Cognitive Phonetics: The Transduction of Distinctive Features at the Phonology-Phonetics Interface

Veno Volenec & Charles Reiss

We propose that the interface between phonology and phonetics is mediated by a transduction process that converts elementary units of phonological computation, features, into temporally coordinated neuromuscular patterns, called 'True Phonetic Representations', which are directly interpretable by the motor system of speech production. Our view of the interface is constrained by substance-free generative phonological assumptions and by insights gained from psycholinguistic and phonetic models of speech production. To distinguish transduction of abstract phonological units into planned neuromuscular patterns from the biomechanics of speech production usually associated with physiological phonetics, we have termed this interface theory 'Cognitive Phonetics' (CP). The inner workings of CP are described in terms of Marr's (1982/2010) tri-level approach, which we used to construct a linking hypothesis relating formal phonology to neurobiological activity. Potential neurobiological correlates supporting various parts of CP are presented. We also argue that CP augments the study of certain phonetic phenomena, most notably coarticulation, and suggest that some phenomena usually considered phonological (e.g., naturalness and gradience) receive better explanations within CP.

Keywords: phonology-phonetics interface, Cognitive Phonetics, distinctive features, transduction, neurobiology of language

1. Introduction

The aim of the present paper is to elucidate the nature of a cognitive system that takes as its input a representation made of distinctive features (i.e., the output of the phonological module) and generates a representation directly interpretable by the neuromuscular system associated with speech production. This system we will call 'Cognitive Phonetics' and representations[1] it generates 'True Phonetic Representations'. Here we will concentrate solely on speech (pre)production, leaving the perceptual direction of this system aside whenever possible. This inquiry is a resuscitation of certain proposals (see §2) made by Eric Lenneberg 50 years ago, recast in a modern biolinguistic research program advocated

[1] The way we use the term 'representation' here is slightly different than is customary in generative linguistics, where a representation is taken to be an abstract characterization of implicit linguistic knowledge. We use the term in a broader sense, as a scientific abstraction in general, similar to how H_2O 'represents' water in formal stating of chemical processes. The main difference between a surface representation and a true phonetic representation, as will be shown in greater detail in §4, is that the former represents knowledge (competence), and the latter represents information feeding speech production. This more general sense of usage is in line with Marr's (1982/2010: 20) definition of 'representation' as "a formal system for making explicit certain entities or types of information together with a specification of how the system does this".

by David Poeppel and colleagues as an attempt to unify theoretical linguistics and cognitive neuroscience.

Our point of departure is a fairly well-established claim: Surface (also known as 'phonetic') representations of the phonological component of a generative grammar are matrices of distinctive features (where columns represent segments). During most of the 1960s, it was usually assumed that the features of underlying and surface representations are entities of a different kind, the former being binary, the latter gradual scales (Chomsky & Halle 1968: 297). However, one aspect of Postal's (1968) 'naturalness condition' — the statement that a surface representation is identical (and therefore composed from *the same* set of representational elements) to its underlying representation except as requested otherwise by phonological rules — seems to have been, often tacitly, adopted over the following decades, after a brief period of uncertainty. Thus in early 1970s, in an influential compendium on the then pertinent issues in phonological theory, Maran (1973: 73), discussing classificatory (phonological) and phonetic features, wrote in conclusion that "[w]e do not, however, claim at this stage that the set of abstract phonological features is identical in membership to the set of phonetic features. There are many things which remain unclear." But already by late 1970s a consensus seems to have emerged that underlying and surface representations do consist of the same vocabulary of features:

> Assuming that utterances are best represented as a string of feature matrixes at the phonetic level, we can raise the question of how sounds are represented for the purpose of phonological description (i.e., in the UR and at all intermediate levels). (...) [A] fundamental tenet of generative phonology has been that sounds are most properly represented at these levels in the same way they are phonetically—namely, as feature matrixes in which each feature describes an articulatory and/or acoustic property of the sound. (Kenstowicz & Kisseberth 1979: 239)

Given that URs and SRs belong to the same cognitive module, that is, the phonological module, and since a 'module' may operationally be defined as an encapsulated computational system that operates over a particular kind of abstract units (Boeckx 2009: 125–127), it would indeed be most plausible to express all levels of phonological representation with the same set of primitives (Hale & Kissock 2007: 83). Thus the output of the phonological module, the surface representation, is also a matrix of distinctive features.

We understand distinctive features here as a particular kind of substance-free units of mental representation, neither articulatory nor acoustic in themselves, but rather having articulatory and acoustic *correlates*, as Halle (1983/2002: 108–109) and Reiss (2017: §7) have pointed out. Many influential phonological texts have stated over the last several decades that features serve as a bundle of information that the brain sends to the articulators (if speech is the chosen modality). Here are three examples of such statements:

> In articulatory terms each feature might be viewed as information the brain sends to the vocal apparatus to perform whatever operations are involved in the production of the sound, while acoustically a feature may be viewed as the information the brain looks for in the

> sound wave to identify a particular segment as an instance of a particular sound. (Kenstowicz & Kisseberth 1979: 239)
>
> (...) [T]he distinctive features correspond to controls in the central nervous system which are connected in specific ways to the human motor and auditory systems. (...) In producing speech, instructions are sent from higher centers in the nervous system to the different feature boxes in the middle part of (5) ['tone', 'vocal', 'labial' etc. ⟶ vv & cr] about the utterance to be produced. (Halle 1983/2002: 109)
>
> The (...) featurally specified representation constitutes the format that is both the endpoint of perception - but which is also the set of instructions for articulation. (Poeppel & Idsardi 2011: 179)

If one thinks about how exactly features trigger the articulatory system, it becomes apparent that there is a substantial conceptual gap between features and neural structures or activities. At present there is no way to link either the general concept 'distinctive feature' or any of the particular features (e.g., [CORONAL]) to any known neural structure (e.g., dendron, neuron, cortical column etc.) or activity (e.g., long term potentiation, oscillation, synchronization etc.) (Embick & Poeppel 2015). In fact, there seems to be very little understanding of how the brain exactly represents and computes any of the units or processes that are part of linguistic competence (Chomsky 2000a; Gallistel & King 2010; Mausfeld 2012). In other words, the units of linguistic computation and the units of neurological computation — as currently understood — are mostly incommensurable. This problem was therefore dubbed 'the ontological incommensurability problem' by Poeppel & Embick (2005). The proposed solution to it is to decompose a particular linguistic domain (e.g., phonology) into formal units and operations that are as basic and as generic as possible, and then formulate biologically plausible and scientifically productive 'linking hypotheses' across the fields of linguistics and neuroscience (Poeppel & Embick 2005; Poeppel 2012; Embick & Poeppel 2015).

The main goal of this paper is to formulate a hypothesis about the 'intermodular bridge' (Pylyshin 1984: 147) from the symbolic and substance free (phonology) to the physical and substantive (phonetics). By pursuing this line of inquiry a modest attempt is made to formulate a theory of the phonology-phonetics interface in strict biolinguistic terms, that is, in such a fashion that it can be linked to the kind of neurobiological phenomena that we might plausibly find in a neuromuscular system.

