disagreement with proponents of both Perspectives 1 and 2.

We comment briefly on the particular version of Perspective 3 that we will defend. We will make claims about the character of language forms (discrete phonetic gestures and linkages among them) that an experienced language user has developed the capacity to use, we will show that these forms serve as segmental primitives both of the phonological system of the native language and of temporally overlapped vocal tract actions in speech production. We will show that these actions causally structure acoustic signals in ways that preserve their discreteness, and we will show that listeners track the acoustic signatures of segmental speech actions in perception. We will not address how best to characterize the ways that experienced language users carry their relevant learning histories around with them to exploit what they have learned in those ways.[6]

In the next section, we argue for phonetic segments as components of the language that individuals know and have the capacity to use, whether or not they are literate in an alphabetic writing system. We show that phonetic segments as known are not merely abstractions floating above the richly detailed memories of the more contemplative of language users or generalizations across the "speech corpus" of a community of speakers (Port, 2010a). Rather, they function in individual acts of talking and individual acts of listening.

In subsequent sections of the article, we address the reality of segments in articulation, as specified in acoustic speech signals, and as perceivers may detect them. We suggest that, in articulation, segments are phonetic actions of the vocal tract ("phonetic gestures"; Browman & Goldstein, 1986; defined under "Gestures and synergies" below) or coordinated pairs or triads of phonetic gestures for multigestural segments.[7] Segments are articulatory and dynamic in nature; they are not the static entities suggested by a written transcription of what is said. Because gestures are dynamic events that overlap temporally, speech production researchers should not seek static, temporally discrete entities in articulation that serve as segments (cf. MacNeilage & Ladefoged, 1976, as quoted earlier). Instead, segments have a character that is shaped by capabilities of vocal tract action. Although gestures are not discrete along the time axis, they are separate and ordered actions. The acoustic signal can provide perceptible information for separate, ordered segments, because each segment is produced by distinct actions that wax and wane in their acoustic effects each in its own time frame. In a final section of the article, we summarize the most central conclusions of our arguments.

Segments in Language as Known (in "Competence"[8])

Liberman and coauthors of PSC did not question the reality of segments in the language itself as individual language users know it. However, because other researchers have done so as we have noted, we review nine lines of evidence converging on the view that they do exist in that domain. However, before turning to that evidence we offer a comment specifically on Port's contrary view.

[4] Leaving it mysterious for Port and others (Massaro, 1987, 1998) why they share a letter when they are spelled.

[5] The hedge ("largely") is because, as we make clear next, we do not claim that phonemes are representations in the heads of language users. Rather, individuals who know a language, say, English, have the capacity, however it is manifest in a brain, to produce and perceive segments in a native English-like way.

[6] We have already commented that language users preserve information about speech episodes, and we will show that they also can access information abstracted from particular autobiographical incidents. We will discuss effects of experience on procedural kinds of skills such as speech production, but also on metalinguistic awareness, mirroring, perhaps a distinction between procedural and declarative memory (Squire, 1986). However, we have no special proposals to offer about how language users preserve their prior histories so as to exploit their history in those ways.

[7] We acknowledge an irony. The theoretical perspective that we will present shares much with, and borrows from, that of our colleagues Browman and Goldstein (Browman & Goldstein, 1986, 1992) with an important exception. Browman and Goldstein (1990a) do not agree that phonetic gestures constitute phonetic segments or link with one another to form segmental units. In agreement with proponents of Perspective 2 above, they wrote (Browman & Goldstein, 1990a, p. 418): the basis for [phonetic segmental] units seems to be in their utility as a practical tool rather than in their correspondence to important informational units of the phonological system. We disagree (and ask why they have practical utility if they are not real), but we nonetheless borrow their important idea that phonetic gestures constitute primitives both of the language itself and of speech production.

[8] "Competence" is borrowed from Chomsky (1965) and refers to the knowledge that a language user has of the language that permits language use ("performance").

We dwell on his concerns, because we think that the confusions they represent are not unique to his point of view.

Our comment is to reject Port's argument that segments are disconfirmed because memories for speech are episodic in nature. As memories for episodes, they are richly detailed, and linguistic information is intertwined with nonlinguistic information, for example, about the talker's voice quality, emotional tone, and so forth (and even, weakly, about the talker's appearance; Sheffert & Fowler, 1995). As noted, there is compelling evidence for this (Goldinger, 1998; Palmeri, Goldinger, & Pisoni, 1993). Port (2010a, p. 43) infers from the evidence that: "people actually employ high-dimensional, spectro-temporal, auditory patterns to support speech production, speech perception and linguistic memory in real time."

However, the memories cannot be spectrotemporal. True, spectrotemporal representations (similar to spectrographic displays) are products of auditory processing to the extent that variation over time in energy concentrations in different parts of the spectrum has been tracked. However, *perception* is something quite different from this. In perception, listeners extract episode-specific information *from* these patterns *about:* the talker's utterance, voice quality, emotional tone, and rate of speaking (to provide a sampling), in short about a speech episode. That information is *potential* in a spectrotemporal display but has not been extracted from it.

That listeners do extract speech-episode information from auditorily processed acoustic signals is shown clearly in the literature that Port (2010a) cites. For example, consider the finding that listeners can correctly judge (with 90% accuracy) whether a word was presented previously in a recognition memory task with voice information preserved, but also (with 80% accuracy) without voice preservation (Palmeri, Goldinger, & Pisoni, 1993). In both conditions, the listeners perceived the words, but should not have if they had recourse only to a spectrotemporal display. In short, evidence that memory is episodic is orthogonal to the issue of whether or not the memories for words include phonetic segments.

Following are nine lines of evidence that converge on a conclusion that segments are real components of the language; moreover, some of those lines show that segments are part of individuals' capacity to use language (and so, are not only generalizations drawn over the language activities of a speech community; Port, 2010a).

The particulate principle: Segments as meaningless particles. The unbounded semantic scope of language rests on its combinatorial hierarchy of phonology and syntax. At the lower level, phonology evades the limits of a finite vocal apparatus by sampling, permuting, and combining a few discrete articulatory actions to construct an unbounded lexicon of words. At the higher level, syntax permutes and combines words to represent an infinity of relations among objects, events and concepts.

The principle of a combinatorial hierarchy is not unique to language. Both Jakobson (1970), a linguist, and Jacob (1977), a biologist, remarked on the isomorphism between verbal and genetic codes. Both also emphasized that the basic units of such systems must be devoid of meaning. In language only if phonetic units have no meaning[9] can they be commuted across contexts to form new words with new meanings. Jacob further observed that the hierarchical principle ". . . appears to operate in nature each time there is a question of generating a large diversity of structures using a restricted number of building blocks. Such a method of construction appears to be the only logical one" (Jacob (1977), p. 188).

Jacob did not spell out the logic, however. That was left to Abler (1989), who independently extended to other domains Fisher's (1930) argument concerning the discrete combinatorial (as opposed to blending) mechanisms of heredity. He recognized that a combinatorial hierarchy was a physically and mathematically necessary condition of all natural systems that "make infinite use of finite means," including physics, chemistry, genetics and language. He called it "the particulate principle of self-diversifying systems."

As one of us has previously written (Studdert-Kennedy, 2005, pp. 52–53):

> [T]he principle holds that all such systems necessarily display the following properties: (i) Discrete units drawn from a finite set of primitive elements (e.g., atoms, genes, phonetic segments) are repeatedly permuted and combined to yield larger units (e.g., molecules, proteins, syllables/words) above them in a hierarchy of levels of increasing complexity; (ii) at each higher level of the hierarchy . . . units have structures and functions beyond and more diverse than those of their constituents from below; (iii) units that combine into a . . . [higher] unit do not disappear or lose their integrity: they can reemerge or be recovered through mechanisms of physical, chemical, or genetic interaction, or, for language, through the mechanisms of human speech perception and language understanding.

In short, the particulate principle rationalizes the combinatorial hierarchy that standard linguistic theory takes as a language-specific axiom, and thus draws language within the domain of the natural sciences. In what follows we provide evidence for consonants and vowels as basic units of linguistic function, analogous to molecules, and in the following section ("Segments in articulation, in the acoustic signal, and in perception") for the gestures, that compose them as analogous to atoms.

The alphabet. Without the alphabet (or some other written mode of phonological analysis, such as a syllabary), linguists would have had no way of notating spoken utterances. Until the advent of X-rays, magnetic resonance imaging (MRI) and other modern methods of observing articulation, the alphabet was the only hard evidence, beyond intuition (and speech errors; see below), that we form meaningful words by repeatedly sampling, permuting and combining a small number of meaningless phonetic segments.[10]

The earliest forms of writing were syllabic (Gelb, 1963): a graphic symbol stood for the sound and meaning of a syllable. The first step toward recognizing that speech dissociates sound and

[9] In their typical function, phonetic units have no meaning. However, contrasting sounds may pair systematically with lexical and/or physical contrasts in meaning (Nygaard, 2010; Remez, Fellowes, Blumenthal, & Nagel, 2003). For example, people are slightly better than chance in identifying which of two antonyms in languages they do not know carries which meaning (Brown & Nuttall, 1959; Kunihira, 1971; Nygaard, 2010). These are weak statistical tendencies that do not challenge the requirement for their generative function that they be meaningless.

[10] Although we agree with Port (2010b) and Linell (2005) that literacy in an alphabetic writing system affects both knowledge of the phonology and intuitions about it, our claim here is that the relation is two-way. That is, recognition of the phonological structure of speech underlay invention of alphabetic writing systems, and (see "Phoneme awareness" below) its recognition by early readers facilitates learning to read.

meaning was discovery of the "rebus principle" (De Francis, 1989, p. 50). In rebus writing the symbol for a syllable with a certain sound (e.g., *sole*, fish) also stands for another syllable with the same sound, but a different meaning (*soul*, spirit), leaving the semantic ambiguity to be resolved by context. Isolation of the syllable as a unit of sound led, in due course, to recognition of its onset and coda in the Phoenician consonantal orthography of the second millennium B.C.E., from which all the world's alphabets ultimately derive. The Greeks later completed the alphabet by adding vowels (Gelb, 1963).

All full writing systems (De Francis, 1989, Chapter 2) represent the sounds of speech as either syllables or segments. No full writing system represents meaning directly without phonological mediation. Chinese, Chinese-influenced writing systems, such as Japanese, and a few African and Amerindian writing systems use syllabaries. All other written languages use alphabets. Use of a syllabary does not mean that the language cannot be written alphabetically. Indeed. both Chinese and Japanese have romanized alphabetic alternatives (*pinyin* and *romaji*, respectively) that are easier to learn and may eventually supersede traditional syllabaries. Moreover, with the International Phonetic Alphabet, consisting of about 160 symbols (letters and diacritics), a competent listener can transcribe and a competent reader can recover any utterance in any of the world's languages.

Given the universal reach of the alphabet, it seems unlikely that its symbols stand for fictional language forms.[11] The evidence of the alphabet added to the logic of the particulate principle demands physical definition of the phonetic segment. We will suggest that, contra PSC, speech and its perception is, in a real sense, alphabetic.

Phonemic awareness. For the most part, preliterate individuals, whether children (I. Y. Liberman, Shankweiler, Fischer, & Carter, 1974), or adults (Lukatela, Carello, Shankweiler, & Liberman, 1995; Morais, Cary, Alegria, & Bertelson, 1979) perform poorly when they are asked to count the phonemes in a spoken word, to add a phoneme, or delete it, or otherwise to manipulate spoken words at the level of phonemes. Ability to perform these tasks generally improves with acquisition of literacy in an alphabetic writing system. Lack of awareness of phonemes can mean either that in speech there are no segments to be aware of or that, for any of several possible reasons,[12] there are segments, but they are difficult focus attention on.

Despite these general findings, there are reports of preliterate children or adults unfamiliar with alphabetic writing who do well on tests of phonemic awareness (Lundberg, 1991; Mann, 1991). For example, Lundberg (1991) reports that of 51 nonreading children in a larger sample of preschoolers, nine performed perfectly on a phoneme segmentation task. (Moreover, the others could be taught to recognize segments using procedures that did not rely on printed materials.) Compatibly, Mann (1991) reviews findings from longitudinal studies of beginning readers, studies of skilled readers of nonalphabetic scripts, and of learners of word games who show phonemic awareness. For example, she cites Chao, (1931), who describes word games in Cantonese and Mandarin that involve shifts, additions, and substitutions of phonemes. Use of these games predates the development of pinyin, the alphabetic script for Chinese. Likewise, she cites McCarthy's (1982) description of a game in Hijaze Bedouin in which consonants in a triconsonantal root morpheme can be swapped with one another leaving interleaved vowel infixes and the word's prosodic template intact. Also compatibly, Mattingly (1987) refers to the oral tradition of morphological and phonological analysis of Sanskrit that preceded Panini (who wrote it down).

Related evidence comes from the oral verse of illiterate ancient Greek bards. Classical scholars generally agree that Homer (and others) wrote the Iliad and the Odyssey in the late 8th and early 7th centuries B.C.E. They also agree, since the work of Milman Parry (Parry, 1971), that Homer wrote in the dialect, style, and meter of a centuries old tradition of oral verse, composed and sung by professional bards (Finley, 2002).

Homeric verse is written in dactylic hexameters, lines of six metrical feet in a pattern of short and long syllables. The first four feet of the hexameter are either dactyls (long, short, short) or spondees (long, long), the fifth is a dactyl and the sixth is either a trochee (long, short) or a spondee. Syllable length is determined by the vowel. Ancient Greek has two long vowels (long e and long o) and five ambivalent vowels corresponding to English (a, e, i, o, and u). For the present discussion, it is enough to know that the length of the latter vowels in verse is largely determined by the number of immediately following consonants, if any, within a line. Broadly, they are short if followed by none or by one consonant, long if followed by two or more, either in the same word or distributed across the offset and onset of consecutive words. Relevant to Browman and Goldstein's claim (Browman and Goldstein, 1992) that basic phonetic units are gestures, not segments, a bigestural segment such as /m/ counts as one consonant, not two. In short, illiterate ancient Greek bards, and presumably at least some of their illiterate listeners, paid scrupulous attention to the pattern of consonants and vowels in spoken verse. Specifically, determination of ambivalent vowel length required counting consonants (zero or one vs. two or more).

Adult visual word recognition. That phonemes are not fictional for skilled readers of alphabetic writing systems is supported by the literature on skilled word recognition. This literature (e.g., see Frost, 1998, for a review) shows that readers access the pronounced forms of words even when that access is detrimental to performance. For example, in a Stroop procedure in which participants name the color of the ink in which printed letter strings are written, *grean* slows performance if the ink color is not green (Dennis & Newstead, 1981). Decisions whether a printed form is the name of a flower are slowed by homophones of flower names (e.g., ROWS) in research by van Orden (1987).

A variety of findings indicate that the pronounced forms of words accessed by readers have segmental structure. For example, Lukatela, Turvey and colleagues (L. B. Feldman & Turvey, 1983; Lukatela, Savic, Gligorjevic, Ognjenovic, & Turvey, 1978) studied word recognition by readers highly familiar with the two alphabets used to write Serbo-Croatian words. The two alphabets (Roman and Cyrillic) have letters in common, some of which are associated with the same and some with different phonemes. In that research line, lexical decision (i.e., the response time to decide whether a letter string is a word or not) is slowed if a letter in the printed

[11] However, see Faber (1992).

[12] For example, listeners are accustomed to focusing most of their attention on what a spoken communication means, not on its internal structure, perhaps because their knowledge of language is mostly procedural, not declarative.

word (or nonword) is associated with different pronunciations in the two alphabets (L. B. Feldman & Turvey, 1983; Lukatela et al., 1978). This can be despite the fact that all words and nonwords are presented to participants in just one of the two alphabets. That response time is slowed by a letter, say in a word printed in Roman, which has a different pronunciation in Cyrillic, shows that the bivalent letter is mapped to its two pronunciations during reading. A further study showed (L. B. Feldman & Turvey, 1983; L. B. Feldman, Kostic, Lukatela, & Turvey, 1983) that response time is slowed more by the presence of two such bivalent letters.