Distinctive feature theory was initially outlined by Roman Jakobson in a lecture delivered in 1928 (see Jakobson 1971: 3–6) and in an often overlooked paper from the late 1930s (Jakobson 1939), and subsequently elaborated by Jakobson, Fant & Halle (1952) and Jakobson & Halle (1956). The idea of a 'distinctive feature' was founded upon purely phonological, that is, non-biological and non-cognitive, insights about phonemic oppositions in the vein of Trubetzkoy (1939/1969), as shown in the following passage:

> Any minimal distinction carried by the message confronts the listener with a two-choice situation. Within a given language each of these oppositions has a specific property which

> differentiates it from all the others. The listener is obliged to choose either between two polar qualities of the same category, such as *grave* vs. *acute*, *compact* vs. *diffuse*, or between the presence and absence of a certain quality, such as *voiced* vs. *unvoiced*, *nasalized* vs. *non-nasalized*, *sharpened* vs. *non-sharpened* (*plain*). The choice between the two opposites may be termed *distinctive feature*. The distinctive features are the ultimate distinctive entities of language since no one of them can be broken down into smaller linguistic units. The distinctive features combined into one simultaneous or (...) concurrent bundle form a *phoneme*. (Jakobson, Fant & Halle 1952: 2)

Despite many revisions of the theory during the following decades (e.g., Chomsky & Halle 1968: 298–329; Halle & Clements 1983; Clements 1985; Clements & Hume 1995), it stands to reason that distinctive feature theory was never meant to face one of the more difficult questions of modern biolinguistics and of cognitive neuroscience in general, namely, how to bridge the gap between a cognitive faculty, in this case phonological competence partly represented by features, and brain. The existence of features themselves should not be in question — they have withstood almost a century of rational and empirical scrutiny and are considered "to be a scientific achievement on the order of the discovery and verification of the periodic table in chemistry" (Jackendoff 1994: 60). Also obvious is the fact that features *are* somehow interpreted by the sensorimotor (SM) system because utterances are effectively externalized and perceived/parsed. Therefore, a question that logically follows from these facts is how exactly to get from discrete, timeless, abstract cognitive entities (features) to temporally arranged articulatory movements, and ultimately to continuously varying sound waves.

Here we will adopt the position that cognition, including linguistic cognition, is best understood as a set of modules (see Chomsky 1984 and Curtiss 2013 for justification), each of which is characterized by mappings involving inputs and outputs in a particular format (Reiss 2007: §2.1). Modules are connected via 'interfaces' — configurations in which the outputs of one module serve as the inputs to another module. We argue that the interface between the phonological component of the grammar and phonetics (in this case starting with the neurophonetics of speech production, that is, with sending efferent neural commands to speech organs) is mediated by a system that transduces features into True Phonetic Representations — arrays of temporally coordinated neuromuscular information directly interpretable by the speech production motor system. An assumption that is interleaved in this proposal is that distinctive features, as currently conceived in modern literature, are not *directly* intelligible to the SM system. It is a non-trivial matter to show why this is so, and we return to this issue in §3. Thus our research question is that of transduction of distinctive features at the phonology-phonetics interface, which necessarily precedes speech production. A convenient and productive way to fractionate this question and begin to approach it is to adopt Marr's (1982/2010) three level perspective that specifies — for any cognitive information-processing system — its computational level ('What is computed and why?'), algorithmic level ('How is it computed?'), and implementational level ('How is it realized physically?'). It should be noted that these three levels of analysis *do not* state some fundamental truth about cognitive systems in general (e.g., that every cognitive system consists of three levels); these are explanatory devices, a

convenient way of dividing a cognitive system in order to study it, or in Marr's (1982/2010: 24) words, these are "the different levels at which an information-processing device must be understood before one can be said to have understood it completely". Since the cognitive system under study is essentially an information-processing device, we will frame our discussion in Marr's terms.

The rest of the paper is structured as follows. In §2 we revisit Lenneberg's (1967) *Chapter Three* where he introduces abstract neuromuscular schemata to account for the transformation of basic phonological units, segments in his case, into muscular events. In §3 we state in more detail some general properties of Cognitive Phonetics, our proposed interface theory; we show how it can be constrained by both phonological and phonetic considerations; and we provide arguments for why features need to be transduced before a representation can be legible to the SM system. In §4 we define the transduction of features into True Phonetic Representations following Marr's (1982) tri-level approach and we explore its neurobiological substrate. In §5 we pursue several direct consequences of viewing the phonology-phonetics interface this way and introduce the concept of 'intrasegmental coarticulation'. We conclude (§6) by summarizing our results and by pointing out some further research strategies that follow directly from our insights.

2. Lenneberg's neuromuscular schemata

Lenneberg (1967: 89–90) was well aware of the complexity of the relationship between discrete, logically ordered phonological units (phonemes, segments) on the one hand, and continuous articulatory movements with concomitant acoustic results on the other. He pointed out that although some acoustic discontinuities corresponding to segment transitions are detectable in a spectrogram, in general, these boundaries are not apparent, and the acoustic record of speech provides very limited information about phonological organization. This complexity is of course mirrored in speech production, since discrete sequences of segments correspond to continuous movements of physical systems: "[w]hen we think of the entire musculature of the speech apparatus in activity, we realize that there is a continuous waxing and waning in states of contraction throughout these muscles" (p. 90). The relation between phonological units and articulatory movements is further complicated by various directions, scopes and types of segmental coarticulation: "[t]he muscular activity associated with one phoneme is influenced by the phonemes that precede and follow it" (p. 92). As was already understood at that time, and as subsequent research has confirmed, coarticulation is a ubiquitous phenomenon that obliterates the neat, beads-on-a-string-like succession of phonological segments. A further problem that Lenneberg emphasized is that the order and duration of events at different levels of phonetic organization — perceptual, acoustic, neural — are not perfectly aligned:

> The perceptual order of speech sounds need not be identical with the order of acoustic correlates (we may ignore or fail to hear certain acoustic phenomena); the order of acoustic events need not be identical with the order of motor or articulatory events (movements occur that do not produce sound or sound-changes); the order of central neuronal events

> may be different from the order of peripheral motor events (certain nervous impulses must be initiated in advance of others because traveling time to the periphery is longer for some pathways [e.g., the recurrent nerve supplying the muscles of the larynx — vv & cr] than others [e.g., the trigeminal nerve innervating the muscles of the jaw — vv & cr]). (p. 93)

Lenneberg's discussion illustrates how segmental units of surface representations radically differ from their realizations. The former are discrete, timeless, neatly ordered mental abstractions, the latter continuous, dynamic, overlapping, coordinated movements of respiratory, phonatory and articulatory organs. The magnitude of this mismatch is even greater when we take into account the tremendous complexity of the neuromuscular mechanisms by which mental representations are realized. The production of speech is the most complex neuromuscular activity human beings ever come to master, requiring temporal coordination of over 100 muscles controlled by more than 1400 motor commands per second (Stetson 1951; Lenneberg 1967: 91–92; Laver 1994: 1). Stated this way, it becomes apparent that the mental unit represented as [t] on the one hand, and the sound of producing that unit on the other, are separated by a considerable gap. The problem, then, is to explicitly relate the two sides, taking into account their fundamentally different natures.

Lenneberg (1967: 98–107) proposed a two-step process which, essentially, transmutes segments into real-time muscular activity. A few caveats are due before sketching his proposals. First, Lenneberg's discussion is based on the production of idealized utterances. His examples are not drawn from observed speech, but are models of the process of speech production applied to hypothetical tokens. A related second point is that Lenneberg's proposal is not intended as part of a psycholinguistic theory of language use, what is sometimes called a 'psychologically real' model of speech production. Similar to the components of Marr's (1982) tri-level analysis, the components of Lenneberg's model are "theoretical stages that help us visualize the complications of speech production" (Lenneberg 1967: 99). Third, Lenneberg takes segments, not distinctive features, to be the basic phonological units, and uses a traditional structuralist terminology — 'phonemes' for abstract segmental distinctive units, 'phones' for their intended realizations. One of our primary goals in this paper is to show how Lenneberg's insights can be further developed by combining them with a finer level of phonological representation using distinctive features.

Lenneberg's model, as shown in *Fig. 1*, takes a string of phones as its input and applies two operations: 1. assigns muscle activity to each phone; 2. orders that muscle activity temporally.

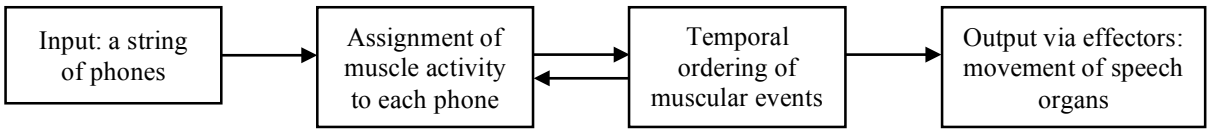

Figure 1. Diagram of a hypothetical transduction processes involved in speech production. Based on Lenneberg (1967: 99).

Both medial processes of *Fig. 1* may be represented in a form of a schema. Lenneberg represented the assignment of muscle activity to each phone with a table where columns stand for successive phones, and rows for muscles relevant for their production (*Fig. 2*).

		string of phones					
		I	II	III	IV	V	VI
muscles	*a*	+	+	+	0	+	0
	b	0	+	+	+	0	+
	c	+	+	0	+	+	0
	d	+	0	0	+	+	+
	e	+	0	+	+	0	+
	f	0	+	0	0	+	+

Figure 2. Schema of the process of assigning muscle activity to a string of phones. Based on Lenneberg (1967: 100).

This schema is intended as a matrix indicating which muscles are to be contracted in order to produce a given speech sound. Rows correspond to specific muscles (abstractly labeled from *a* to *f*), columns to phones; '+' means contraction of a given muscle, '0' means relaxation. For example, the schema in *Fig. 2* indicates that in order to produce phone IV it will be necessary to contract muscles *b*, *c*, *d*, *e*. Naturally, in actual cases of realization of phones, many more muscles are involved. The next step in transduction is to order muscular activity from *Fig. 2* temporally. This process is illustrated in *Fig. 3*. A simplifying assumption is that the relevant muscles may be grouped into classes, here denoted as α through δ, ranked according to the time it takes neural impulses to travel from the brain stem and to reach the muscles in each class. Thus the α class of muscles has an activation latency that is four times greater than the δ class, three times greater than γ, and two times greater than β.

muscles grouped by activation latency				temporal segments								
α	β	γ	δ	1	2	3	4	5	6	7	8	9
			a				+	+	+	0	+	0
b				0	+	+	+	0	+			
	c				+	+	0	+	+	0		
		d				+	0	0	+	+	+	
e				+	0	+	+	0	+			
		f				0	+	0	0	+	+	

Figure 3. Schema of the process of temporal ordering of muscle activity for a given string of phones. Based on Lenneberg (1967: 101).

A further simplification is that in this schema all phones are assumed to be of equal duration.[2] Based on the classification of relevant muscles into latency groups, shown in the left table of *Fig. 3*, the schema from *Fig. 2* is rearranged to obey this relative temporal order. The table on the right in *Fig. 3* shows that if a string of phones I to VI is to be realized correctly, then the first neuromuscular event to occur is the the firing of impulses for contraction of muscle *e*; after that muscles *b* and *c* contract but *e* relaxes, and so on. Due to temporal shifting of the muscles associated with particular phones, the columns in this schema can no longer be put into one-to-one correspondence with the segments in the phonological string. It is here that the phonemic 'Easter eggs' are smashed (Hockett 1955: 210)[3] and coarticulatory effects begin to emerge. Therefore, each column in the right schema of *Fig. 3* corresponds to a 'temporal segment' which indicates, for a given point in time, which muscles need to be contracted or relaxed. Unfortunately, Lenneberg does not discuss the details of this temporal arrangement. For example, he leaves unresolved the question of how much time does one cell denote — 5 ms, 10 ms, 20 ms? Time is represented abstractly in *Fig. 3*, from 1 to 9, a reflection of the hypothetical and tentative nature of his discussion, i.e., "merely stat[ing] what the neuronal firing order is on some given level in the brain" (Lenneberg 1967: 102).

The result of both steps in the transduction of phones into a neuromuscular schema is given in *Fig. 4*. For each unit of time (abstractly denoted here as a temporal segment), the schema specifies which muscle needs to contract and across how many such units, that is, for how long. Within each column, events are assumed to be simultaneous. Notice that for example a_{I} in the fourth temporal segment, which is a muscle contraction associated with the phone ordered first in the string of *Fig. 2*, is preceded by four muscle contractions unrelated to that phone (b_{II}, b_{III}, c_{II}, c_{III}).

---

[2] This is a curious assumption/simplification on Lenneberg's behalf since four pages prior to describing the transduction of segments into neuromuscular schemata he discusses timing problems arising from differences in segmental duration (cf. Lenneberg 1967: 96–97). In fact, temporal discrepancies on various levels of phonetic organization are what initially prompted him to devise such a model of transduction.

[3] "Imagine a row of Easter eggs carried along a moving belt; the eggs are of various sizes, and variously colored, but not boiled. At a certain point, the belt carries the row of eggs between the two rollers of a wringer, which quite effectively smash them and rub them more or less into each other. The flow of eggs before the wringer represents the series of impulses from the phoneme source; the mess that emerges from the wringer represents the output of the speech transmitter. At a subsequent point, we have an inspector [i.e., a hearer — vv & cr] whose task it is to examine the passing mess and decide, on the basis of the broken and unbroken yolks, the variously spread-out albumen, and the variously colored bits of shell, the nature of the flow of eggs which previously arrived at the wringer. Note that he does not have to try to put the eggs together again—a manifest physical impossibility—but only to identify." (Hockett 1955: 210)

temporal segments							
1	2	3	4	5	6	7	8
e_{I}	b_{II}	b_{III}	a_{I}	a_{II}	a_{III}	d_{IV}	a_{V}
	c_{I}	c_{II}	b_{IV}	c_{IV}	b_{VI}	f_{V}	d_{VI}
		d_{I}	e_{IV}		c_{V}		f_{VI}
		e_{III}	f_{II}		d_{IV}		
					e_{VI}		

Figure 4. A neuromuscular schema as a result of transduction of a string of phones into information directly interpretable by the SM system. Based on Lenneberg (1967: 102).