Likewise, Stone, Vanhoy, and Van Orden (1997) showed that lexical decision in English is slowed by both "feedforward" and "feedback" inconsistency (respectively, different ways to pronounce a spelling and, remarkably for visually presented stimuli, different ways of spelling a given pronunciation). Accordingly, time to decide that *mint* is a word is slowed by the existence, for example, of *pint*. Time to decide that *deep* is a word is slowed by the existence, for example, of *heap.* These findings show that the nature of subword letter-to-phoneme mappings routinely has an impact on word reading.

Systematic phonological and morphological processes. The phonological systems of languages exhibit systematic processes. Two examples from English are vowel lengthening before voiced obstruents (compare the vowels in *cap* and *cab*) and aspiration of voiceless stops in stressed syllable-initial position (compare the breathiness of the /k/s in *cab* and *scab*). Although we agree with Port (2010b) that these processes emerge in languages in the course of community-wide language use, even so, they are sustained by individual language users, and they are productive in individual language use. For example, were a new English word /kIb/ to be coined (by an individual language user), it would be [k^hIb] (with an aspirated initial /k/), not [kIb] (with unaspirated /k/).

Some phonological processes involve manipulation of individual phonemes. In relevant examples of metathesis, phonemes swap places. In some cases of epenthesis or syncope, whole phonemes are inserted or deleted, respectively. Two of these processes, phoneme syncope and phoneme metathesis, are exemplified by Hanunoo, a language of the Philippines (Gleason, 1955, as presented in Kenstowicz & Kisseberth, 1979). In this language, *two* (*duwa*) becomes *twice* (*kadwa*) by inserting *ka* and deleting *u* (a phoneme in *duwa*, not a syllable). *Three* (*tulu*) becomes *thrice* (*katlu*) in the same way. Metathesis occurs when *four* (*ʔupat*) becomes *four times* (*kapʔat*). The glottal stop (/ʔ/) swaps places (metathesizes) with a following consonant, /p/. In these examples, the units involved in productive phonological processes (/u/ deletion, /ʔ/-/p/ metathesis) are phonemic.

Some morphological processes likewise involve productive manipulation of individual consonants and vowels. An example comes from Semitic languages. In Classical Arabic (McCarthy, 1982), verb root morphemes are triconsonantal, for example, *ktb*, referring to writing. They can be inflected or derived by inserting vocalic morphemes in accordance with a prosodic template. The perfective active form of the verb is *katab*, with insertion of the morpheme *a* into *ktb* in accordance with a template CVCVC. The perfective passive is *kutib* with insertion of *ui* into the same template. Insertion of *a* into *ktb* via a template CVCCVC gives *kattab* (meaning *cause to write*); into CVVCVC gives *kaatab*, *correspond.*

Because all of the foregoing processes are productive in applying to newly coined words or to existing words produced or understood by a speaker-hearer for the first time, they provide evidence that individual language users have a functional capacity to manipulate individual consonants and vowels.

Speech errors. There is evidence provided by a variety of experimental paradigms (see, Levelt, Roelofs, & Meyer, 1999, for a review; also see Meyer, 1991) and modeling (Levelt et al., 1999; Roelofs, 2014) for a role of phonemes or phonetic segments in planning for speech production. Here we focus only on speech errors, which provide a sufficiently compelling case.

We first characterize speech errors as they have been described in transcription-based studies and then turn to complexities that reinterpretations of that evidence (Browman & Goldstein, 1990a) and that laboratory studies of errors (Mowrey & MacKay, 1990; Pouplier, 2003) have introduced. We argue that none of the complexities substantially challenges the observations from earlier studies that phonemes are units of speech planning,

Spontaneous errors of speech production occur when a talker intends to produce an utterance, and is capable of producing it, but instead produces something else. Prominent kinds of errors are noncontextual substitutions in which one language form replaces another that was not part of an intended utterance (e.g., intending to say *summer*, but saying, *winter* instead or intending to say *material*, but saying *maserial* instead[13]) and movement errors, in which intended forms appear in unintended locations. The latter occur as anticipations, perseverations, and exchanges. An example of a word perseveration is *I want to use that book as a bookstop* (intended: . . . as a doorstop); a segment exchange is *Kotcher kishen* for intended *Kosher kitchen*. Because the units involved in these errors are inserted or moved on their own, an inference has been drawn (Dell, 1986; Garrett, 1980) that they are discrete units of language planning. Notably for present purposes, only some imaginable language units participate in errors with any frequency. Those that do are words, morphemes, and phonemes (with morphemes participating in error patterns somewhat different from those of words and phonemes). Syllable errors occur rarely. Errors that are unambiguously feature errors (Fromkin's (1973) *glear plue sky* for *clear blue sky*) are also rare.

These findings suggest that phonemes (and words and morphemes) are units of speech production planning that can be misselected in substitution errors and mislocated in movement errors. If the findings were unchallenged, speech errors would constitute strong evidence for the reality of segments as functional units in language. However, they are not unchallenged.

Before addressing the challenges, we make an observation about transcribed phoneme errors that we find compelling. It is that phonemes and whole words participate in the same kinds of errors: noncontextual substitutions, anticipations, perseverations, and exchanges. Whereas single segments are subject to mishearing and other sources of bias, whole words are less so. If the whole-word data (e.g., *We have a laboratory in our own computer* for *We have a computer in our own laboratory*) can be trusted as revealing words as units, most likely so can phoneme data, despite the

[13] Errors reported here either occur in publications (Dell, 1986; Fromkin, 1973; Garrett, 1980; Shattuck-Hufnagel, 1983) or else in a corpus collected by the first author and students.

greater likelihoods of mishearing and bias, be trusted as revealing segments as units.

A challenge to the interpretability of transcription-based corpora derives from the fact that findings are based on errors obtained outside the laboratory by individuals who write down what they hear. As such, they are subject to biases of many kinds. Some errors can be missed, because they are so subtle as to be inaudible, listeners are known to repair errors perceptually (Marslen-Wilson & Welsh, 1978), and they have a lexical bias, that is, a tendency to report hearing real words rather than nonwords (Ganong, 1980). Additionally, some errors that are audible may not be transcribed because error collectors do not know how to write them down. This last source of bias may inflate the impression that speech errors are tidy in the units they involve.

In recognition of these sources of bias, researchers have brought errors into the laboratory (Dell, 1986; Motley & Baars, 1976; Mowrey & MacKay, 1990; Pouplier, 2003, 2007; Pouplier, Chen, Goldstein, & Byrd, 1999). Of course, procedures employed to induce errors in the laboratory may lead to errors with properties that are not identical to those that occur in spontaneous speech. However, an advantage is that induced errors can be recorded acoustically, and, in some cases, speakers' muscle activity or articulations have been recorded. Important findings from laboratory research have been that subphonemic errors are common (Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007; Mowrey & MacKay, 1990: Pouplier, 2003) although gestures that are coupled into nasal segments (the only multigesture segments examined) move together in errors with greater than chance likelihood (Goldstein et al., 2007). In addition errors are gradient in magnitude, many are inaudible, and many violate phonotactic legality (Goldstein et al., 2007; Pouplier, 2003), a rarity in transcribed errors.

In addition to these findings is an important assessment of transcribed-error corpora by Browman and Goldstein (1990a). They remark that phonemes involved in movement and substitution errors (e.g., *shocks and shoes* for *socks and shoes; just finished mating it* for *just finished making it*) are typically featurally similar. This makes such errors ambiguous as to whether what moves or substitutes is a phoneme or a feature. Although Shattuck-Hufnagel and Klatt (1979) did purport to show that exchange errors that ostensibly involve consonants, in fact do involve consonants, not features, Browman and Goldstein (1990a) dismiss their demonstration on grounds that the majority of consonant exchanges occur in word onsets. Possibly these are the relevant chunks. In favor of Shattuck-Hufnagel and Klatt's conclusion (1979); however, is that errors like *glear plue sky* that are unambiguously feature errors are rare. Second, not all segments involved in exchanges *are* featurally similar (e.g., *heft lemisphere*). Third, Nooteboom and Quene′ (2014) find that, taking opportunities to make (phonotacticallly permissible) errors into account, word onsets are no more likely to be involved in slips than are vowels or consonants in medial and final word positions. Accordingly, ostensible phoneme errors in word onsets are probably genuine phoneme errors just as they are more unambiguously elsewhere in a word.

To address whether literacy has fostered extraction of phonemes from continuous speech, it would be helpful to know whether illiterate people make phoneme errors, unbiased by knowledge of alphabetic units. Until relatively recently no one apparently thought to ask, despite a history of studies investigating the metalinguistic abilities of illiterate adults to manipulate phonemes (the phoneme awareness studies discussed earlier). On one view the expectation is no; phonemic organization of speech is considered to be a result of alphabetic literacy (A. E. Fowler, 1991). In that event speech errors of illiterates should manifest a coarser segmentation than those of literate people. Fortunately, Castro-Caldas and associates (Castro-Caldas, Petersson, Reis, Stone-Elander, & Ingvar, 1998) have obtained data that allow at least a provisional answer. They studied errors in repetition of lists of spoken trisyllabic words and pseudowords by illiterate villagers in Portugal comparing them with modestly literate people from the same rural villages. There were two principal findings: Illiterates made more errors overall than literates but chiefly on pseudowords, for which meanings were not available to reinforce recall of phonological structure. Second, the speech errors patterned similarly in the two groups; some of the repetitions in each group were single phoneme substitutions (eplara going to eflara; lipalio to lifalio). That the illiterates were less accurate in repeating pseudowords than the literates does require explanation, but whatever that explanation may be, it can be concluded that speech productions of both literate and illiterate people on this task were (on at least some occasions) phonemically organized.

In summary, some findings from speech errors support the reality of segments as units of speech planning. Although this evidence has been challenged, and laboratory investigations of errors has added new information about error properties, in our view, none of the evidence challenges that segment-sized language forms are components of planning for talk.

Neighborhood density and probabilistic phonotactics. Pierrehumbert (2006) offers a variety of reasons for concluding that language users develop both episodic memories and a level of segmental structure abstracted from them. Some of her reasons in favor of segments in memory overlap with some of our reasons already provided (e.g., speech errors, the particulate principle[14]). Another, however, is a finding of Vitevich and colleagues (e.g., see Vitevitch, Luce, Pisoni, & Auer, 1999, for a review). Words and nonwords can differ in respect to their probabilistic phonotactics, that is, the frequency with which a segment or sequence of segments occurs in different positions in a word across the words of the language. Nonwords created to include high probability sequences are rated more word-like and are named (repeated) faster than nonwords with low probability sequences (Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997). For their part, words can be classified by their "neighborhood density," that is, the number of words phonologically similar to them in memory. In a variety of tasks (word identification in noise, lexical decision and auditory naming (Experiments 1–3 of Luce & Pisoni, 1998) words in high density neighborhoods are responded to less accurately and more slowly than those in low density neighborhoods. A puzzle is that high probability phonotactics, which helps performance on nonwords, and high neighborhood density, which slows performance on words, are highly correlated, because phonotactics are highly probable when many words share them. To explain how nonword

[14] Pierrehumbert (2006) refers instead to the "phonological principle" that captures the notion of generative recombination of meaningless language forms, but not the commonality among phonology, physics, genetics, and chemical compounding that is central to the particulate principle.

processing can be facilitated by a variable highly correlated with one that slows word processing in the same task of auditory naming appears to require that a distinction be made between the levels at which the effects arise, lexical (episodic) and sublexical (phonemic).

Historical sound change. Another argument of Pierrehumbert (2006) in favor of a segmental level in language use is historical sound change, the continual changes in how words are pronounced by members of language communities. Although some sound changes may be lexically gradual, occurring earlier in high than low frequency words, "historical change does not have the character of random drifts of the pronunciation patterns for individual words " (Pierrehumbert, 2006, p. 522). Rather, pronunciation of the same segment undergoes the same kind of shift in different words. If change did not occur in this way, as she comments, the phonological principle (or, as we prefer, the particulate principle) would gradually disappear in language.

Experimental sound change: Perceptual learning. It has been proposed that there is better evidence for segments as planning units in speech production than for segments as units extracted in perception (Dell, 2014; Hickok, 2014). This may be true, but there is nonetheless sufficient information that listeners extract information about phonetic segments in words they identify. In a sense, the evidence is like that of historical sound change just discussed. It is that perceptual learning about a segment, say an unfamiliar variant of [f], provided by exposures to a subset of words in which that segment occurs, generalizes to other contexts, as if, for listeners, it is the same [f] everywhere.

In research by Norris, McQueen, and Cutler (2003; McQueen, Cutler, & Norris, 2006), Dutch listeners were exposed to words ending in a sound ambiguous between [f] and [s]. For some participants, the ambiguous sounds were the final consonants of words that end in [f] (in English, say, *rebuff).* These listeners also heard other words ending in clear [s]. For other participants the ambiguous fricative ended [s]-final words (say, *remiss*), and these listeners also heard words ending in clear [f]. These exposure trials should lead learners to retune what counts as [f] or [s] at least in the words to which they were exposed. If retuning occurs, and the finding generalizes to [f] and [s] in nonword contexts or in new words, there is evidence for segments as entities in memory that are retunable and are accessed in perception. In one study (Norris et al., 2003), following exposure to these words, listeners identified consonants along an [ɛf] to [ɛs] continuum. Boundary shifts reflected listeners' exposure experience. Those exposed to [s]-like [f]s identified more continuum members as [f]; the other group showed the opposite shift. McQueen et al. (2006) showed generalization to new [f]- and [s]-final words, that is, other familiar words to which they had not been exposed in the learning phase of the experiment.

Summary

We have summarized nine kinds of evidence that converge on a claim that the phonologies of languages have phonemic structure for individual language users. For the authors of PSC, this is uncontroversial, and did not require the justifications we have offered. However, as noted, the claim has been repeatedly challenged (Browman & Goldstein, 1990a; Faber, 1992; Port, 2010b). Addressing that issue is a necessary precursor for going on, as we do next, to evidence converging on the idea that phonetic segmental units of the language can be manifest in articulation. When they are, they can be specified acoustically, and, therefore, as we will show, can be perceived by listeners.[15]

Segments in Articulation, in the Acoustic Signal, and in Perception

In what follows we turn to speech production and perception. However, we begin by providing a conceptualization of segments as known that departs from that represented in PSC and in most linguistic and psycholinguistic descriptions. We suggest that consonants and vowels as *known* are articulatory.[16] Next we show that they are identifiable units in speech articulation, they have signatures in acoustic speech signals, and they can be perceived by listeners, who detect those signatures. Indeed, research has shown that listeners can track the acoustic consequences of distinct but temporally overlapping gestures.

The Character of Segments as Known Is Articulatory

As noted earlier (pp. 8–9), MacNeilage and Ladefoged (1976) reject that phonemes exist in articulation on grounds that the "discrete, static, context-free" segments of phonological descriptions cannot be found there (cf. Faber, 1992). The authors of PSC characterized coarticulation as introducing both spatial and temporal sources of context sensitivity that likewise eliminated the discreteness and context-free nature of phonetic segments as known. It is true that discrete, static, context-free segments are absent in articulation of speech. However, Browman and Goldstein (1986) proposed a phonological theory in which primitives of the system are not the discrete, static, context-free elements assumed in PSC or by MacNeilage and Ladefoged (1976) or, for that matter, by phonologists generally (Chomsky & Halle, 1968; Gleason, 1955; Prince & Smolensky, 2008). Rather, primitives of the language itself are phonetic gestures that create and release constrictions in the vocal tract. They are discrete in being separate constriction-release actions, they are dynamic, and they are sufficiently context-free to preserve their identity over phonetic contexts. We elaborate these characterizations below ("Speech production").

Although articulatory phonology is a linguistic theory, Browman, Goldstein, and colleagues implemented it as a computational model as one way of testing the viability of their theoretical account. One version of the model (Browman & Goldstein, 1990b) is shown in Figure 1A. Given an intended utterance, the linguistic gestural component of the model computes a "gestural score"

[15] We have made somewhat weaker claims than we might have. We do not necessarily claim that talkers always produce words as sequences of their component phonemes. For example, if someone answers a question such as *What are you doing tonight?* by producing a continuous vocalic utterance with an appropriate fundamental frequency contour to mean, *I do not kno*w, there is no phonemic structure, phonetic segments will not be specified acoustically, and will not be perceived. We restrict our attention here to careful speech (Lindblom, 1990).

[16] In PSC, the authors proposed that there is a 1:1 correspondence between linguistic features and commands to muscles (see their Figure 5 and discussion of it). Therefore, there is a sense in which consonants and vowels were articulatory in that version of the motor theory.

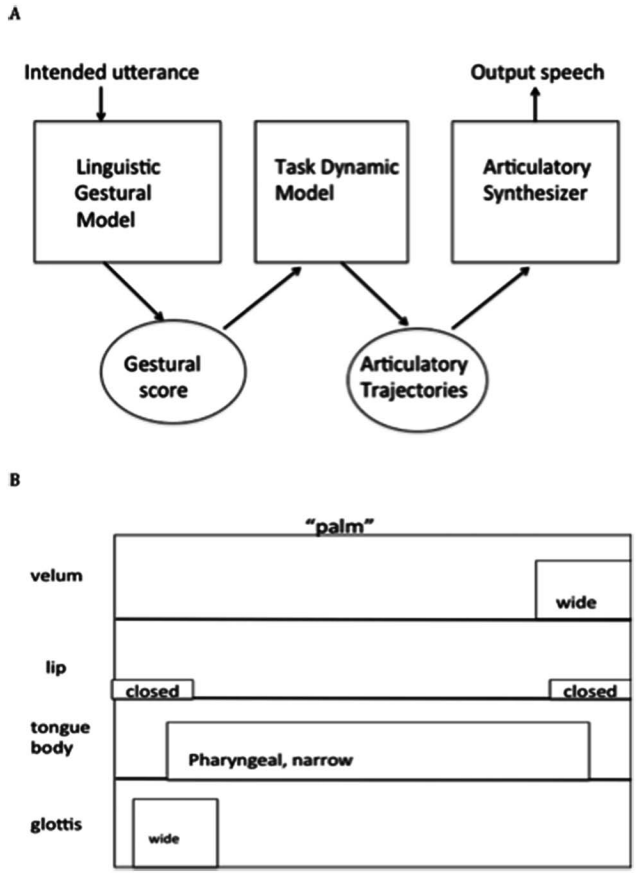

Figure 1. Panel A: Representation of the computational model of Browman and Goldstein and colleagues (redrawn from Browman & Goldstein, 1990b). The model takes an intended utterance as input, and using generalizations about component gestures and their phasing in the language (in the Linguistic Gestural Model), generates a gestural score (Panel B, redrawn from Browman & Goldstein, 1990b; reproduced with permission), a depiction of the component gestures of the utterances, including their constriction locations and degrees, and the gestures' phasings. (Time is represented along the horizontal dimension of the gestural score.) This serves as input to the task dynamic model of Saltzman and colleagues (Saltzman & Munhall, 1989), which generates dynamical systems for each gesture. The systems include the relevant articulators and the coordinative relations that will achieve required constriction locations and degrees in an equifinal way. The trajectories of articulators that this model outputs is input to an articulatory synthesizer that produces an acoustic speech output. Adapted from "Specification Using Dynamically-Defined Articulatory Structures," by C. Browman and L. Goldstein, 1990, *Journal of Phonetics, 18*, pp. 299–320. Copyright 1990 by Elsevier.

based on generalizations about gestural phasing in utterances of the language. In a gestural score (Figure 1B), phonetic gestures are temporally overlapped in the systematic ways that the language being modeled manifests temporal overlap. The gestural score is input to Saltzman and Munhall's (1989) "task dynamic model" of speech production. This model generates planned gestures of the gestural score as successively activated coordinated actions, themselves products of dynamical systems (or synergies) that achieve the gestures' associated constriction locations and degrees in an equifinal way (as described below under "Speech production.") These planned actions drive an articulatory synthesizer (Rubin, Baer, & Mermelstein, 1981) that produces an acoustic speech signal.

Can dynamic phonological primitives substitute for the static forms of other phonological theories in a viable phonological theory? They do in at least three ways. First, articulatory phonology provides an alternative proposal about how phonological word forms are known as lexical items, a goal of many phonological theories, beginning with descriptive linguistic theories (Gleason, 1955; Trager & Smith, 1951). Second, it offers an account of "contrast," also a major goal of descriptive linguistic phonological theories (Gleason, 1955; Trager & Smith, 1951). In those approaches, to find the set of phonemes of a language, linguists sought "minimal pairs" of words that had different meanings (were not just different variants of the same word) and differed in just one segment (e.g., *pat*, *bat*, in English). These different segments were identified as different phonemes in the language. Third, articulatory phonology offers an account of systematic phonological processes that languages exhibit (e.g., longer vowels before voiced than voiceless obstruents in English, *cab* vs. *cap*). Finding illuminating characterizations of these processes has been a major focus of attention of many generative phonologies (Chomsky & Halle, 1968; Prince & Smolensky, 2008).

As for word representations, a sample gestural score is shown in Figure 1B. Gestural scores can replace the arrays of feature columns of descriptive linguists or of earlier generative approaches or can replace the more complex geometric structures of autosegmental approaches (Goldsmith, 1976) and others (Clements, 1985). Components of gestural scores are phonetic gestures that are inherently dynamic and overlap temporally.

Regarding contrast, Browman and Goldstein (1992) showed that a phonology with dynamic primitives can serve that function. For example, the words *pat* and *bat* are "minimal pairs," of words that differ in meaning and differ by just one gesture, a devoicing gesture that is present only in *pat*.

Other phonological theories (e.g., that of Chomsky & Halle, 1968; and optimality theories, e.g., Prince & Smolensky, 2008) have shifted attention away from contrast and focused instead on systematic phonological processes (such as vowel lengthening in English before voiced obstruents or processes of metathesis, syncope and epenthesis as discussed above, p. 26). In regard to those processes, Gafos (1999, 2002, Gafos & Benus, 2006) has shown that Browman and Goldstein's gestural primitives provide a better account of some of them than do the discrete, static segments of other phonologies.

For example (Gafos, 2002), he showed for Moroccan Arabic that some phonological processes require reference to timing, which is absent from the static phonological primitives in theories other than that of Browman and Goldstein. In other research (Gafos, 1999), he has shown that the relative frequency of two phonological processes, vowel harmony and consonant harmony, and the characteristics of consonant harmony when it occurs, parallel the general viability of their likely articulatory sources, vowel-vowel coarticulation across a consonant or consonants and consonant-consonant coarticulation across a vowel, respectively. This is as expected if phonological primitives and processes are articulatory in nature. Finally, Gafos and Benus (2006) showed that phonological processes can have graded effects as in final devoicing in German. This is unexpected from the perspective of

other phonological theories where phonological processes either apply or do not, but it is expected from the gestural perspective of articulatory phonology (cf. the gradience of speech errors in the research by Pouplier, 2003, and others summarized earlier).

The force of the preceding discussion is that a shift in theoretical perspective about segments of the phonology itself is required to understand how segments can be preserved in speech. The shift is to recognize that consonants and vowels *as linguistic entities* are dynamic and articulatory in nature. They are adapted to their use in between-person speech communications.

Speech Production

In this section, we show how it is possible to understand speech articulation as composed of serially ordered segments despite the occurrence of coarticulation. In the preceding section, we made one important move in this direction in arguing that phonetic segments are fundamentally dynamic-articulatory, not static, in nature. Accordingly, theorists should not look at speech production as, ideally, achievement of successive target vocal-tract *configurations* each associated with a phonetic segment, with intervening transitions between them being physically required, but otherwise unwanted. Rather, they should look at speech production as successions of overlapping *action*s, each action appropriate for a phonetic segment or for one of its gestures.

Such a view corresponds quite closely to that of the authors of PSC, who were endeavoring to understand how speech could be routinely produced and perceived at rates of 10–15 phonemes/second, an order of magnitude higher than the most efficient nonspeech acoustic cipher, Morse code. Morse code rates of perception (and so of production), with an average of about three acoustic pulses/phoneme are limited by the temporal resolving power of the ear for which discrete pulses at a rate of 20/second merge into a low-pitched buzz. The solution of PSC for speech production was parallel processing of successive phonemes, packaging them into syllables by spatiotemporal distribution of their constituent "features" (in our terms, gestures). Phonemes "are taken apart into their constituent features and the features belonging to successive phonemes are overlapped in time. Thus, the load is divided among the largely independent components of the articulatory system" (PSC, p. 455.) In short, PSCs account of production was, in our view, broadly correct. What it lacked was a conception of gestures as phonemic primitives rather than features and a coherent theory of speech action in which gestures of a segment can cohere despite not being produced synchronously. We turn to that account next.

Here, we simplify the picture of production in several ways as we try to maximize clarity by focusing on central observations. In what follows, we will pretend sometimes that segments are composed of single gestures. This is true in English for voiced stops and fricatives and many vowels in Browman and Goldstein's articulatory phonology (Browman & Goldstein, 1986, 1992, 2000) on which we will rely, but it is not true of voiceless segments, nasal stops, rounded vowels, or /l/, /w/, /r/. Under "Intergestural links," we will address the evidence that some complexes of gestures form segments via linkages among the gestures, but we acknowledge that there is little relevant articulatory evidence. Motivation for making the effort in this direction is provided by the nine converging lines of evidence summarized earlier and the theoretical commitment to identify segments as known and as produced as the same, if possible.

We will further simplify the discussion of gestures by focusing on the main articulatory actions that compose them, ignoring the finer detail in their production. We will argue that the best way to understand coarticulation is as temporal overlap of gestures for segments rather than as context-sensitive changes in them. However, we acknowledge that, in casual styles of speaking especially, context-sensitivity looms large. We will generally ignore casual speech registers in which "targets" of segment production (constriction locations, degrees in the theory of articulatory phonology) may not be achieved or even pointed to. We assume that gestural reductions and omissions can occur without sacrificing intelligibility when the content of a message is sufficiently redundant. We will not consider further how these phenomena affect speaking and listening.

Gestures and synergies. We follow Browman and Goldstein (1986, 1992) in identifying consonants and vowels with actions of the vocal tract (*phonetic gestures*) that create and release constrictions. Two properties of the constriction actions are the location at which the constriction is made (*constriction location*) and the relative openness or closeness of the constriction (*constriction degree*). For example, /b/ involves a complete closure at the lips; /i/ involves a more open constriction in the palatal region.

The coordinated systems that achieve constriction locations and degrees are dynamical systems (Saltzman & Munhall, 1989) sometimes called *synergies*. The vocal tract actions of these systems generally reflect transient establishment of coordinative linkages between two or more articulators; for example, the jaw and lips for /b/, the tongue tip and jaw for /d/, the jaw and tongue body for /i/.

Synergies exhibit "equifinality," which means that their associated constriction locations and degrees can be achieved in flexible ways. For example, in a syllable such as /bi/ the coarticulating synergy for the high vowel is compatible with the synergy for /b/ in that both tend to raise the jaw. However, in /ba/, the synergy for the low vowel exerts a downward pull on the jaw as the synergy for /b/ exerts an upward pull. In the latter case, the lips contribute more to lip closure than in /bi/ because the jaw occupies a relatively lower position. In both cases, lip closure is achieved, an equifinal outcome.

The equifinality property of synergies is revealed, for example, by perturbation studies (Abbs & Gracco, 1984; Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984; Shaiman, 1989). In the perturbation procedure, an articulator is perturbed online as a speaker talks. For example, in the research of Kelso et al. (1984), a jaw puller tugged the jaw down on an unpredictable 20% of trials as a talker produced the closing jaw movement for the final /b/ in /bæb/ or the /z/ in /bæz/. For /b/ where it is compensatory, but not for /z/ where it is not, the perturbation was followed by additional activity of an upper lip muscle (compared statistically to control, unperturbed, productions,) leading to increased downward motion of the upper lip motion with a very short (20–30 ms) latency after the perturbation onset. The short latency provides good evidence that the compensation is a low-level property of synergies (automatically enabled by the transient, physiological, coordinative linkages, not by cognitive problem solving processes).

The equifinality property of synergies allows *context-sensitive* movements of individual articulators to achieve *invariant* or *context-free* constriction locations and degrees for a given segment

across the coarticulatory contexts in which it occurs. For example as noted, in /ba/ and /bi/, lip closure is invariantly achieved, but with different, contextually determined, contributions of the jaw (more for /bi/ than /ba/) and lips (more for /ba/ than /bi/).

Coarticulation as temporal overlap. In the literature, coarticulation itself is described in different ways. It is often characterized (Daniloff & Moll, 1968; Kühnert & Nolan, 1999) as adjustment of a segment to its context. Ohala (1981) has referred to its distorting effects. The authors of PSC agreed with Hockett (1955) that coarticulation destroyed segmental integrity. Coarticulation is also characterized as temporal overlap of the production of gestures for successive consonants and vowels (Bell-Berti & Krakow, 1991; Boyce, Krakow, Bell-Berti, & Gelfer, 1990; Fowler, 1980; Keating, 1990; Öhman, 1966). We suggest that temporal overlap captures the essential nature of coarticulation. In careful speech, context-sensitivity is largely of the sort we have just described. Gesture production is equifinal across phonetic contexts because, at the level of description at which vocal tract actions achieve constriction locations and degrees, there is invariance. Despite this invariance, at a more fine-grained level of description, the relative contributions of different articulators to those context-free achievements vary across contexts. As Keating (1990) comments, temporal overlap creates the *appearance* of context sensitivity. It does so, because overlap causes assimilatory acoustic consequences in neighboring segments.[17] However, those consequences belong to their gestural causes in production and also in perception as we will show. That is, language particles (see "The particulate principle: segments as meaningless particles" above) do not blend in speech production or perception any more than they do in the language as language users know it.

In a series of investigations, Bell-Berti and colleagues (Gelfer, Bell-Berti, & Harris, 1989) demonstrated that an interpretation of anticipatory coarticulation of lip rounding and nasalization as "coproduction"[18] better accounts for findings, when proper controls are in place, than either an interpretation that invokes context sensitivity (e.g., a look-ahead interpretation; Daniloff & Moll, 1968; Henke, 1967) or a hybrid model (including a look ahead component and a coproduction component, e.g., Bladon & Al-Bamerni, 1982; Perkell, 1986; Perkell & Chiang, 1986). The look ahead or feature-spreading account (Daniloff & Moll, 1968; Henke, 1967) is that a feature, such as lip rounding of a rounded vowel (e.g., /u/), or nasalization from a nasal consonant (e.g., /n/), is attached to any preceding segment that is unspecified for that feature. Thus, lip rounding attaches to all consonants preceding a rounded vowel in a VC_nR string (in which V is an unrounded vowel, the second vowel (R) is rounded and C_n refers to a consonant string of length *n*). The feature spreading can occur because English consonants do not contrast phonemically in respect to rounding (i.e., there are no consonant pairs whose members are featurally the same except that one has a rounding feature and one does not). Therefore, if a consonant becomes rounded in context, it maintains its identity. Likewise, in the look-ahead account, in English, which lacks contrastive vowel nasality, a [+nasal] feature will be attached to all versus in a string preceding a nasal consonant *N* in a CV_nN string. Although some researchers reported evidence favoring that account (Daniloff & Moll, 1968; Lubker, 1981), Gelfer et al. (1989) noted that interpretation of the findings is muddied by researchers' failure to include important control utterances. For example, Gelfer et al. (1989) showed that some consonants (e.g., /s/) are associated with lip protrusion movements. Therefore, evidence of lip protrusion at the onset of a string of consonants may reflect lip activity for the consonant, not anticipation of the lip feature of the rounded vowel. To control for that possibility, comparison utterances are required that are matched to VC_nRs except that final vowels are unrounded (i.e., VC_nV). That way, rounding because of the consonants themselves can be distinguished from coarticulatory rounding from a final vowel.