The anticipation of future events emphasizes the need for a model of speech preproduction that feeds the sensorimotor system with "a hierarchic plan in which events are selected (...) as an integration of all elements within units of several seconds duration" (p. 103). For reasons discussed at length (see esp. pp. 102–107), Lenneberg (p. 106) explains that a 'sequential chain model' that scans the surface representation from 'left to right', interpreting linearly ordered segments, is not a viable model for relating phonology to phonetics. Instead, what is needed is a 'central plan model' of speech preproduction, which Lenneberg described as follows:

> On the lowest level, muscular contractions belonging to different speech sounds intermingle and therefore their sequencing cannot be programmed without considering the order of the speech sounds to which they belong. But the choice and sequencing of speech sounds cannot take place without knowledge of the sequence of morphemes to which the sounds belong. [Compare the two different pronunciations of article *the* depending on whether the following morpheme begins with a consonant or a vowel — vv & cv] (...) On the next higher level, the level of morphemes, we encounter again the phenomenon of intermingling of elements and an impossibility to plan the sequence without insight into the syntactic structure of higher constituents. (...) On a still higher level, the level of immediate constituents, (...) syntactic elements cannot be ordered without knowledge of the entire sentence. (p. 106)

The need for a hierarchical central plan for speech production is thus just a specific example of a more general requirement for all levels of linguistic computation and behavior, a requirement that probably extends into other behavioral domains such as navigating through space.

In summary, Lenneberg (1967: §3) recognized the complexities involved in transforming a mental representation of a string of phones into a temporally coordinated sequence of muscular contractions. The result of this transduction may be understood as a neuromuscular schema as in *Fig 4*. The sequential arrangements of muscular events require preplanning with anticipation of later events. Therefore, the occurrence of some events is contingent upon other events yet to come, which may be adduced as proof that sequencing on a neuromuscular level is not accomplished by a sequential chain model (i.e., by scanning and interpreting a string of segments), but rather by a complex central plan model. The

observed interdigitation of muscular correlates of a given phone is mirrored on higher levels of organization, for which a central plan model is also required. The importance of Lenneberg's work, foundational to biolinguistics, derives from his capacity to invoke and synthesize concepts and results from domains as diverse as phonology, phonetics, physiology and neurology.

3. Phonology-Phonetics Interface (PPI)

One of the points that emerged from the previous discussion is that relating phonology and phonetics is a non-trivial and complex task. Lenneberg's views were generally a step in the right direction because he understood the need to explicitly address the conceptual gap between the units and operations characteristic of these two systems. Yet there is room for further improvement by adopting ideas and findings that were mostly unavailable in the 1960s. In particular, the discussion of the phonology-phonetics interface (PPI) can be constrained from 'both sides', that is, by strictly adopting a constrained phonological theory which feeds the interface in production (§3.1), and by using insights from modern models of speech production which are fed by this interface (§3.2).

3.1 Phonology

On the phonological side, we assume a generative substance-free approach (Hale & Reiss 2000a; 2000b; 2008; Reiss 2017; Bale & Reiss 2018). Phonology is understood here as a component of the language faculty that involves formal computations over discrete symbolic units such as distinctive features, syllables, feet etc. Considering phonology is a part of the knowledge of language, it is the case that "all the work in phonology is internal to the mind/brain" (Chomsky 2012: 48). Furthermore, representations involved in phonology are abstract and symbolic, that is, devoid of articulatory, acoustic, typological, statistical etc. information; computations involved in phonology treat features and other phonological units as arbitrary symbols (Hale & Reiss 2008: 169). All representational levels of the phonological component of a generative grammar — underlying, surface, and intermediate — consist of distinctive features (and perhaps markers of other segmental and suprasegmental structure, such as syllable or foot boundaries etc., which need not detain us here). This means that features are part of the 'representational alphabet' of the phonological module. Representational levels are related by ordered phonological rules which serve as the computational aspect of phonology (Vaux 2008).

It is important to distinguish between computation and transduction. Computation is the formal manipulation (reordering, regrouping, deletion, addition etc.) of representational elements *within* a module, and *without* a change in the representational alphabet. Transduction is a process of converting an element in one form into a distinct form, that is, a mapping between dissimilar formats. For example, in the process of hearing, air pressure differentials are transduced into biomechanical vibrations of the tympanic

membrane and the ossicles of the middle ear, which are transduced via the oval window into fluidic movements within the cochlea, which are in turn transduced by the organ of Corti into electrical signals which are passed on for further processing in the nervous system. The distinction between computation and transduction facilitates conceptualizing the notion of modularity. A module can be thought of as a device which takes input representations and computes over them, generating thereby an output in the same representational alphabet. Modules of the mind (and of organic systems more generally) are linked by transducers which convert information in one form into a form required by the computational module fed by the conversion process. An interface between modules is therefore defined by (1) the form of the input, (2) the form of the output, and (3) a set of transformations that relate (1) to (2).

By virtue of the form of its representations and operations, each module imposes 'legibility conditions' at its interface: if some information is to be legible to a given module, that information must come in a specific form in which that module operates (Chomsky 2000a: 9–14). Otherwise, that information would either not be received by that module at all or would be treated as noise,[4] perhaps as human speech is noise to dogs, which lack the needed cognitive modules and transducers, even though their auditory system is far superior to that of humans. The SM system imposes certain legibility conditions to phonology, the component of the grammar with which it interfaces, most notably the condition that information must have a linear arrangement (one cannot produce eleven words in parallel) with certain temporal properties (one cannot produce a polysyllabic word in three nanoseconds). Linearity is a complex notion (see Cairns & Raimy 2011; Idsardi & Raimy 2013). For example, in phonological representations, several tiers may be distinguished (segmental, moraic, prosodic etc.), leading to a kind of multi-linearity characteristic for autosegmental phonology; also, in speech, many overlapping articulatory events may be detected, as will be shown in more detail in §4. Nonetheless, the general idea of linearity, namely, that sequential ordering and precedence relations among basic units play an important role, seems to hold for both phonology and phonetics, unlike for syntax (Chomsky 1995: 334–340; Everaert *et al.* 2015). Another condition, to which we will return in more detail below, is the condition of bi-directionality: if the same phonological architecture is to be employed in both language comprehension and in speaking, that is, if it is *not* the case that humans use completely different grammatical devices for each direction,[5] then the atomic representational units of phonology, features, must somehow integrate acoustic *and* articulatory correlates. If a feature were defined exclusively in terms of, say, its articulatory correlates, as the feature [CORONAL] is, then in principle such a feature could not be used in phonological decoding.