Likewise, Bell-Berti (1980) reviewed evidence that vowels have lower velum positions than oral consonants, so that there will be velum lowering in the first of a string of versus after an oral C that is because of the vowel itself. Accordingly, in assessments of anticipatory nasal coarticulation, control utterances of the form CV_nC have to be compared with the critical CV_nNs to eliminate a misleading source of velum lowering.

When appropriate control utterances are included (Boyce et al., 1990; Gelfer et al., 1989), evidence for context-sensitivity of lip rounding and nasal gestures (in the form of look ahead) disappears, leaving clear evidence for coproduction. More important, the finding is that the lip rounding or nasalization gesture begins in a time frame that is relatively invariant with respect to the rounded vowel or the nasal consonant (Bell-Berti & Harris, 1981). That is, it remains attached to the coarticulating segment itself; it does not become attached to the segments in the string, and so, for example, it may begin *within* a preceding segment, not at its onset (so that only part of it is rounded or nasalized).

Coarticulatory actions other than rounding and nasalization also reflect temporal overlap. A helpful picture in this regard is provided by Figure 2 redrawn from Lindblom and Sussman (2012). The figure shows X-ray tracings of sagittal views of the vocal tract for the syllables /da/ and /di/ overlaying two points in time. The jaw is the V-shape to the right of the figure, so the back of the head is to the left. Tongue shapes are shown during consonant closure (plain solid lines) and later during the vowel (dashed lines). In both syllables, the constriction location and degree for /d/ are achieved (in that the tongue tip makes complete closure behind the teeth; see arrows in the figure). However, notably, during consonantal closure, the tongue body configurations show that vowel production has begun. The configurations of the tongue body during consonant closure are different in the two syllables, and both are in the direction of the configuration manifest later during the vowel. That is, the figure shows temporal overlap of consonant and vowel gestures during consonant closure with invariant, equifinal, achievement of the constriction location and degree of the consonant.

This degree of invariance of consonantal constrictions is not always present. For example, in production of /g/, which shares both the jaw and the tongue body with coarticulating vowels, the point along the palate at which the tongue body makes contact with

[17] For example, lip rounding lowers F2 in the domain of a rounded vowel and in preceding segments where the rounding gesture begins to be produced.

[18] The word "coarticulation" has just the right morphological composition to refer to temporally overlapping production of segments. Even so, one of us (Fowler, 1977) coined the term "coproduction" because the older term had been coopted to refer instead to distortions or obliterations of segments or to context-sensitive adjustments in the production of a segment (e.g., feature spreading) to accommodate it to its context, definitions that do not encompass temporal overlap.

the palate during /g/ closure shifts with the constriction location of the coarticulating vowel (Dembowski, Lindstrom, & Westbury, 1998). This kind of context-sensitivity reflects a vector-like averaging of the pulls of the consonant and vowel synergies on the same articulators. That is, it still reflects temporal overlap. In the next section, we see that this kind of context effect is systematically restricted.

Coarticulation resistance. There are other instances, besides production of /g/ or /k/ with vowels, in which synergies for gestures cannot readily provide compensation for perturbations by temporally overlapping gestures. Generally, in those cases, research suggests that overlap is restricted. Bladon and Al-Bamerni (1976) first identified this characteristic of speech, known as *coarticulation resistance*, and it has been studied more extensively especially by Recasens and colleagues (1984, 1985, 1989, 1991; Recasens & Espinosa, 2009).

Using both articulatory (electropalatographic) and acoustic measures, Recasens (1984; see also, Recasens, 1989) showed, for example, that consonants resist coarticulatory overlap by preceding and following vowels to a degree that varies directly with the extent that consonants make use of the tongue body, a primary articulator for the vowels.[19] Accordingly, to the degree that vowel production during a consonant would interfere with the consonant's gestural goals, its temporal overlap with the consonant is curtailed.

Consonants and vowels can be generally characterized as higher or lower in resistance to coarticulatory encroachment by neighbors. Consonants that use the tongue (excepting /g/, /k/) tend to be higher in resistance to overlap from vowels than consonants that do not. This is shown clearly in the articulatory measures of Fowler (2005) who tracked a single speaker's speech using a magnetometer (EMMA). In Figure 3 (Fowler, 2005, redrawn based on her Figure 1), tongue height measures are shown from four successive measurement points in a schwa vowel (Panels 1–4) in a schwa-CV disyllable, during consonant closure (Panel 5) and during the final (stressed) vowel (Panel 6). Consonants included three high resistant and three low resistant consonants. Measures in Figure 3 are averaged separately over the three high (/d/,/ð/, and /z/) and low (/b/,/v/, /g/) resistant consonants (for a view of each consonant separately, see Fowler, 2005). Vowels were one of six. Panel 6 shows that tongue height varies in the expected ways across the six vowels. Panels 1–5 show how that *pattern develops over time* as the stressed vowel begins to be produced during production of the preceding schwa-C. Beginning weakly in Panel 1 (schwa onset) and growing clearer over the next four panels, two quite distinct patterns emerge. When the consonants are low in resistance (gray lines), tongue height during the schwa vowel and during consonant closure follow the upside-down V height pattern that is clearest during the stressed vowel itself (Panel 6). When the consonants are higher in resistance (black lines), however, the tongue shows very little height variation across the six stressed vowel contexts, including during schwa. Tongue fronting measures show a similar pattern as do acoustic measures of F2 (Fowler, 2005).

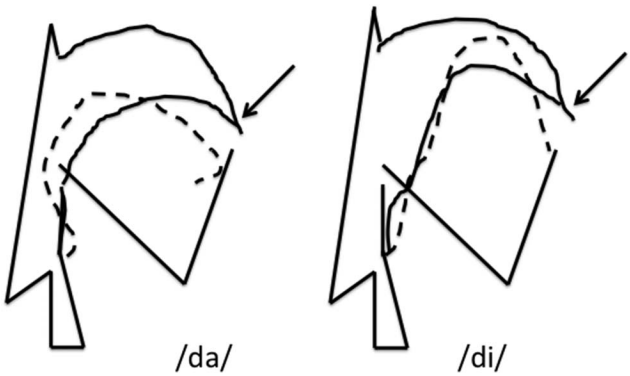

Figure 2. Superimposed sagittal X-ray views of the vocal tract at two time points in the syllables /da/ and /di/. The jaw (V shape) and lip region of the vocal tract are to the right in each display. Solid lines: tongue shape during /d/ constriction showing common tongue tip constriction locations for the two syllables (arrows). Dashed lines show later tongue positions for low back vowel, /a/ (left) and high front vowel, /i/ (right). Adapted from "Dissecting Coarticulation: How Locus Equations Happen," by B. E. F. Lindblom and H. M. Sussman, 2012, *Journal of Phonetics, 40*, p. 7. Copyright 2012 by Elsevier.

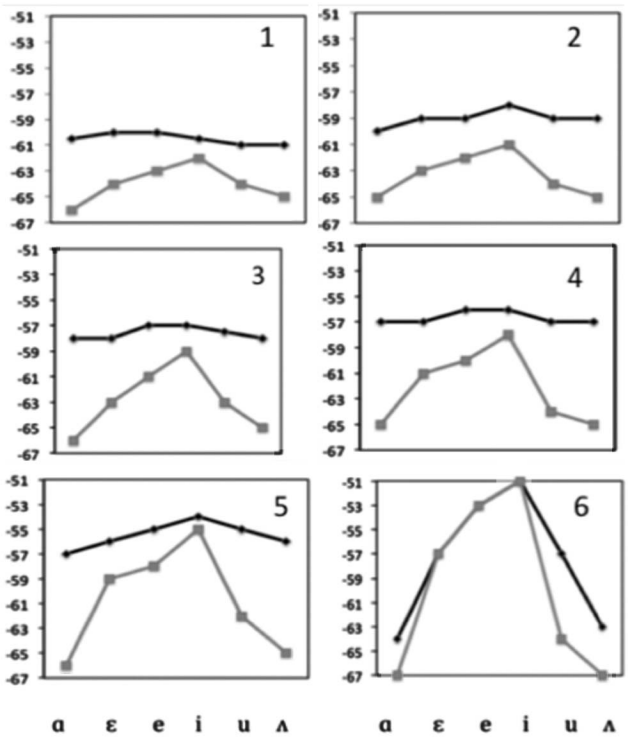

Figure 3. Redrawn from Fowler (2005; Figure 1) showing variation in tongue height (in mm) for schwa-CV disyllables with the six versus displayed along the *x*-axis. Measures were taken at six points in time: Four points during schwa (Panels 1–4). Panels 5, 6: during C closure and in mid V. Cs are either low in coarticulation resistance (/b v g/, gray lines with squares) or high (/ð *d* z/, black lines with diamonds). Anticipatory coarticulatory fronting because of V production is present to a greater extent for low- than high-resistant consonants.

[19] In English, /g/ and /k/ are exceptions in that they permit vowel overlap despite using the tongue body. Their low resistance to coarticulation possibly reflects the lack of other stops (e.g., uvular or palatal stops) that might be confused with velars because of effects of coarticulatory overlap.

Findings of coarticulation resistance are important, because they show that Hockett's view (endorsed by PSC) that coarticulation is destructive is wrong, and they show why it is wrong. Hockett suggested that effects of coarticulation in speech are analogous to the fate of an array of brightly colored Easter eggs (consonants, vowels) sent through a coarticulatory wringer, which destroys the integrity of segments. This is a misleading analogy: Instead, coarticulation is fundamentally temporal overlap of gestures produced by synergies with an equifinality property that permits invariant achievement of gestural goals (constriction locations, degrees) despite coarticulatory perturbations by nearby segments. When the interfering effects of a nearby segment would, despite equifinality, interfere with achievement of gestural goals, the gesture resists coarticulatory overlap (see Figure 3), thereby preserving segmental information.

Intergestural Links

As noted earlier, we accept most of the claims of articulatory phonology as Browman and Goldestein (1986, 1995, 2000) have presented them. That is, we agree that primitive phonological units are vocal tract gestures that are coupled in various ways in word and utterance production. Our sole disagreement with Browman and Goldstein's theoretical perspective is that we claim that linkages between recurrent patterns of gestures form phonetic segments. For Browman and Goldstein and colleagues (Browman & Goldstein, 2000; Goldstein, Byrd, & Saltzman, 2006), gestures in syllable onsets and gestures in onsets and the following vowel are coupled. However, these linkages may or may not create phonetic segmental units.

Our reasons for disagreeing in principle have been provided under "Segments in language as known (in 'competence')." That is, they are reasons why segments are necessary components of language as it is known; this is coupled in our approach to an idea that speaking conveys language forms intact. However, we acknowledge that there is, as yet, little direct articulatory evidence that gestural complexes corresponding to conventional segments emerge from coordinative links between gestures.

One kind of evidence is provided by Gelfer and colleagues (1989; Bell-Berti & Harris, 1981), briefly mentioned above. Their investigations of anticipatory coarticulation of lip rounding and nasality suggested to them that lip rounding and velum lowering gestures are tied temporally to primary articulations for the rounded or nasal segment (tongue body constrictions for rounded vowels, lip, tongue tip, or tongue body gestures for /m/, /n/ and /ŋ/, respectively). With proper controls in place, their results disconfirm the claim that rounding or nasality begins at the onset of segments preceding the rounded vowel or nasal consonant that are unspecified for the rounding or nasal feature.

A second kind of finding, also previously described, was evidence in research by Goldstein et al. (2007) that in elicited nasal consonant errors the velum gesture controlling nasality moved with the place constriction gesture with greater than chance likelihood. However, we note that in their stimuli syllable codas were single consonants; therefore, the coupling between the gestures can reflect either coupling between the gestures of a consonant or between the gestures of a syllable coda.

A final piece of evidence makes use of the perturbation procedure, described earlier. Munhall, Lo͡ɛqvist, and Kelso (1994) used a lip paddle to perturb the lower lip. They delayed the lower lips's contribution to lip closure for the first /p/ in /ipip/. This delay was accompanied by a delay in laryngeal abduction, the other gesture of /p/ besides lip closure, revealing a coupling between the two gestures. The following vocalic gesture was not delayed, however. This finding appears to show tighter coupling between the two gestures of /p/ than between the oral constriction gestures of /p/ and /i/, the following vowel.

These findings do not, by any means, constitute strong evidence that bi- or multigestural phonetic segments are articulatory units. The three sources of evidence that we summarized (that are the only relevant findings of which we are aware) are not decisive. However, we are not aware of counterevidence. Reasons to expect eventual confirmation are the nine kinds evidence favoring the reality of segments in individual language users' language capacity coupled with a commitment to a view that components of the language capacity are realized in public use of language.

The Acoustic Speech Signal

Preservation of segmental structure in speech production does not guarantee that segmental structure is available to speech perceivers. For it to be available to them, acoustic speech signals have to provide sufficient information for segments. However, there is a view (Atal & Hanaver, 1971; Diehl, Lotto, & Holt, 2004; but see Iskarous, 2010) that inverse transforms (i.e., mapping from an acoustic speech signal to its articulatory causes) are indeterminate, even "intractable" (Diehl et al., 2004, p. 172).

It is true that snapshots of an acoustic signal are consistent with more than one configuration of the vocal tract. Accordingly, recovery of articulatory states from acoustic snapshots by listeners is indeterminate. For some researchers (Diehl et al., 2004), this is evidence enough against a theory of perception in which articulations and, therefore, gestural segments are recovered by listeners. However, the objection loses force if the perceiver is supposed to track gestural information over time.

We preface the next part of our discussion by remarking that the ideas we present here are inspired by Gibson's (1966, 1979) general theory of perception. The general idea is that perceptual systems, whether visual, auditory, haptic, and so forth, enable organisms to be in direct contact with their environment; moreover, they do so in the same general way across the modalities. Objects and events in the environment (e.g., speech actions) causally structure media (light, air, the surfaces of the body, etc.) to which sense organs are sensitive. Distinctive properties of objects and events tend to cause distinctive structuring of the media so that the structuring, characteristically over time, provides information for their causal source. It serves as information to perceivers who intercept the structure perceptually. They use the information as such, that is as information for its causal source.

If gestures are phonological primitives as we are supposing, then listeners should seek acoustic structure that has been caused by gestures (and constellations of gestures that form segments) and that, therefore, serves as information for them. This will be information that develops over time in the acoustic signal. In particular, as we have noted, production of a gesture first waxes then wanes in the vocal tract. Most gestures begin in a time period during which a preceding gesture is still being produced. At that time, the gesture that is starting will have a smaller impact on the acoustic

signal than will the ongoing gesture. Over time, however, the first gesture, having reached its gestural goals, will wane in its dominance in the vocal tract as the next gesture grows in dominance. This succession of waxing then waning articulatory patterns should have roughly parallel waxing-waning acoustic consequences. That is, phonetic information develops over time, much as Joos (1948) conceived it in the seminal work of modern acoustic phonetics. Listeners aiming to identify successive phonetic segments in speech, then, should track these patterns. That is, ideally the acoustic consequences of producing successive phonetic gestures should be roughly that in Figure 4 with coarticulation resistance modulating temporal and shape details of the overlap.

The idea is that listeners begin hearing gesture A as soon as its acoustic effects become audible. Listeners track the gestural action that is gradually being signaled by the waxing wave of acoustic consequences of the gesture. They begin hearing B when it begins to have audible consequences during production of A. The acoustic consequences of A and B are distinct in participating in different strengthening and weakening sets of effects. Overlap does not produce blending. Rather than hearing an acoustic blend during periods of gestural overlap, say, between gestures A and B, listeners track each separate gesture as it is separately signaled by its strengthening and weakening acoustic consequences. Listeners do not use acoustic snapshots to recover a succession of configurations; instead, they track a succession of coherent, temporally overlapping, waves of acoustic consequences to recover a succession of temporally overlapping phonetic gestures.

That acoustic specification of speech gestures takes the form of a wave of increasing impact on the acoustic signal is suggested by acoustic measures of anticipatory coarticulation. For example, Fowler and Brancazio (2000) had two speakers produce schwa-CV disyllables in which consonants were each of six voiced obstruents and stressed vowels, V, were /i/ and /ʌ/, the vowels in *heat* and *hut*. Figure 5 shows differences in F2 (/i/ minus /ʌ/) at two points during schwa and at consonant release into the stressed vowel. That differences are positive means that information for the stressed vowel gesture (/i/ having a higher F2 than/ʌ/) is available before voicing begins for the vowel. It is available because the vowel gesture begins well before that. Moreover, the acoustic differences grow over time marking the waxing of the vowel gesture and its acoustic consequences.

There is an indication, a kind of existence proof, that growth of information about a specific gesture in an acoustic wave is detectable and is separable from other temporally overlapping acoustic waves caused by other gestures. The existence proof is provided by a version of Elman and McClelland's (1986) TRACE model. In this lesser-known version of their TRACE model (McClelland & Elman, 1986) only prelexical (feature and phoneme) levels were instantiated. Input came from acoustic signals processed to detect evidence for different features in the input. Because of the overlapping effects of production of both a consonant and vowel in their CV acoustic inputs, features of more than one segment were detectable in overlapping time frames by the model. Figure 6 is a redrawn and simplified picture of McClelland and Elman's Figure 17.6A showing activation at the model's phoneme level at one point in time during input of the acoustic signal for /ba/. (The simplification omits for the sake of clarity showing activation of other phoneme representations in the model. However, the two we show were never lower in activation than those that we have omitted. The two we show were the main contenders for being perceived.) The figure shows the relative activations of two phonemes, /b/ and /a/. In the TRACE model, each feature and phoneme is represented in memory at multiple time slices so that, wherever in a spoken utterance a feature or phoneme is produced,

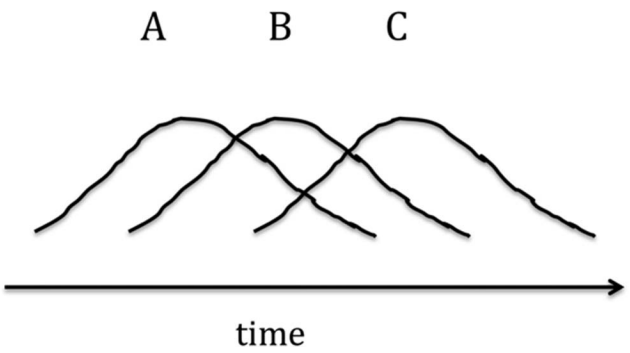

Figure 4. Schematic picture of waxing-waning patterns of both articulatory and acoustic prominence of overlapping speech gestures A, B, and C.

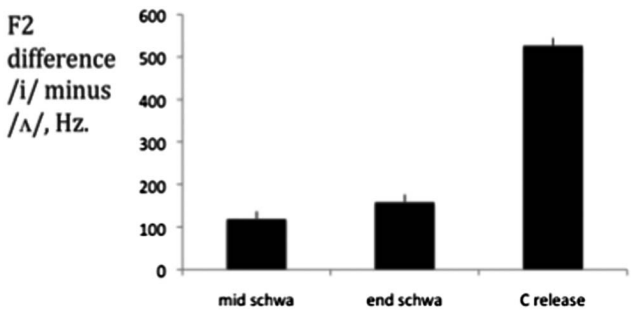

Figure 5. Development of acoustic evidence for a stressed vowel within the preceding schwa-C interval in the research of Fowler & Brancazio, 2000.

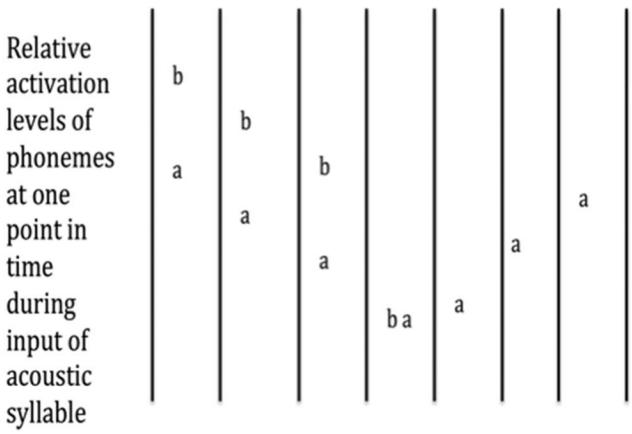

Figure 6. Schematic depiction of activation patterns over time of phoneme representations /b/ and /a/ when acoustic syllable /ba/ is input in the TRACE model of Elman and McClelland (1986). The model uses the overlapped waxing-waning acoustic consequences of production of these phonemes to recover the phonemes' identities and serial order.

its occurrence can be picked up and mapped onto the appropriate representation in memory. Relevant to the present discussion is that, over the time slices, *both* /b/ and /a/ are active. This is obviously not because the bilabial stop /b/ and the vowel /a/ have confusable acoustic consequences; they do not. Rather, it is because there is distinguishable acoustic evidence for each distinct segment in overlapping time frames in the input. Initially, /b/ is more active than /a/; later /a/ becomes more active, and /b/ activation drops out. That is, the model appears to track the separate waxing-waning acoustic waves for the consonantal and vocalic gestures and separately detects /b/ and /a/ even though the phonemes are affecting the acoustic signal in overlapping time frames. Moreover, their distinct impacts on the acoustic signal provide serial order information. (That is, the syllable is identifiably /ba/, not /ab/.)

We are not presenting the TRACE model as a model of perception. Rather, we use its pattern of activation over time slices to illustrate how the separate, but overlapping acoustic impacts of separate, ordered gestures are available in acoustic input to a perceptual system and can be detected as such. In the next section, we show that human listeners can perceive speech at the level of segments in just that way.

Speech Perception

Ironically, PSCs central error lay in the treatment of its titular topic, perception. The error was driven by a mistaken understanding of its own account of production and by underestimating perception. PSC assumed that the interleaving of "features" [gestures] from successive phonemes by distribution of those features across different articulators, destroyed the phoneme's articulatory coherence (as Hockett, 1955, believed). Based on those considerations coupled with positive evidence that listeners track articulation (Liberman et al., 1952, 1954), the authors of PSC proposed a "motor theory of speech perception," in which perception is effected by a specialized decoder. The decoder enabled listeners to find their "way back to the articulatory gestures that produced [a phoneme] and thence, as it were, to the speaker's intent" (PSC, p. 453).

PSC adduced "categorical perception" in support of the decoder. We can now see, however, that this was an unwarranted extrapolation from patterns of identification and discrimination along synthetic speech continua to the categories of natural speech perception. Variations along such continua, being confined to a single phonetic context and a single synthetic speaker, in no way reflect the unbounded diversity of variation in natural speech. Moreover, Fujisaki and Kawashima (1969) and later Pisoni (1973; see also Pisoni & Tash, 1974), in a more substantial series of experiments, demonstrated that the critical difference in discrimination between "encoded" consonants and "unencoded" vowels was because of differences in short term memory, not to differences in their supposed encoding.

Finally, PSC was encouraged to posit a specialized decoder in part by the then recent finding based on dichotic listening studies that the left cerebral hemisphere was specialized not only for semantic and syntactic aspects of language, but for perception of nonsense syllables formed from meaningless phonemic segments. (Shankweiler & Studdert-Kennedy, 1967). The dichotic studies formed a link in the claim that speech is integral to language, and they helped to buttress the idea central to the motor theory that speech is not like other sounds. Although the pioneering dichotic studies have continued to stimulate much research, more recently by direct neurophysiological means, the origin of the perceptual asymmetries is still unresolved (Poeppel, 2003; Zatorre, Belin, & Penhune, 2002). The general conclusion, however, that the left hemisphere specialization for language processes includes the acquired phonetic components of language is well-attested by research (Shtyrov, Pihko, & Pulvermuller, 2005).

We have acknowledged (Footnote 15) that talkers do not always produce phonemically structured utterances. However, the occurrence of phoneme speech errors and the persistence of the particulate principle in language both suggest that they must do so with some frequency. Likewise, we recognize that listeners may not always listen at a level of detail sufficient to extract information for overlapping segments even when it is present in the speech they hear. The research that we report next shows that they *can* extract segments from coarticulated speech when segments are present. Moreover, some language users must do so with some regularity. For example, children have to recover segments to learn them. Importantly, as Pierrehumbert (2006) noted, listeners must produce and detect segmental structuring of words regularly given the persistence of the phonological (or, in our terminology, particulate) principle in languages despite ongoing sound change.

From our characterization of speech production and the acoustic speech signal, we can extract some expectations about how listeners recover phonetic segments from speech when they do. A general expectation is that, because (dynamic) gestures create changing acoustic structure over time, these parts of the signal should be important information sources compared to steady-state parts of the signal. Indeed, it has been known for a long time (Liberman et al., 1954) that formant transitions provide critical information for consonants. Research also shows, however, that dynamic portions of the acoustic signal provide valuable information for vowels as well, even though vowels, intuitively, being continuant segments can provide steady-state information during intervals of minimal coarticulatory overlap with consonantal gestures.

Pioneering studies of speech production indicated that the different vowels can be characterized by specific configurations of lips, tongue, and pharynx that vary the shape of the vocal tract and give rise to different resonant frequencies (formants) in speech (Chiba & Kajiyama, 1941). Accordingly, vowels were thought to be perceived by their "target" formant frequencies taken to be the lowest two or three formant peaks observed when sustained vowels are spoken in isolation (Delattre Liberman, Cooper, & Gerstman, 1952; Fairbanks & Grubb, 1961).

However, in conversational speech, many syllables do not reach steady-state formant frequencies at any point during their temporal course (Joos, 1948; Lindblom, 1963). That would seem to indicate that vowel targets, as conventionally conceived, are idealizations that do not correspond to what happens in ordinary speech. How, then, is vowel perception achieved when listeners must extract vowel information from continuously varying acoustic signals (Shankweiler, Strange, & Verbrugge, 1977)? From experimental manipulations simulating the effects of speech rate on peak formant frequencies, Lindblom and Studdert-Kennedy (1967) showed that listeners use time-varying information in identifying vowels in rapid speech by taking account of "undershoot" of formant fre-

quencies. Undershoot (i.e., a failure to achieve "target" formant values) occurs, in the present perspective, because of gestural overlap.

Later research suggested that listeners identify vowels more successfully from time-varying formants than from steady-state formants at target frequencies. Strange, Verbrugge, Shankweiler, and Edman (1976) found that medial vowels in naturally produced consonant-vowel-consonant (CVC) syllables extracted from sentences spoken by multiple talkers were more accurately identified by listeners than were vowels spoken in isolation by the same talkers. In subsequent experiments that further undermined the vowel target idea, Strange, Jenkins, and Johnson (1983) showed that vowels can be conveyed accurately even when all the acoustic information at the center of CVC syllables is deleted (by an acoustic manipulation). The doctored, "silent-center" syllables contained no steady-state portion, and so no vowel targets. However, vowels in these silent-center syllables could be perceived about as well as in intact syllables even when 65% of each syllable's acoustic center was discarded and replaced by silence or noise (duration adjusted for short and long vowels), preserving only a only a portion of the formant transitions out of and into the flanking consonants. These findings show that vowels, like consonants, are conveyed by time varying information resulting from phonetic gestures, and that listeners pick up and use this information in perceiving running speech. Listeners do not perceive vowels (or consonants) from static targets.

Another expectation is that, like Elman and McClelland's model (Elman and McClelland, 1986), listeners should use the waxing-waning acoustic consequences of gestures as critical sources of information that permit them to extract gestures from coarticulated speech.[20] Tracking those acoustic patterns should mean that listeners begin to extract information from the acoustic signal as soon as the gesture causes audible acoustic consequences.

That they do has been shown in numerous studies, beginning, with findings of Martin and Bunnell (1981, 1982). These investigators pioneered a cross splicing technique with utterances in which anticipatory coarticulatory information should be present before the portion of the acoustic signal identified with a target segment. In their technique, pairs of utterances are cross spliced so that anticipatory coarticulatory information from one utterance is spliced onto a new context where it provides misleading information about the identity of the context. For example, Martin and Bunnell (1981; see also Martin & Bunnell, 1982) cross spliced /stri/ and /stru/ utterances so that the primary acoustic manifestations of /s/ or /st/ from /stri/ were spliced onto /ru/ of stru and those of /s/ or /st/ from /stru/ were spliced onto /ri/. In these cross spliced syllables, coarticulatory information in the /s/ or /st/ was misleading about the identity of the following vowel. Listeners identified the vowels in a speeded task. Findings were that, relative to intact syllables, listeners were slower to identify the vowel in cross spliced syllables and were slowest when information was misleading in both prior consonants.[21]

More recently, Beddor, McGowan, Boland, Coetzee, and Brasher (2013) have used eye tracking to examine listeners' use of anticipatory information for a forthcoming nasal consonant in preceding vowels. They presented listeners with a spoken word (e.g., *bed*, or *bend*) embedded in a sentence: *Now look at*—_____. On a computer screen were two pictures one either side of a fixation point. In the example, there was a picture of a bed and a picture that participants were trained to identify with *bend*. Acoustic syllables with nasal consonants were manipulated to have either the final 40% or the final 80% of the target-word's vowel nasalized. On trials on which the words with nasal consonants were presented acoustically, first looks to the correct picture (e.g., of bending) had shorter latencies the earlier nasalization began in the vowel. Estimates of when, relative to the onset of nasalization in the vowel, the choice where to look was made suggested that it was made in the vowel, before the onset of the closure for the nasal consonant.

The collection of findings just summarized is consistent with an interpretation that listeners begin to hear a segment as soon as it begins to have audible acoustic consequences even when that point in time precedes the conventionally determined acoustic onset of the segment. However, by itself, the findings are ambiguous. By our interpretation, listeners are tracking the waxing of a wave of acoustic consequences of production of a gesture. However, if coarticulation is interpreted as context sensitivity, not temporal overlap, the results merely show that listeners use the context coloring of a preceding segment to identify its context (e.g., in findings of Beddor et al. [2013], they used a (partially) nasalized vowel as information that a nasal consonant is forthcoming).

Fowler (1981, 1984; Fowler & Smith, 1986) provided evidence that these effects reflect true tracking of a wave of acoustic information by testing for a characteristic of speech perception that only should accompany the cross splicing and eye tracking findings under that interpretation. If listeners are tracking separate waxing-waning acoustic waves caused by production of overlapping gestures, then, even though acoustic effects of two segments blend acoustically, listeners should not hear the blending. For example, they should not hear vowels as nasalized or not in research by Beddor et al. (2013). Fowler (1981; see also Fowler & Smith, 1986) used a discrimination procedure to test this prediction. In brief, they asked listeners to discriminate pairs of schwa vowels (/ə/) presented as gesture B in Figure 4 above, that is, in a coarticulatory context in which preceding and following stressed

[20] A reviewer asked why, even if listeners do track waxing-waning acoustic patterns, the perceptual objects need to be supposed to be phonetic gestures instead of, say, such "underlying representations" as allophones or syllables. As we argued under "The acoustic speech signal," our claim is that listeners extract information from acoustic signals that *is* information because it was lawfully caused by events in the world, actions of the vocal tract. In speech, those actions are phonetic gestures. We are not making a proposal about underlying representations, but about events in the world that listeners perceive by detecting the structuring they cause in a medium, here largely air, to which a perceptual system, here the auditory system, is sensitive.

[21] Whalen (1984) added an important control condition to the procedure. Instead of comparing response times to cross spliced and intact utterances, the comparison was between cross spliced stimuli and control stimuli (henceforth "spliced" stimuli, to contrast with "cross spliced") in which splicing occurred between utterances of the same type (e.g., /st/ from one token of /stri/ would be spliced onto /ri/ of another). In that way, cross spliced and control, spliced, utterances were both subject to a splicing operation that might in itself slow response times to them; however, only cross splicing created a misleading context. Using this control, Whalen (1984); Fowler (1984), and Fowler and Brown (2000) replicated findings of Martin and Bunnell (1981, 1982) with new stimuli involving identification of vowels in the context of spliced and cross spliced consonants and identification of consonants in the context of spliced and cross spliced vowels.

vowels would overlap with them. Listeners heard acoustically identical schwas as different if the contexts (e.g., high vowels as gestures A and C in one case [/ibəbi/] and low vowels as A and C in the other [/ab1368>ba/]) perceptually extracted from them should leave different gestures B. They heard acoustically different schwas as the same if subtracting the stressed-vowel contexts should leave similar gestures B. Compatible findings for perception of stop consonants in different vocalic contexts was reported by Fowler (1984) and for vowels in the context of oral or nasal contexts by Fowler and Brown (2000).