In the phonological theory we adopt, features themselves are substance-free cognitive units (see Reiss 2017: §7 for justification), that is, they do not contain information on the temporal coordination of muscle contractions, on the spectral configuration of the

4 "To be usable, the expressions of the language faculty (at least some of them), have to be legible by the outside systems. So the sensorimotor system and the conceptual-intentional system have to be able to access, to 'read' the expressions; otherwise the system wouldn't even know it is there." (Chomsky 2000b: 17)

5 "The processes of comprehension and production of speech have too much in common to depend on wholly different mechanisms." (Lashley 1951: 186)

acoustic target to be reached, and so on. Yet without this information, the respiratory, phonatory and articulatory systems cannot produce normal speech. The motor system for speech production requires information about substance and time in order to arrange the articulatory score properly, therefore this information has to be integrated into a representation before being fed to the motor system. The most plausible way to escape this deadlock (i.e., phonology is substance free, but the SM system needs information about substance to produce speech) is to abandon the idea of a *direct*, unmediated interface between grammar/phonology and SM system, and posit a cognitive phonetic transduction system that converts distinctive feature matrices into True Phonetic Representations that provide the SM system with legible information needed to produce speech.

In summary:

- Outputs of the phonological module, surface representations (SRs) consisting of substance-free features, do not contain substantial and temporal information.
- The SM system requires articulatory, auditory and temporal information in order to produce speech.

$\therefore$ SRs are not legible to the SM system and phonology cannot in principle feed speech production *directly*.

$\therefore$ Interface between phonology and the SM system is mediated by transduction.

Before turning to the nature of this transduction system, let us review how modern models of speech production further constrain this system.

3.2 Speech Production

On the side of speech production, modern models such as DIVA (Guenther 1995a; 1995b; Guenther *et al.* 1998; 2006; Tourville & Guenther 2011; Guenther & Vladusich 2012), HSFC (Hickok 2012), LRM (Levelt *et al.* 1999; Indefrey & Levelt 2004), and MAPL (Poeppel & Idsardi 2011) provide several theoretical and empirical constraints on the nature of representations that directly feed the SM system during speech. In constructing his model of transduction of phones into neuromuscular schemata, Lenneberg (1967: §3) made the assumption that this process involves reaching specific articulatory targets and took into consideration only the distribution of muscle contractions in time. However, more recent research showed that these targets include auditory information as well. Speech production is a mechanism in which feedforward and feedback processes are tightly and intricately related, as witnessed by the general architecture of the Directions Into Velocities of Articulators (DIVA) model, currently the most elaborate and empirically validated model of speech production (see *Fig. 58.3* of Guenther & Hickok (2016: 728)). Manipulating a speaker's auditory feedback during speech production results in substantial compensatory changes in motor speech acts compared to undisturbed speech (Yates 1963, Guenther *et al.* 1998; Houde & Jordan 1998; Larson *et al*. 2001; Purcell & Munhall 2006; Hickok & Poeppel 2016: §25.2.2.1). For example, if a subject is asked to produce one vowel and the

feedback that she or he hears is manipulated so that it sounds like another vowel, then the subject will change the vocal tract configuration so that the feedback sounds like the original vowel. In other words, speakers will readily modify their articulations to hit an auditory target, suggesting that the goal of speech production involves an intricate relation between articulatory and auditory configurations. Furthermore, although individuals who become deaf as adults can remain intelligible for years after they lose their hearing, they show some speech production impairments immediately, including the inability to adjust pitch and loudness in different listening conditions, and over time they exhibit substantial articulatory decline (Walstein 1990; Perkell *et al.* 2000). The fact that speakers are able to repeat speech acts that they heard, even when given speech acts are *ad hoc* inventions such as "zlurb", suggests that people effortlessly map between articulatory and auditory systems (see the work on the Memory-Action-Perception Loop by Poeppel & Idsardi (2011) for further discussion).

The Hierarchical State Feedback Control (HSFC) model (Hickok 2012) provides further corroboration for the view that features integrate both articulatory and auditory information by showing that speech production involves parallel activation of both auditory and motor units corresponding to the information provided by an appropriate mental representation, and also a sensory-motor coordinate transform network mediating auditory and acoustic programs. It has been well established that surface representations of the phonological module, spelled out in terms of features, serve as both the starting point of speech production and as the end-point of speech perception (Poeppel & Idsardi 2011; Idsardi & Monahan 2016). In an indirect manner, the groundwork for these findings was already laid by the Motor Theory of Speech Perception (Liberman *et al.* 1967; Liberman & Mattingly 1985), which posits that speech perception involves translating acoustic signals into motor gestures that produce them, and by the Acoustic Theory of Speech Production (Fant 1960; Stevens 1998), which highlights the importance of acoustic or auditory targets in the process of speech production. It follows logically from all this that distinctive features allow for mapping from auditory input to words and from words to action, and therefore must properly be defined via abstract articulatory *and* auditory correlates.

Modern neuropsychological and neurophysiological evidence indicates that the cognitive aspect of externalizing language through speech has two distinct stages, phonological and phonetic, lending further support for the necessity of cognitive phonetics as a mediating system between phonology and the SM system. The LRM model, named after its creators Levelt, Roelofs & Meyer (1999), explicates the successive computational stages of spoken word production, and clearly distinguishes between cognitive phonological computation and cognitive phonetic encoding. Indefrey & Levelt (2004) reviewed data from 82 imaging experiments and found that phonological operations are independently conducted within the average time window of 205 ms, *followed by* an average of 145 ms of cognitive phonetic processing. Evidence from aphasia also supports the dichotomy between phonological and phonetic cognitive processing (Buchwald & Miozzo 2011; 2012). Consider the words *pill* and *spill* in English. Both are assumed to contain the segment /p/ in their underlying representations; in the surface representation the former has [p^h] and the latter [p]. It is of interest to determine what exactly happens when

an aphasic patient simplifies a consonant cluster so that /s/ does not get realized in a word like *spill*. Will the resultant realization of /p/ be aspirated, consistent with the notion that the deletion of /s/ occurred within the phonological module (i.e., before motor plans for a cluster are implemented), or will it be produced without aspiration, reflecting the conception that the phonological mapping /sp/ → [sp] was left intact and that the deletion of the fricative occurred after phonological computation? Buchwald & Miozzo (2011) measured VOT productions of two aphasic patients who did not realize /s/ in /sp/, /st/, /sk/ clusters and compared these with realizations of correctly produced consonants. Results showed two different patterns of production, with one patient producing the initial stop consonant with a long VOT ([p^h]), and the other producing it with a short VOT ([p]). These findings have been taken to suggest that the errors of the former patient were phonologically based and the errors of the latter patient were phonetically based and "are consistent with an account of spoken production containing at least two processing levels that can be selectively impaired by brain damage: one processing stage [i.e., cognitive phonological] with context independent representations and another [i.e., cognitive phonetic] with context-specific representations" (Buchwald & Miozzo 2011: 1118). Similar results emerged in examination of durational properties of nasal consonants when deleted in /sn/ and /sm/ clusters (Buchwald & Miozzo 2012).

In summary, modern research into speech production, and to a lesser extent speech perception, constrain the PPI insofar as they show (1) that the target of speech production is a complex representation that integrates both articulatory and auditory information; (2) that speech production is strongly influenced by auditory and somatosensory feedback; (3) that features have abstract articulatory and acoustic correlates, as demanded by (1) and (2); (4) that cognitive aspects of externalizing language through speech have two distinct stages: a substance free computational stage (phonology) and a substantial transduction stage (cognitive phonetics).