Like the cross splicing and eye tracking findings, the discrimination findings just summarized are amenable to more than one interpretation. We have interpreted them as companion findings to the cross splicing findings. If listeners extract information for segments from acoustic signals by tracking the waxing-waning acoustic effects they cause, then the cross splicing and eye tracking findings reflect the perception of a waxing acoustic wave. The discrimination findings show that blended acoustic effects during overlap between two waves do not blend perceptually. Rather, listeners parse or pull apart the overlapping acoustic information for distinct gestures.

However, it is important to rule out another interpretation of the discrimination findings, namely, that they reflect auditory contrast. We know that perceivers show contrast effects broadly (Warren, 1985). Asked to judge the heaviness of an intermediate weight after hefting a heavy one leads to a judgment that the weight is lighter than if the same weight is hefted after the perceiver hefts a light weight. Plunging the hand into room-temperature water after plunging it into hot water leads to a judgment that the water temperature is cooler than after experience with cold water. The finding of Fowler (1981) can be explained that way. In the context of flanking high (/i/) vowels with high F2 and F3, a schwa vowel with intermediate formant values will sound lower than in the context of flanking low (/a/) vowels.

These competing interpretations have been tested in multiple studies. Before presenting our conclusion that the contrast interpretation has been vanquished, we comment that, if the contrast account is correct, then the findings of listener sensitivity to anticipatory coarticulatory information summarized earlier in this section and the discrimination findings just summarized would have two independent explanations. The cross splicing and eye tracking findings occur because listeners use their sensitivity to the context provided by preceding segments to predict the identity of an adjoining segment. The discrimination findings are contrast effects that erase evidence of context sensitive acoustic structure in perception. Instead, under our account, there is just one explanation for both findings; both reflect listener's tracking the waxing-waning acoustic patterning caused by gesture production.

We cannot review here the entire array of research designed to distinguish a contrast account from a gesture tracking account of listeners' mode of perceiving coarticulated speech. Instead, we will review why, in our view, the data disconfirm the contrast view and confirm the gesture tracking interpretation.[22]

Mann (1980) introduced the term "compensation" for coarticulation to capture a finding in which listeners appear to do just that. She presented listeners with members of a /da/ to /ga/ acoustic continuum in which syllables differed only in the onset frequency of the third formant transition (high for /da/, low for /ga/). Precursor syllables were /al/ or /aɹ/ (/ɹ/ being the "r" sound in English). Findings were that listeners reported more ambiguous syllables as "d" when the precursor was /aɹ/ than when it was /al/.

A plausible coarticulatory effect of /ɹ/ on /d/ is to pull its point of constriction back; a plausible coarticulatory effect of /l/ on /ga/ is to pull its point of constriction forward. Accordingly, listeners who compensate for these effects will identify syllables that are ambiguous in their point of constriction as "da" in the context of /aɹ/, but as "ga" in the context of /al/. Put in the terms we have been using, they track what should be the waning edge of the acoustic consequences of /l/ and /ɹ/ into the domain of the /da/-/ga/ continuum members, ascribing them to /l/ and /ɹ/ gestures. Acoustic evidence of achievement of a back constriction for /ɹ/ and a front constriction for /l/[23] begins before evidence of stop consonant production begins and continues during evidence of stop production. Because listeners have already begun perceiving a backing or fronting gesture, they do not ascribe acoustic evidence consistent with backing/fronting of the liquid to the stop. In the stop-vowel syllables, a point of constriction that is fronted for /g/ and backed for /d/ provide evidence consistent with continuation of the liquid consonant gestures. This account is precisely what is required to explain the discrimination findings of Fowler (1981) and Fowler and Smith (1986) described earlier. /d/ and /g/ are not perceived as context-sensitive (/ɹ/- or /l/-colored); when listeners track the waves of acoustic information for coarticulatory gestures, they "compensate" for those coarticulating gestures in identifying the stops.

Mann (1980) provided the foregoing interpretation of her findings and also an alternative interpretation. The alternative was that the ending frequency of the third formant (F3) of /aɹ/ or /al/ might exert a contrastive effect on the onset F3 of members of the stop continuum. The low ending F3 of /aɹ/ can make the onset frequency of continuum syllables appear higher (and so more /da/-like); and the high ending F3 of /al/ can make the onset frequency of the following syllable appear lower and so more /ga/-like.

If the findings were, in fact, because of contrast, the contrast would be based on a far more subtle stimulus property, say, than is operative by plunging one's hand in water that is everywhere hot

[22] The issue is, by some accounts (Kingston, Kawahara, Chambless, Key, Mash, & Watsky, 2014), not resolved in our favor, however. Kingston et al. show in one experiment that compensation for coarticulation occurred in stimuli of theirs designed after those of Mann (1980) described next in the text above. This involved presentation of disyllables in isolation to which participants gave identification responses. In a second experiment, they presented pairs of disyllables from the same stimulus set chosen to maximize or minimize contrast differentially within pairs. (That is, if H and L refer to high and low frequencies in relevant parts of each syllable of a disyllable, disyllable pairs that should be maximally different in terms of contrast would be HL vs. LH. The H in the first disyllable should enhance the lowness of the L in the next syllable, etc. Disyllable pairs that should be less discriminable would be HH vs. LL.) They found better discrimination between members of pairs in which contrast should enhance perceived differences. They infer that, because contrast was induced between pairs of disyllables in Experiment 2 and affected discrimination judgments, which contrast underlay the absolute identification responses to individual disyllables in Experiment 1. However, nothing other than use of the same stimulus set and propinquity in the article is provided to buttress that inference. We invite readers to assess this finding in relation to those we present here against the contrast account.

[23] /ɹ/ and /l/ are complex segments, and there is one (for /l/) or two (for /ɹ/) other constrictions involved in their production. Those on which we focus are presumed to be the effective ones in this context.

or cold or than lifting an object that is everywhere heavy or light. In the case of the speech disyllables, the source of contrast, is a five ms portion (or 50 ms if the entire transition is effective) of just one part of an acoustic signal that is otherwise not in a contrastive relation with the following continuum syllables.

Nonetheless, some evidence seems to support an interpretation that the finding cannot be true compensation for coarticulation. It is that effects qualitatively like those of the precursor syllables can be obtained for nonspeech precursor sinewave tones with frequencies set at the ending F3 of /al/ and /aɹ/ (Lotto & Kluender, 1998; but see below, Viswanathan, Magnuson, & Fowler, 2013).

It is difficult to distinguish the contrast and compensation accounts, because coarticulation causes acoustic effects on neighboring segments that resemble the acoustics of the segment itself (i.e., coarticulation has assimilatory acoustic consequences; see (Footnote 17). Compensation ascribes those assimilatory effects to the coarticulating segments so that they are not heard as part of the affected segment; contrast works to erase perception of the effects in the domain of the affected segment.

Even so, the competing accounts of the compensation-like findings have been distinguished:

1. In two studies (Johnson, 2011; Viswanathan, Magnuson, & Fowler, 2010), investigators found stimuli in which predictions of contrast and compensation accounts could be directly contrasted. Both found that listener responses reflected compensation for coarticulatory effects on consonant place of articulation; they did not reflect contrast.

2. Viswanathan, Fowler, and Magnuson (2009) isolated the very part of the context syllables /al/ and /aɹ/ that, in contrast accounts, exerts a contrastive effect on /da/-/ga/ perception; that is, they filtered the precursor syllables to isolate the F3 transitions. These nonspeech precursors exerted no significant effect on /da/-/ga/ judgments. (A further study, Viswanathan, Magnuson, & Fowler, 2013, showed that the sinewave tones found to be effective in earlier research [Lotto & Kluender, 1998] did not exert contrast effects either. Their effectiveness derived from their very high intensity, which led to energetic masking.)

3. Compensation for coarticulation has been found to occur when the only information distinguishing the relevant context (e.g., /al/ vs. /aɹ/ in one set of findings) is visual (exploiting the "McGurk effect"; McGurk & MacDonald, 1976), whereas the information distinguishing target continuum members is auditory (Fowler, Brown, & Mann, 2000; but see Fowler, 2006; Lotto & Holt, 2006 for an independent demonstration, see Mitterer, 2006).[24] Because contrast effects are intramodal, they do not predict the cross modal findings of these studies. Instead, the findings converge with the others summarized on a conclusion that speech perceivers do compensate for coarticulation. Compensation is not a fortuitous accident of perceptual contrast.

4. Compensation for coarticulation occurs when there is no preceding (or following) context to exert a contrastive effect. For example (Silverman, 1987), listeners compensate for the "intrinsic f0" of vowels (with high vowels having higher f0 than low vowels) when they judge the height of intonational peaks on the vowels.

The preceding four kinds of evidence converge on the conclusion that listeners track the overlapping waxing-waning acoustic consequences of gesture production (or visual evidence of gesture production) and thereby in fact compensate for coarticulatory overlap in segment perception. More generally, the perceptual evidence discussed in this section is, overall, consistent with a view that listeners begin to hear production of a gesture when the acoustic signal begins to provide audible acoustic evidence for it. They do not hear intervals in which effects of two gestures converge as blended. Rather, they separately track distinct waxing-waning acoustic consequences of gestural causes. This leads, among other perceptual consequences, to compensation for coarticulation. This strategy would explain how phonetic segments can be extracted from speech signals despite coarticulatory overlap.

There is a final issue to address about speech perception. In our earlier discussion of Port's (2007, 2010a, 2010b) reasons for rejecting phonemes as known, we noted that he invokes findings that listeners learn about speech *episodes* in which, say, a word was produced, by a male or female speaker, an adult or child, with a distinctive voice quality, in some speech register, perhaps in some emotional state (perhaps wearing a hat; Sheffert & Fowler, 1995). As previously discussed, information of all of those kinds is coupled in memory. Because we have focused exclusively on listeners' tracking phonetic gestures, it may appear to readers of the foregoing discussion as if we are arguing instead for a more classical view of speech perception in which phonetic information is stripped from other speech-episode information and somehow is isolated in memory. However, that is not the case. We have shown how listeners can extract information specifically about overlapping phonetic gestures. In a sense, we have shown how "normalization" (Goldinger, 1998, and references there) can occur; that is, how a listener can identify, say, a complete constriction-release gesture at the lips, whether the talker was an adult or a child, male or female, talking fast or slowly, and so forth. This is because the waxing-waning acoustic-information wave is qualitatively the same under all of these conditions. However, it does not follow that the listener is not, at the same time, detecting acoustic structure that identifies characteristics of the speaker, the speaker's voice, speech register, emotional state or speaking rate. Any property of a speech action that causes distinctive and detectable structuring of the acoustic signal has the chance of being detected, perceived, and learned about as part of a speech episode. It is, in our view no different than saying that an observer can see: a table, a round table, a round mahogany table, at some distance, in some direction from the observer. Detecting and learning about one of those properties does not preclude, at the same time, detecting and learning about others.

[24] Explaining cross modal speech perception requires an augmentation of the idea that perceivers track waxing-waning acoustic consequences of gestures. However, the augmentation is a natural one. Perceivers extract information for gestures, however, that information may be conveyed across the perceptual modalities (Fowler & Dekle, 1991; Gick & Derrick, 2009).

Summary and Conclusions

Our purpose has been to address a paradox in the literature on speech and language that underlay development of the motor theory of speech perception in PSC and is a persisting legacy of that article. On the one hand, there is evidence in favor of the existence of phonetic segments in language. On the other hand, there is evidence that has led theorists to reject their existence in public implementations of language. The authors of PSC did not question this ostensible state of affairs. Pioneering work that preceded PSC (Cooper, 1950; Studdert-Kennedy & Liberman, 1963) on development of a reading machine for the blind began with a supposition that there were segments in the head and in the signal: Speech was a sound alphabet (see Shankweiler & Fowler, 2015, for an account of this earlier history). This research that seemed to show the "encoded" nature of speech (Liberman et al., 1952, 1954) caused a major shift in their thinking to a view that segments are not present in acoustic speech signals as coherent entities, because coarticulation necessarily eliminates them. The authors of PSC did not use their findings as a springboard to reconsider the conventional linguistic characterization of segments as static units, and to bring it, if possible, into better alignment with speech as implemented. In the present article, we have done just that.

Language forms as known provide the means within language for communicating with others; they should, if possible, be adapted to public language use. Therefore, the idea that phonetic segments are in the language as language users know it, but not in public implementations of an utterance is paradoxical. We have reconsidered the nature of linguistic segments, their implementation in speech production, their acoustic signatures, and perception by listeners, all with a motivation to determine whether all can be brought into close alignment. We believe that our characterization promotes an understanding of segments as well-adapted to their public use in speech communication, because they persist throughout communicative exchanges: in intentions to speak (as revealed, e.g., by speech errors), as articulated, as signaled acoustically, and as perceived.

In PSC, the authors devoted considerable attention to the idea that that acoustic signal is a code on the phonemes of the language. We suggest here that, instead, the acoustic signal is closer to an alphabetic cipher than to a code, the position rejected by PSC, but supported here with new reasons.

To counter the view that segments do not exist even in language as known, we summarized nine kinds of evidence that converge to show that segments are real components, not only of language as a "social institution" (Port, 2010b), but also as a capacity of individuals to use language. The particulate principle is a crucial aspect of language that underpins its generativity at the level of the lexicon. It is perpetuated in language use, as Pierrehumbert (2006) remarks, despite ongoing language change, because when a segment undergoes diachronic phonetic change it tends to undergo that same change across all words in the lexicon that contain the segment albeit at different rates for words of different frequencies. Segments reveal themselves as components of individuals' language capacity, for example, in spontaneous errors of speech production, in generative use of phonological processes such as metathesis, and in generative use of morphological processes such as infixing in Classical Arabic and in the existence of alphabetic writing systems.

In our view, the implicit ignoring of the paradox that accepts the foregoing evidence for segments as known, but rejects segments as making public appearances is unsustainable. It is unsustainable, because speech is intrinsic to the evolved capacity for language (e.g., PSC; Liberman, 1970; Shankweiler & Studdert-Kennedy, 1967). Excepting in special circumstances such as deafness, language is universally spoken. Although signed languages have expressive power equivalent to that of spoken languages, it is not a coin toss whether a culture will have a spoken or a signed language. In spoken languages, phonological forms are the means within language for making communications sharable by making them public. In our view, it is implausible that languages universally develop forms (i.e., segments) that cannot be made public without being destroyed (Hockett, 1955) or distorted (Ohala, 1981) or encoded.

We proposed a way to understand how segments may be preserved in speech production despite coarticulation. A first move in this direction, following PSC, is to recognize that coarticulation is fundamentally temporal overlap of speech actions. It is not fundamentally adjustment of language forms to their contexts. Nor is it distortion of the forms or destruction of them. In moving ahead, we followed Browman and Goldstein's (1986) proposal that primitive phonological forms are gestures of the vocal tract. These gestures are implemented as dynamical systems or synergies that have an equifinality characteristic. This means that the gestures can achieve their critical constriction locations and degrees flexibly despite perturbations because of coarticulatory overlap with other gestures. Moreover in coarticulatory contexts in which achievement of gestural properties would be difficult because of conflicting demands on shared articulators, the research shows (Recasens, 1984, 1989; Recasens & Espinosa, 2009) that gestures resist coarticulatory overlap. All of this implies that speech production can be seen (in clear speech) as a succession of temporally overlapping gestures.

It would be unhelpful to communication if that overlapping preserved segmental identity in production but did not permit identification of segments in perception. Indeed, in PSC, "decoding " the speech signal required help from a neural specialization for speech, a decoder (see also Liberman & Mattingly's [1985] phonetic module) that incorporated "complete" knowledge of coarticulation and its acoustic effects. We suggested that the acoustic speech signal can be seen as less inimical to segmental perception than has been represented in the literature if we look for acoustic evidence for successions of overlapping gestures rather than for acoustic snapshots (or "cues") that might permit recovery of associated vocal tract configurations. Acoustic consequences of overlapping gestures should be successions of waxing then waning patterns. That such patterns can be tracked is suggested by the behavior of Elman and McClelland's (1986) TRACE model that appears to do just that.