3.3 An Interface Theory: Cognitive Phonetics

Cognitive Phonetics (CP) is a theory of the phonology-phonetics interface (PPI). It is motivated by the conceptual distance between the characteristics of phonology as shown in §3.1 on the one hand, and the characteristics of the speech production mechanism as shown in §3.2 on the other. CP proposes that the output of the grammar is transduced into a representation that contains substance-related information required by the SM system in order to externalize language through speech. *Fig. 5* illustrates the general architecture of the PPI and the place of CP within it.

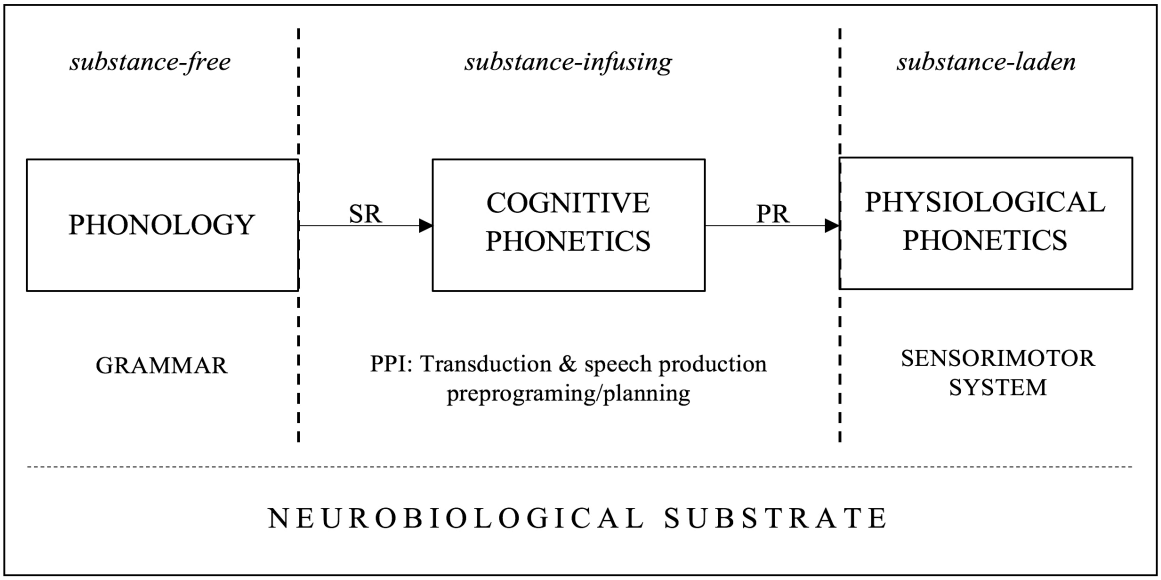

Figure 5. The architecture of the phonology-phonetics interface and the place of cognitive phonetics within it.

Recall that our present focus on speech externalization, without discussion of speech perception and phonological comprehension, is a matter of expository convenience, not a claim about the purview of CP. As the interface between phonology and phonetics, CP is a bi-directional system, thus also relevant for transduction in the direction of perception, that is, for decomposing, parsing, and mentally representing the sound of speech (Reiss 2007: §2.5; Poeppel *et al.* 2008). Therefore, in the 'input' direction, CP serves as "the bridge from the physical to the symbolic" (Pylyshyn's 1984: 152). In the 'output' direction, which is our focus here, CP is the bridge from the symbolic to the physical, relating the substance-free (phonology) to the substance-laden (physiological phonetics).

CP is fed by the output of the phonological grammar, and directly feeds the sensorimotor (SM) system associated with speech production. CP is substance-infusing in the sense that it provides the means to externalize language through speech in real time using human neurophysiological machinery. The movements of various organs and the subsequent acoustic consequences comprise the substance-laden aspect of speech traditionally associated with articulatory and acoustic phonetics. CP is a transduction system, which means it changes inputs of one ontological type into outputs of another. The input to CP is a mental representation comprised in part of abstract distinctive features. The output is a representation that contains information on the auditory target to be reached, the muscles necessary to realize a given input, and their temporal arrangement. Outputs of phonology are interchangeably called in the literature 'surface' representations and 'phonetic' representations, while representations from which these are derived are called 'underlying' or 'phonological' representations (Kenstowicz 1994: 60). Since both are *phonological* representations, that is, encoded in the primitives of the phonological module, it is misleading to call only one representational level phonological. Therefore, in line with

our ideas regarding the PPI and CP, we propose a terminological clarification. Inputs to phonology, typically conceived of as strings of concatenated morphemes, we will call 'underlying phonological representations' (UPR); outputs of phonology, which are the inputs CP, will be called 'surface phonological representations' (SPR); and outputs of CP 'true phonetic representations' (TPR); or, for short, 'underlying representations' (UR), 'surface representations' (SR), and 'phonetic representations' (PR), respectively. URs and SRs are part of phonology; PRs are in principle extragrammatical, non-phonological entities.

It is an understatement to say that progress in solving the ontological incommensurability problem in all cognitive domains has been modest. In this light, the fact that we are still talking about theoretical abstractions (e.g., PRs) and not solely in terms of neurobiological processes does not reflect a commitment to any sort of dualism. It reflects instead the position that theoretical cognitive models are crucial for understanding neurobiology of any cognitive domain, including language (Gallistel – King 2010; Poeppel 2012). However, provided that we decompose models of various aspects of cognition — language and speech programing included (Boeckx *et al.* 2014) — into elementary units and operations, it is a logical necessity that for these units and operations to be 'real' in any coherent sense of that word, they must have a neurobiological substrate, as reflected by *Fig 5*. For phonology, works like *Phillips et al.* (2000), Binder *et al.* (2000), Hickok & Poeppel (2000a; 2004; 2007), Indefrey & Levelt (2004), Obleser *et al.* (2004), Mesgarani *et al.* (2008; 2014), Idsardi & Raimy (2013), Monahan *et al*. (2013), Idsardi & Monahan (2016) provide some useful information on what this substrate might be and how to look for it. For neurobiological substrate of cognitive aspects of speech perception and production see Hickok & Poeppel (2000b; 2016), Poeppel *et al.* (2008), Poeppel & Hackl (2008), Poeppel & Monahan (2008), Poeppel & Idsardi (2011), Blumstein & Baum (2016); Guenther & Hickok (2016); Tremblay *et al.* (2016). The neurobiological substrate for CP will be explored in §4.

CP shares its name and some conceptual commitments with the theory of cognitive phonetics by Tatham (1984; 1987; 1990) and Morton (1987), although there are substantial differences. While both approaches reject the notion of a direct interface between phonology and phonetics, and argue for a cognitive approach to certain phonetic phenomena, their theory (henceforth 'CP-TM') offers a different view of what phonology is and how it works. Although CP-TM was somewhat sympathetic to contemporary developments in generative phonology (Tatham 1990: §3.1), the most important difference from our approach is that CP-TM did not fully commit to the generative architecture of the human language faculty, and therefore did not inherit all the implications (and results) that the generative framework entails. In particular, while CP-TM acknowledge the existence and *phonological* importance of features (*ibid.*), as soon as they reach the phonetic level (albeit a cognitive one), they, like most phonetic models, tacitly shift attention to the realization of segments (Tatham 1990: §6). In contrast, we are interested in decomposing SRs into phonological primitives, features, and in exploring how these might be implemented neurobiologically in real time. A further difference is that CP-TM has no commitments to neurobiology and keeps the discussion strictly in the cognitive domain. In

fact, CP-TM resolutely banishes neurobiological considerations and maintains a curious form of an "extreme dualist view" (Tatham 1990: 11).

The positing of a cognitive aspect of phonetics in no way blurs the competence/performance distinction. Phonology is competence; phonetics, even its cognitive aspect, is performance by definition, since only mental grammar is defined as competence. The transduction process modeled by CP (§4) does not entail 'knowledge' (e.g., 'knowing how' to produce speech) in any useful sense of the word (see Chomsky (1980: 101–102) for a relevant discussion on this matter). Transduction of SRs into PRs entails a set of neuromuscular skills. Its ontogenetic development most likely follows the development of performance systems in general (Lenneberg 1967: §4.II). These skills are most properly conceived as 'automatic synergisms', "whole trains of events that are preprogrammed and run off automatically", and that "form the basis of all motor phenomena in vertebrates" (Lenneberg 1967: 92; see also Lorenz & Tinbergen (1957; 1970) for the seminal investigation of innate egg rolling automatisms in greylag geese). That they are cognitive, at least partially, despite being part of performance should also not be controversial.[6] CP by definition has access to cognitive representations generated by phonology, as shown by the left portion of *Fig. 5*, and it is in this respect that the epithet 'cognitive' is justified; what CP generates, phonetic representations (PFs), are instructions for the SM system on how to execute neuromuscular commands, obviously no longer cognitive. One of the main characteristics of a transducer is that it changes the format of its input, and in our case the input is a cognitive entity.

4. The Inner Workings of CP: Transduction

In this section, we turn to the primary research question for Cognitive Phonetics (CP): How are phonological features related to human neurobiological structures? In other words, how can we bridge the symbolic and the physical in the domain of speech? As we have indicated, this means exploring the structure of the transducer that converts SR-type information into PR-type information. Clearly, our chances of understanding a transducer are better if we have a good understanding of the transducer's inputs and outputs. The relatively robust results of generative phonology, as compared with other domains of cognition, provide us with an anchor for such explorations — we have a fairly explicit model of the nature of SR-type information as linearly ordered strings of feature matrices. Models of comparable detail are not available for the other two aspects of CP, the transduction procedures and PR-type information, and it is to those topics that we now turn.

Marr's (1982/2010) tri-level theory, which we will adopt in further discussion, has been widely accepted as a means to gain insight into information processing systems (IPS) such as CP. Marr proposes that IPSs are best analyzed in terms of three conceptual levels, each corresponding to a specific set of questions. These levels include the 'computational level',

[6] It should be noted that this is mostly a definitional matter; by 'performance' in this context we merely mean 'not grammar'.

the ‘representational and algorithmic level’, and the ‘implementational level’ (p. 22–27), defined by the following questions:

- Computational level: What does the process do? Why does the process do it?
- Representational/algorithmic level: How does the process work? In particular, what are the input and output representations and what is the algorithm for the transformation?
- Implementational level: How are the output representation and the algorithm realized physically? In particular, what is the neurobiological substrate of the mapping in question?

Before proceeding, let us clarify a confusing terminological ambiguity. The fact that we are describing transduction, as distinct from computation, and yet still can talk about the *computational level* of a transducer does not reflect an intellectual inconsistency, but rather just two different uses of a term. As was stated in §3.1, the main difference between a computational module and a transducer is that the former is a mapping between entities in the same format (e.g., feature matrices to feature matrices), and the latter is a mapping between entities of dissimilar formats (e.g., feature matrices to muscle commands, or sound vibrations to neural impulses). However, both modules and transducers are IPSs, therefore both are amenable to Marr’s tri-level analysis, and both can be analyzed at the *computational level* in Marr’s sense.

So, what implications does Marr’s theory have for our research question? First, it calls for maximal conceptual decomposition of the representations and operations posited by linguistics. For a long time, the cognitive neuroscience of language was (and to a certain extent perhaps still is) focused on exploring the neurobiological correlates of rather complex linguistic entities or domains, such as syntax (so for example, “Broca’s area underlies syntax” would be a common assertion in such a tradition), phonology, lexical semantics, and so on (Poeppel 2012: 36–49). However, Marr (1982/2010) argued that IPSs are best studied by decomposing them into representational and computational primitives, and then by building a bottom-up understanding of them. It is partly from this method that the success of his theory of vision derives, and it is a success that has inspired much of the recent work in computational neuroscience of language. Second, Marr’s theory encourages us to seek an explanation for an IPS’s nature from several different sources (for example, linguistics, cognitive science more broadly, neurobiology, formal computational theory) and facilitates explicitly connecting cognitive primitives with neurobiological structures. Therefore, it serves as a general framework for positing linking hypotheses across the fields of linguistics and neurobiology.

Let us now turn to defining transduction — the operational aspect of CP — at the phonology-phonetics interface in terms of these three levels. Firstly, we want to address the ‘what’ and ‘why’ questions of the computational level. What does transduction in CP do? It transforms a representational format that is necessary for the coding of phonological knowledge into a representational format adequate for instructing the neuromuscular system on what it must accomplish in articulatory terms. Why does CP carry out transduction? In general, the answer to this question follows directly from the theoretical

and empirical considerations of §3.1, namely, that outputs of phonology, SRs consisting of substance-free features, lack crucial substantial and temporal information and are thus not legible to the SM system; therefore, phonology cannot in principle feed speech production *directly*, but only through transduction. The very fact that phonology and phonetics constitute two distinct domains that share an interface logically implies the necessity of transduction between them. In the absence of CP, a mental expression could not be externalized through the human SM system. The transduction maps between properties of the mind — mental representations composed at the most basic level of discrete, timeless, symbolic elements — and the functioning of the motor system, which works in terms of gradual, dynamic, temporally arranged neuromuscular activity. Since we *do* speak, the existence of transduction is confirmed.

We now turn to the representational and algorithmic level of transduction.[7] The first step at this level is to state the representations involved in transduction. The input representation, SR, is a matrix of distinctive features. Each feature is transduced and receives interpretation by the SM system. Features[8] are elementary units of phonological computation, stored in long term memory, that represent articulatory and acoustic information in a highly abstract manner. Each feature may abstractly be schematized as shown in *Fig. 6*, which is an extension of the Memory-Action-Perception Loop of Poeppel & Idsardi (2011). The input representation thus involves a set of idealized acoustic targets at which the neuromuscular system will aim, as corroborated by studies discussed in §3.2, and a set of idealized articulatory configurations needed to achieve these goals. It should be emphasized that these 'targets' are not precise, physically invariant acoustic measurements, as features are substance-free units; they are coarse mental representations of acoustic spaces. It is an basic finding of psychoacoustic phonetics that what a speaker deems a repetition of the same category may in fact reflect a wildly different acoustic signal (Liberman 1957). The cognitive unity between acoustic and articulatory correlates of features seems to be so strong that hearing someone utter something excites a corresponding motor program, regardless of whether the hearer has the intention to speak (Cooper & Lauritsen 1974; Fadiga et al. 2002).