Finally, we cited evidence that human listeners track the waxing-waning patterns as well. First, they extract information preferentially from dynamical acoustic change rather than static signals as shown in research by Strange and colleagues (Strange et al., 1976, 1983). A further sign of their extraction of information for overlapped gestures is that they begin to hear a segment (e.g., a nasal consonant in research by Beddor et al., 2013) when it

begins to have audible acoustic consequences, even though that is in the domain of a preceding segment (a vowel in the research of Beddor et al., 2013). Finally, in context, listeners do not hear stretches of an acoustic signal in which effects of two segments are intermixed as blended, that is, as context-sensitive (Fowler, 1981 and research on "compensation for coarticulation").

We believe that we have shown that the proper resolution of the segment paradox is that segments exist as a capacity to use them among individual users of a language, but also as units of speech production that listeners can perceive, because they are adequately signaled acoustically. Therefore, we can understand how the particulate principle, so crucial to productivity in language use, can be preserved in public language use.

We acknowledge that we are writing about clear speech. In social contexts, highly predictable utterances (such as "I don't know" implemented as a continuous vocalic segment as described earlier) may not have segmental structure. Speakers may only speak as carefully as they need to get their message across. For their part, listeners may not always attend to speech at the level of segmental structure. However, we note following Pierrehumbert (2006), that segmental structure persists in language; its generative powers remain. Accordingly, segments are produced and perceived frequently enough for that.

We also acknowledge that, because language structures and processes emerge, develop, and change in the course of their public use, the structures and processes are untidy. They do not have the neatness of theoretical linguistic accounts of them. There may be issues whether some collections of gestures are segments or clusters (e.g., /sp/, /st/, /sk/ [Fudge, 1969] or the first and last consonants of *church* and *judge*) that may be unresolvable just as it may be undecidable (and unimportant) sometimes whether a sequence of syllables should count as one word or two (e.g., *crossmodal* or *cross modal*). Finally, as noted (Footnote 9), experiments have shown that contrasting segments may carry a degree of meaning in some contexts. Despite this untidiness, the logic of the particulate principle and the evidence of the alphabet demonstrate that the elements of spoken language are necessarily functionally meaningless.

None of these sources of fuzziness in the language changes the "bottom line." It is that segments are essential to the communicative power of language and are well adapted to their participation in public communication by means of speech.

References

Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology, 51,* 705–723.

Abler, W. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures, 12,* 1–13. http://dx.doi.org/10.1016/0140-1750(89)90015-8

Atal, B. S., & Hanaver, S. L. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *The Journal of the Acoustical Society of America, 50,* 637–655. http://dx.doi.org/10.1121/1.1912679

Beckman, M. E., & Kingston, J. (1990). Introduction. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 1–16). Cambridge: Cambridge University Press.

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America, 133,* 2350–2366. http://dx.doi.org/10.1121/1.4794366

Bell-Berti, F. (1980). Velopharyngeal function: A spatiotemporal model. In N. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 4, pp. 291–316). New York, NY: Academic Press.

Bell-Berti, F., & Harris, K. S. (1981). A temporal model of speech production. *Phonetica, 38,* 9–20. http://dx.doi.org/10.1159/000260011

Bell-Berti, F., & Krakow, R. A. (1991). Anticipatory velar lowering: A coproduction account. *The Journal of the Acoustical Society of America, 90,* 112–123. http://dx.doi.org/10.1121/1.401304

Bladon, R. A. W., & Al-Bamerni, A. (1976). Coarticulation resistance in English/l/. *Journal of Phonetics, 4,* 137–150.

Bladon, R. A. W., & Al-Bamerni, A. (1982). One-stage and two-stage temporal patterns of velar coarticulation. *The Journal of the Acoustical Society of America, 72*(Suppl. 1), S102.

Boyce, S. E., Krakow, R. A., Bell-Berti, F., & Gelfer, C. E. (1990). Converging sources of evidence for dissecting articulatory movements into core gestures. *Journal of Phonetics, 18,* 137–188.

Braine, M. D. (1994). Is nativism sufficient? *Journal of Child Language, 21,* 9–31. http://dx.doi.org/10.1017/S0305000900008655

Browman, C., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology, 3,* 219–252. http://dx.doi.org/10.1017/S0952675700000658

Browman, C., & Goldstein, L. (1990a). Representation and reality: Physical systems and phonological structure. *Journal of Phonetics, 18,* 411–425.

Browman, C., & Goldstein, L. (1990b). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics, 18,* 299–320.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49,* 155–180. http://dx.doi.org/10.1159/000261913

Browman, C. P., & Goldstein, L. (1995). Dynamics and articulatory phonology. In R. F. Port & T. van Gelder (Eds.), *Mind as motion* (pp. 175–193). Cambridge, MA: MIT Press.

Browman, C., & Goldstein, L. G. (2000). Competing constraints on intergestural coordination and self-organizaiton of phonological structures. *Bulletin de la Communication Parlée, 5,* 25–34.

Brown, R., & Nuttall, R. (1959). Method in phonetic symbolism experiments. *The Journal of Abnormal and Social Psychology, 59,* 441–445. http://dx.doi.org/10.1037/h0045274

Castro-Caldas, A., Petersson, K. M., Reis, A., Stone-Elander, S., & Ingvar, M. (1998). The illiterate brain. Learning to read and write during childhood influences the functional organization of the adult brain. *Brain: A Journal of Neurology, 121,* 1053–1063. http://dx.doi.org/10.1093/brain/121.6.1053

Chao, Y. (1931). Fan-quie Yu Ba Zhong (Eight varieties of secret language based on the principle of fan-quie). *Bulletin of the Institute of History and Philology Academia Sinica, 2,* 320–354.

Chiba, T., & Kajiyama, M. (1941). *The vowel: Its nature and structure*. Tokyo: Kaiseikan.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York, NY: Harper and Row Publishers.

Clements, G. N. (1985). The geometry of phonological features. *Phonology, 2,* 225–252. http://dx.doi.org/10.1017/S0952675700000440

Cooper, F. S. (1950). Research on reading machines for the blind. In P. Zahl (Ed.), *Blindness: Modern approaches to the unseen environment* (pp. 512–543). Princeton, NJ: Princeton University Press.

Daniloff, R., & Moll, K. (1968). Coarticulation of lip rounding. *Journal of Speech and Hearing Research, 11,* 707–721. http://dx.doi.org/10.1044/jshr.1104.707

De Francis, J. (1989). *Visible speech: The diverse oneness of writing systems*. Honolulu: University of Hawaii Press.

Delattre, P. C., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. (1952). An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word, 8,* 195–210.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321. http://dx.doi.org/10.1037/0033-295X.93.3.283

Dell, G. S. (2014). Phonemes and production. *Language, Cognition and Neuroscience, 29,* 30–32. http://dx.doi.org/10.1080/01690965.2013.851795

Dembowski, J., Lindstrom, M. J., & Westbury, J. R. (1998). Articulator point variability in the production of stop consonants. In M. P. Cannito, K. M. Yorkston, & D. R. Beukelman (Eds.), *Neuromotor speech disorders: Nature, assessment, and management* (pp. 27–46). Baltimore, MD: Brookes.

Dennis, I., & Newstead, S. E. (1981). Is phonological recoding under strategic control? *Memory & Cognition, 9,* 472–477. http://dx.doi.org/10.3758/BF03202341

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology, 55,* 149–179. http://dx.doi.org/10.1146/annurev.psych.55.090902.142028

Elman, J. L., & McClelland, J. L. (1986). Exploiting lawful variability in the speech wave. In J. Perkell & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 360–385). Hillsdale, NJ: Erlbaum.

Faber, A. (1992). Phonemic segmentation as epiphenomenon: Evidence from the history of alphabetic writing. In P. Downing, S. D. Lima, & M. Noonan (Eds.), *The linguistics of literacy* (pp. 111–134). Amsterdam: John Benjamins Publishing Co.. http://dx.doi.org/10.1075/tsl.21.11fab

Fairbanks, G., & Grubb, P. (1961). A psychophysical investigation of vowel formants. *Journal of Speech and Hearing Research, 4,* 203–219. http://dx.doi.org/10.1044/jshr.0403.203

Fant, G., & Lindblom, B. (1961). Studies of minimal speech sound units. *Speech Transmission Laboratory: Quarterly Progress Report, 2,* 1–11.

Feldman, L. B., Kostic, A., Lukatela, G., & Turvey, M. T. (1983). An evaluation of the "basic orthographic syllable structure" in a phonologically shallow orthography. *Psychological Research, 45,* 55–72. http://dx.doi.org/10.1007/BF00309351

Feldman, L. B., & Turvey, M. T. (1983). Word recognition in Serbo-Croatian is phonologically analytic. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 288–298. http://dx.doi.org/10.1037/0096-1523.9.2.288

Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review, 120,* 751–778. http://dx.doi.org/10.1037/a0034245

Finley, M. I. (2002). *The world of Odysseus*. New York, NY: New York Review of Books.

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford: Clarendon Press. http://dx.doi.org/10.5962/bhl.title.27468

Fowler, A. E. (1991). How early phonological development might set the stage for phoneme awareness. In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman* (pp. 97–117). Hillsdale, NJ: Erlbaum.

Fowler, C. A. (1977). *Timing control in speech production*. Bloomington, IN: Indiana University Linguistics Club.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing control. *Journal of Phonetics, 8,* 113–133.

Fowler, C. A. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research, 24,* 127–139. http://dx.doi.org/10.1044/jshr.2401.127

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics, 36,* 359–368. http://dx.doi.org/10.3758/BF03202790

Fowler, C. A. (2005). Parsing coarticulated speech in perception: Effects of coarticulation resistance. *Journal of Phonetics, 33,* 199–213. http://dx.doi.org/10.1016/j.wocn.2004.10.003

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics, 68,* 161–177. http://dx.doi.org/10.3758/BF03193666

Fowler, C. A., & Brancazio, L. (2000). Coarticulation resistance of American English consonants and its effects on transconsonantal vowel-to-vowel coarticulation. *Language and Speech, 43,* 1–41. http://dx.doi.org/10.1177/00238309000430010101

Fowler, C. A., & Brown, J. M. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception & Psychophysics, 62,* 21–32. http://dx.doi.org/10.3758/BF03212058

Fowler, C. A., Brown, J. M., & Mann, V. A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 877–888. http://dx.doi.org/10.1037/0096-1523.26.3.877

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 816–828. http://dx.doi.org/10.1037/0096-1523.17.3.816

Fowler, C. A., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell & D. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 123–136). Hillsdale, NJ: Erlbaum.

Fromkin, V. (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123,* 71–99. http://dx.doi.org/10.1037/0033-2909.123.1.71

Fudge, E. C. (1969). Syllables. *Journal of Linguistics, 5,* 253–286. http://dx.doi.org/10.1017/S0022226700002267

Fujisaki, H., & Kawashima, T. (1969). On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute, 28,* 67–73.

Gafos, A. (1999). *The articulatory basis of locality in phonology*. New York, NY: Garland Publishing, Inc.

Gafos, A. I. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory, 20,* 269–337. http://dx.doi.org/10.1023/A:1014942312445

Gafos, A. I., & Benus, S. (2006). Dynamics of phonological cognition. *Cognitive Science, 30,* 905–943. http://dx.doi.org/10.1207/s15516709cog0000_80

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13,* 361–377. http://dx.doi.org/10.3758/BF03193857

Ganong, W. F., III. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110–125. http://dx.doi.org/10.1037/0096-1523.6.1.110

Garrett, M. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.) *Language production, Vol. 1*: *Speech and talk* (pp. 177–220). London: Academic Press.

Gelb, I. J. (1963). *A study of writing*. Chicago, IL: University of Chicago Press.

Gelfer, C. E., Bell-Berti, F., & Harris, K. S. (1989). Determining the extent of coarticulation: Effects of experimental design. *The Journal of the Acoustical Society of America, 86,* 2443–2445. http://dx.doi.org/10.1121/1.398452

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature, 462,* 502–504. http://dx.doi.org/10.1038/nature08572

Gleason, H. (1955). *An introduction to descriptive linguistics* (Rev. Ed.). New York, NY: Holt Rinehart and Winston.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251–279. http://dx.doi.org/10.1037/0033-295X.105.2.251

Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics, 31,* 305–320. http://dx.doi.org/10.1016/S0095-4470(03)00030-5

Goldsmith, J. (1976). An overview of autosegmental phonology. *Linguistic Analysis, 2,* 23–68.

Goldstein, L. G., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. Arbib (Ed.), *From action to language: The mirror neuron system* (pp. 215–249). Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511541599.008

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition, 103,* 386–412. http://dx.doi.org/10.1016/j.cognition.2006.05.010

Goodman, K. S. (1986). *What's whole in whole language: A parent-teacher guide*. Portsmouth, NH: Heinemann.

Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics, 31,* 423–445. http://dx.doi.org/10.1016/S0095-4470(03)00051-2

Henke, W. L. (1967). Preliminaries to speech synthesis based on an articulatory model. *Proceedings of the IEEE Speech Conference, Boston,* 170–171.

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience, 29,* 2–20. http://dx.doi.org/10.1080/01690965.2013.834370

Hockett, C. (1955). *A manual of phonetics*. Bloomington, IN: Indiana University Press.

Iskarous, K. (2010). Vowel constrictions are recoverable from formants. *Journal of Phonetics, 38,* 375–387. http://dx.doi.org/10.1016/j.wocn.2010.03.002

Jacob, F. (1977). The linguistic model in biology. In D. Armstrong & C. H. van Schoonefeld (Eds.), *Roman Jakobson: Echoes of his scholarship* (pp. 185–192). Lisse, Holland: Peter de Ridder Press.

Jakobson, R. (1970). Linguistics. In J. Havet (Ed.), *Main trends in research in the social and human sciences* (Vol. 1, pp. 419–463). Paris, The Hague: UNESCO/Mouton.

Johnson, K. (2011). Retroflex versus bunched in compensation for coarticulation. *UC Berkeley Phonology Lab Annual Report, 2011,* 114–127.

Joos, M. A. (1948). Acoustic phonetics. *Language, 24*(Suppl.), 5–136. http://dx.doi.org/10.2307/522229

Keating, P. (1990). The window model of coarticulation: Articulatory evidence. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech* (pp. 451–470). Cambridge: Cambridge University Press.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Converging evidence in support of common dynamic principles for speech and movement control. *The American Journal of Physiology, 246,* R982–R935.

Kenstowicz, M., & Kisseberth, C. (1979). *Generative phonology*. New York, NY: Academic Press.

Kingston, J., Kawahara, S., Chambless, D., Key, M., Mash, D., & Watsky, S. (2014). Context effects as auditory contrast. *Attention, Perception, & Psychophysics, 76,* 1437–1464. http://dx.doi.org/10.3758/s13414-013-0593-z

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7,* 279–312.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122,* 148–203. http://dx.doi.org/10.1037/a0038695

Kühnert, B., & Nolan, F. (1999). The origin of coarticulation. In W. J. Hardcastle & N. Hewlett (Eds.) *Coarticulation: Theory, data and techniques* (pp. 7–30). Cambridge: Cambridge University Press.

Kunihira, S. (1971). Effects of the expressive voice on phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior, 10,* 427–429. http://dx.doi.org/10.1016/S0022-5371(71)80042-7

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38. http://dx.doi.org/10.1017/S0140525X99001776

Liberman, A. M. (1970). The grammars of speech and language. *Cognitive Psychology, 1,* 301–323. http://dx.doi.org/10.1016/0010-0285(70)90018-6

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–461. http://dx.doi.org/10.1037/h0020279

Liberman, A. M., Delattre, P., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *The American Journal of Psychology, 65,* 497–516. http://dx.doi.org/10.2307/1418032

Liberman, A. M., Delattre, P., Cooper, F. S., & Gerstman, L. (1954). The role of consonant-vowel transitions in the perception of stop and nasal consonants. *Psychological Monographs, 68,* 1–13. http://dx.doi.org/10.1037/h0093673

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1–36. http://dx.doi.org/10.1016/0010-0277(85)90021-6

Liberman, I. Y., & Liberman, A. M. (1990). Whole language vs. code emphasis: Underlying assumptions and their implications for reading instruction. *Annals of Dyslexia, 40,* 51–76. http://dx.doi.org/10.1007/BF02648140

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18,* 201–212. http://dx.doi.org/10.1016/0022-0965(74)90101-5

Lindblom, B. E. F. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America, 35,* 1773–1781. http://dx.doi.org/10.1121/1.1918816

Lindblom, B. E. F. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Dordrecht, The Netherlands: Kluwer Academic Publishers. http://dx.doi.org/10.1007/978-94-009-2037-8_16

Lindblom, B. (2000). Developmental origins of adult phonology: The interplay between phonetic emergents and the evolutionary adaptations of sound patterns. *Phonetica, 57,* 297–314. http://dx.doi.org/10.1159/000028482

Lindblom, B. E. F., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *The Journal of the Acoustical Society of America, 42,* 830–843. http://dx.doi.org/10.1121/1.1910655

Lindblom, B. E. F., & Sussman, H. M. (2012). Dissecting coarticulation: How locus equations happen. *Journal of Phonetics, 40,* 1–19. http://dx.doi.org/10.1016/j.wocn.2011.09.005

Linell, P. (2005). *The written language bias in linguistics: Its nature, origins and transformations*. Oxford: Routledge. http://dx.doi.org/10.4324/9780203342763

Lotto, A. J., & Holt, L. L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception & Psychophysics, 68,* 178–183. http://dx.doi.org/10.3758/BF03193667

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification.

Perception & Psychophysics, 60, 602–619. http://dx.doi.org/10.3758/BF03206049

Lubker, J. (1981). Temporal aspects of speech production: Anticipatory labial coarticulation. *Phonetica, 38,* 51–65. http://dx.doi.org/10.1159/000260014

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19,* 1–36. http://dx.doi.org/10.1097/00003446-199802000-00001

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. (1978). Bi-alphabetical lexical decision. *Language and Speech, 21,* 142–165.

Lukatela, K., Carello, C., Shankweiler, D., & Liberman, I. Y. (1995). Phonological awareness in illiterates: Observations from Serbo-Croatian. *Applied Psycholinguistics, 16,* 463–488. http://dx.doi.org/10.1017/S0142716400007487

Lundberg, I. (1991). Phonemic awareness can be developed without reading instruction. In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman* (pp. 47–53). Hillsdale, NJ: Lawrence Erlbaum, Associates.

MacNeilage, P., & Ladefoged, P. (1976). The production of speech and language. In E. C. Carteret & M. P. Friedman (Eds.) *Handbook of perception, Volume VII*: *Language and speech* (pp. 76–120). New York, NY: Academic Press.

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 391–409. http://dx.doi.org/10.1037/0096-1523.33.2.391

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics, 28,* 407–412. http://dx.doi.org/10.3758/BF03204884

Mann, V. A. (1991). Are we taking too narrow a view of the conditions for development of phonological awareness? In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman* (pp. 55–64). Hillsdale, NJ: Lawrence Erlbaum, Associates.

Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10,* 29–63. http://dx.doi.org/10.1016/0010-0285(78)90018-X

Martin, J. G., & Bunnell, H. T. (1981). Perception of anticipatory coarticulation effects in /stri, stru/ sequences. *The Journal of the Acoustical Society of America, 69*(Suppl. 1), S92. http://dx.doi.org/10.1121/1.386029

Martin, J. G., & Bunnell, H. T. (1982). Perception of anticipatory coarticulation effects in vowel-stop consonant-bowel sequences. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 473–488. http://dx.doi.org/10.1037/0096-1523.8.3.473

Massaro, D. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

Massaro, D. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

Mattingly, I. G. (1987). Morphological structure and segmental awareness. *European Bulletin of Cognitive Psychology, 7,* 488–498.

McCarthy, J. J. (1982). Prosodic templates, morphemic templates, and morphemic tiers. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations (Linguistic Models 2)* (pp. 191–223). Dordrecht: Foris.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86. http://dx.doi.org/10.1016/0010-0285(86)90015-0

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748. http://dx.doi.org/10.1038/264746a0

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science, 12,* 369–378. http://dx.doi.org/10.1111/j.1467-7687.2009.00822.x

McNeill, D., & Lindig, K. (1973). The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior, 12,* 419–430. http://dx.doi.org/10.1016/S0022-5371(73)80020-9

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science, 30,* 1113–1126. http://dx.doi.org/10.1207/s15516709cog0000_79

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30,* 69–89. http://dx.doi.org/10.1016/0749-596X(91)90011-8

Mitterer, H. (2006). On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception & Psychophysics, 68,* 1227–1240. http://dx.doi.org/10.3758/BF03193723

Morais, J., Cary, L., Alegria, A., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7,* 323–331. http://dx.doi.org/10.1016/0010-0277(79)90020-9

Motley, M. T., & Baars, B. J. (1976). Laboratory induction of verbal slips: A new method for psycholinguistic research. *Communication Quarterly, 24,* 28–34. http://dx.doi.org/10.1080/01463377609369216

Mowrey, R. A., & MacKay, I. R. (1990). Phonological primitives: Electromyographic speech error evidence. *The Journal of the Acoustical Society of America, 88,* 1299–1312. http://dx.doi.org/10.1121/1.399706

Munhall, K. G., Löfqvist, A., & Kelso, J. A. S. (1994). Lip-larynx coordination in speech: Effects of mechanical perturbations to the lower lip. *The Journal of the Acoustical Society of America, 95,* 3605–3616. http://dx.doi.org/10.1121/1.409929

National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Bethesda, MD: National Institute of Child Health and Human Development, National Institutes of Health.

Nooteboom, S., & Quené, H. (2015). Word onsets and speech errors: Explaining relative frequencies of segment substitutions. *Journal of Memory and Language, 78,* 33–46. http://dx.doi.org/10.1016/j.jml.2014.10.001

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognition Psychology, 47,* 204–238. http://dx.doi.org/10.1016/S0010-0285(03)00006-9

Nygaard, L. (2010). *Cross-linguistic sound symbolism*. Presentation at the Conference on Sound Symbolism: Challenging the Arbitrariness of Language, Emory University, March 26, 2010.

Ohala, J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178–203). Chicago, IL: Chicago University Press.

Öhman, S. E. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America, 39,* 151–168. http://dx.doi.org/10.1121/1.1909864

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 309–328. http://dx.doi.org/10.1037/0278-7393.19.2.309

Parry, A. (1971, Ed.) *The making of Homeric verse: The collected papers of Milman Parry*. Oxford: The Clarendon Press.

Perkell, J. S. (1986). Coarticulation strategies: Preliminary implications for a detailed analysis of lower-lip protrusion movements. *Speech Communication, 5,* 47–68. http://dx.doi.org/10.1016/0167-6393(86)90029-4

Perkell, J. S., & Chiang, C. M. (1986). Preliminary evidence for a "hybrid model" of anticipatory coarticulation. *Proceedings of the 12th International Congress of Acoustics,* A3–A6.

Perkell, J. S., & Klatt, D. H. (1986, Eds.). *Invariance and variability in speech processes*. Hillsdale, NJ: Erlbaum.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics, 34,* 516–530. http://dx.doi.org/10.1016/j.wocn.2006.06.003

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13,* 253–260. http://dx.doi.org/10.3758/BF03214136

Pisoni, D. B. (1985). Speech perception: Some new directions in research and theory. *The Journal of the Acoustical Society of America, 78,* 381–388. http://dx.doi.org/10.1121/1.392451

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15,* 285–290. http://dx.doi.org/10.3758/BF03213946

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as asymmetric sampling in time. *Speech Communication, 41,* 245–255. http://dx.doi.org/10.1016/S0167-6393(02)00107-3

Port, R. F. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology, 25,* 143–170. http://dx.doi.org/10.1016/j.newideapsych.2007.02.001

Port, R. F. (2010a). Rich memory and distributed phonology. *Language Sciences, 32,* 43–55. http://dx.doi.org/10.1016/j.langsci.2009.06.001

Port, R. F. (2010b). Language as a social institution: Phonemes and words do not live in the brain. *Ecological Psychology, 22,* 304–326. http://dx.doi.org/10.1080/10407413.2010.517122

Pouplier, M. (2003). *The dynamics of error*. Proceedings of the 15th International Congress of Phonetic Sciences, pp. 2245–2248.

Pouplier, M. (2007). Tongue kinematics during utterances elicited with the SLIP technique. *Language and Speech, 50,* 311–341. http://dx.doi.org/10.1177/00238309070500030201

Pouplier, M., Chen, L., Goldstein, L. G., & Byrd, D. (1999). Kinematic evidence for the existence of gradient speech errors. *Journal of the Acoustical Society of America, 22,* 106, 2242A.

Prince, A., & Smolensky, P. (2008). *Optimality theory: Constraint interaction in generative grammar*. Oxford: Wiley.

Recasens, D. (1984). V-to-C coarticulation in Catalan CVC sequences: An articulatory and acoustical study. *Journal of Phonetics, 12,* 61–73.

Recasens, D. (1985). Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. *Language and Speech, 28,* 97–114.

Recasens, D. (1989). Long range coarticulatory effects for tongue dorsum contact in VCVCV sequences. *Speech Communication, 8,* 293–307. http://dx.doi.org/10.1016/0167-6393(89)90012-5

Recasens, D. (1991). An electropalatalographic and acoustic study of consonant-to-vowel coarticulation. *Journal of Phonetics, 17,* 177–192.

Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *The Journal of the Acoustical Society of America, 125,* 2288–2298. http://dx.doi.org/10.1121/1.3089222

Remez, R. E., Fellowes, J. M., Blumenthal, E. Y., & Nagel, D. S. (2003). Analysis and analogy in the perception of vowels. *Memory & Cognition, 31,* 1126–1135. http://dx.doi.org/10.3758/BF03196133

Roelofs, A. (2014). Integrating psycholinguistic and motor control approaches to speech production: Where do they meet? *Language, Cognition and Neuroscience, 29,* 35–37. http://dx.doi.org/10.1080/01690965.2013.852687

Rubin, P., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. *The Journal of the Acoustical Society of America, 70,* 321–328. http://dx.doi.org/10.1121/1.386780

Sagey, E. C. (1986). *The representation of features and relations in nonlinear phonology*. PhD Dissertation, MA Institute of Technology, Cambridge, MA.

Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1,* 333–382. http://dx.doi.org/10.1207/s15326969eco0104_2

Savin, H. B., & Bever, T. G. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior, 9,* 295–302. http://dx.doi.org/10.1016/S0022-5371(70)80064-0

Schane, S. (1973). *Generative phonology*. Englewood Cliffs, NJ: Prentice Hall. Inc.

Shaiman, S. (1989). Kinematic and electromyographic responses to perturbation of the jaw. *The Journal of the Acoustical Society of America, 86,* 78–88. http://dx.doi.org/10.1121/1.398223

Shankweiler, D., & Fowler, C. A. (2015). Seeking a reading machine for the blind and discovering the speech code. *History of Psychology, 18,* 78–99. http://dx.doi.org/10.1037/a0038299

Shankweiler, D., Strange, W., & Verbrugge, R. R. (1977). Speech and the problem of perceptual constancy. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 315–345). Hillsdale, NJ: Erlbaum.

Shankweiler, D., & Studdert-Kennedy, M. (1967). Identification of consonants and vowels presented to left and right ears. *The Quarterly Journal of Experimental Psychology, 19,* 59–63. http://dx.doi.org/10.1080/14640746708400069

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *Speech production* (pp. 109–136). New York, NY: Springer-Verlag. http://dx.doi.org/10.1007/978-1-4613-8202-7_6

Shattuck-Hufnagel, S., & Klatt, D. (1979). Minimal uses of features and markedness in speech production: Evidence from speech errors. *Journal of Verbal Learning and Verbal Behavior, 18,* 41–55. http://dx.doi.org/10.1016/S0022-5371(79)90554-1

Sheffert, S., & Fowler, C. A. (1995). The effect of voice and visible speaker change on memory for spoken words. *Journal of Memory and Language, 34,* 665–685. http://dx.doi.org/10.1006/jmla.1995.1030

Shtyrov, Y., Pihko, E., & Pulvermüller, F. (2005). Determinants of dominance: Is language laterality explained by physical or linguistic features of speech? *NeuroImage, 27,* 37–47. http://dx.doi.org/10.1016/j.neuroimage.2005.02.003

Silverman, K. (1987). *The structure and processing of fundamental frequency contours*. Unpublished PhD dissertation, Cambridge University.

Squire, L. R. (1986). Mechanisms of memory. *Science, 232,* 1612–1619. http://dx.doi.org/10.1126/science.3086978

Stone, G. O., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language, 36,* 337–359. http://dx.doi.org/10.1006/jmla.1996.2487

Strange, W., Jenkins, J. J., & Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. *The Journal of the Acoustical Society of America, 74,* 695–705. http://dx.doi.org/10.1121/1.389855

Strange, W., Verbrugge, R. R., Shankweiler, D. P., & Edman, T. R. (1976). Consonant environment specifies vowel identity. *The Journal of the Acoustical Society of America, 60,* 213–224. http://dx.doi.org/10.1121/1.381066

Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. In A. Allport, D. G. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production* (pp. 67–84). London: Academic Press.

Studdert-Kennedy, M. (2000). Imitation and the emergence of segments. *Phonetica, 57,* 275–283. http://dx.doi.org/10.1159/000028480

Studdert-Kennedy, M. (2005). How did language go discrete? Tallerman, M. (Ed). *Language origins: Perspectives on evolution* (pp. 48–67). Oxford: Oxford University Press.

Studdert-Kennedy, M., & Liberman, A. M. (1963). Psychological considerations in design of auditory displays for reading machines. *Proceedings of the International Congress on Technology and Blindness, 1,* 289–304.

Trager, G. L., & Smith, H. L., Jr. (1951). *An outline of English structure (Studies in Linguistics, Occasional Papers 3)*. Norman, OK: Battenberg Press.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 13273–13278. http://dx.doi.org/10.1073/pnas.0705369104

van der Hulst, H., & Smith, N. (1982). An overview of autosegmental and metrical phonology. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations (Part I)* (pp. 1–45). Dordrecht, The Netherlands: Foris Publications.

van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition, 15,* 181–198. http://dx.doi.org/10.3758/BF03197716

Viswanathan, N., Fowler, C. A., & Magnuson, J. S. (2009). A critical examination of the spectral contrast account of compensation for coarticulation. *Psychonomic Bulletin & Review, 16,* 74–79. http://dx.doi.org/10.3758/PBR.16.1.74

Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2010). Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1005–1015. http://dx.doi.org/10.1037/a0018391

Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2013). Similar response patterns do not imply identical origins: An energetic masking account of nonspeech effects in compensation for coarticulation. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 1181–1192. http://dx.doi.org/10.1037/a0030735

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40,* 47–62.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language, 68,* 306–311. http://dx.doi.org/10.1006/brln.1999.2116

Warren, R. M. (1971). Identification times for phonemic times of graded complexity and for spelling of speech. *Perception & Psychophysics, 9,* 345–349-389.

Warren, R. M. (1976). Auditory illusions and perceptual processes. In N. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 389–417). New York, NY: Academic Press.

Warren, R. M. (1985). Criterion shift rule and perceptual homeostasis. *Psychological Review, 92,* 574–584. http://dx.doi.org/10.1037/0033-295X.92.4.574

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics, 35,* 49–64. http://dx.doi.org/10.3758/BF03205924

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences, 6,* 37–46. http://dx.doi.org/10.1016/S1364-6613(00)01816-7

Received March 2, 2015
Revision received June 25, 2015
Accepted June 28, 2015 ■