---

[7] Here we will make two simplifying assumptions. We will assume that features within a single bundle (segment) are parts of an unordered and unstructured set and are not grouped hierarchically so as to mimic the composition of the vocal apparatus. We will also abstract away from the possibility, strongly suggested by evidence presented in Hale & Kissock (2007), that featurally underspecified segments persevere into SRs. Integrating perseverant underspecification into CP will be left aside for future research.

[8] It is doubtful that current expositions of the universal set of features in linguistic literature are either quantitatively or qualitatively adequate. Compare, for example, Kenstowicz & Kisseberth (1979: 241–253), Lass (1984: 82–93), Katamba (1989: 42–51), Carr (1993: 54–66), Gussenhoven & Jacobs (2011: 74–84), Odden (2013: 45–61), Zsiga (2013: 258–270) etc., and notice the tremendous differences in the total number of features, in the way they are classified, in the set of features that made it to the final list, in the assumptions about *n*-arity, and especially in their definitions.

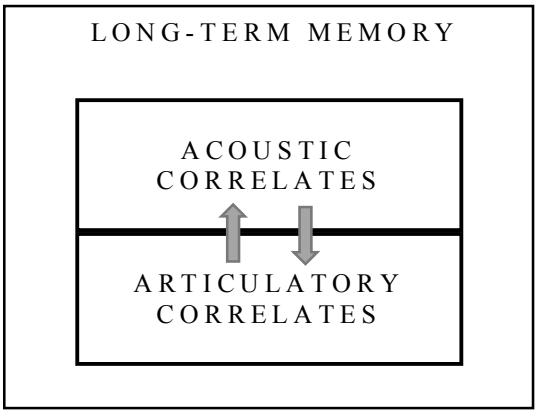

Figure 6. A schematization of a distinctive feature. Features serve as the cognitive basis of the bi-directional translation between speech production and perception, and are part of the long-term memory representation for the phonological content of morphemes, thus forming a memory-action-perception loop (Poeppel & Idsardi 2011) at the lowest conceptual level.

The output representation, called 'True Phonetic Representation', or 'Phonetic Representation' (PR) for short, is a complex array of neural commands that activate muscles involved in speech production. As pointed out in §2, uttering even a single syllable involves hundreds of neuromuscular connections, therefore a detailed description of every neuromuscular event for every single and interacting feature is far beyond the scope of this paper.

Our modest goal here is to sketch the fate of a transduced feature in a few simple and idealized cases. Take, for example, the feature [+ROUND]. Since lip rounding is known to have systematically varying muscular expression (due to interaction with other features, to which we will return below), the Phonetic Representation (PR) has to allow for this variation across contexts. The transduced form of [+ROUND], call it $PR_{[+ROUND]}$, involves at least four muscles: *orbicularis oris*, *buccinator*, *mentalis, levator labii superioris*. The idealized expression, assuming no directly interfering articulatory movements (a relatively rare case in actual speech), is simultaneous contraction of the superior and inferior parts of *orbicularis oris*, contraction of *mentalis* (for protruding the lower lip) and *levator labii superioris* (for protruding the upper lip), and relaxation of *buccinator*. This is the case observed in pronouncing [u]. In [y], on the other hand, $PR_{[+ROUND]}$ in addition to contracting the aforementioned muscles also involves a compressing movement (lips drawn together horizontally) caused by the contraction of the *buccinator*. The difference between protrusion and compression in $PR_{[+ROUND]}$ is dependent on whether $PR_{[+ROUND]}$ is interacting with $PR_{[+BACK]}$ or $PR_{[–BACK]}$ (Catford 1982: 172–173). Of course, various other complications exist, but this suffices to illustrate the general idea. The exact and fully detailed characterization of $PR_{[+ROUND]}$ will thus be possible only after thoroughly studying various possible interactions of transduced features, no doubt a massive phonetic undertaking.

Note that PRs are still *abstractly* related to speech; they are not hi-fi encodings of speech-sound articulations, although they are less abstractly related to speech than SRs. This is because what is actually externalized is further complicated by a great number of

factors. As Hale & Kissock (2007: 85) point out, transduction is followed by other performance factors that have no bearing on neither grammar nor transduction, factors like speech rate, loudness, interruptions due to sneezing, and many other situational effects. We will also have nothing to say here about how other aspects of SRs (e.g., prosodic elements like tone) are transduced.

The algorithm that transforms SRs into PRs has two steps, echoing Lenneberg's (1967) proposals outlined in §2. In the first step (A_1), a feature is related to muscles which need to be contracted in order to produce an appropriate acoustic effect. Since speech occurs in real time, the second step (A_2) will entail temporal coordination of muscular activity demanded by A_1. A tremendous amount of complexity arises in relation with the second step of transduction. The main resultant phenomenon of this step is coarticulation (see Hardcastle & Hewlett (1999), Farnetani & Recasens (2013) and Volenec (2015) for surveys) — temporal overlapping of various aspects of PRs. Neurobiological studies on speech perception have uncovered that the human perceptual system consistently uses two time scales to analyze a continuous speech signal, a segmental time-frame of roughly 10 – 80 ms, and a syllabic time-frame of 100 – 500 ms (Poeppel *et al.* 2008; Poeppel & Idsardi 2011; Chait *et al.* 2015):

> There are two critically important chunk-sizes that seem universally instantiated in spoken languages: segments and syllables. Temporal co-ordination of distinctive features overlapping for relatively brief amounts of time (10 – 80 ms) comprise segments; longer coordinated movements (100 – 500 ms) constitute syllabic prosodies. (Poeppel & Idsardi 2011: 182)

However, transduced features often 'spill over' these temporal borders, crossing segmental and sometimes even syllabic boundaries in both directions, thus leading to coarticulation. Our decision to examine the transduction of [+ROUND] to $PR_{[+ROUND]}$ is useful since this aspect of speech relies on several muscles and is known to show great propensity for temporal overextending, especially in the anticipatory direction. Lisker (1978: 133) states that "lip-rounding and nasalization are segmental features of English that refuse to be contained within their 'proper' segmental boundaries, as these are commonly placed". (Note that Lisker's example should not be specific to English if it derives from universal transducer properties.) Likewise, according to Benguerel & Cowan (1974) $PR_{[+ROUND]}$ may be evident several consonants in advance of the rounded vowel for which it is required: in French, labial coarticulation can extend up to 6 segments in the anticipatory direction. Lubker *et al.* (1975) showed, using electromyography, that in Swedish $PR_{[+ROUND]}$ can start up to 600 ms ahead of a rounded vowel. Both directions of temporal overextending of $PR_{[+ROUND]}$ are observed in English, as demonstrated by Laver's (1994: 321) clever example [$h^wud^wtʃ^wuz^wp^wɹ^wun^wdʒ^wus^w$] (*Who'd choose prune juice?*). While our step A_2, the temporal distribution of muscular activity, can obscure segmental and syllable boundaries, phonological word boundaries seem to be universally impenetrable by it, thus imposing the upper limit to its effect. The neurobiological mechanisms underlying transduction algorithms are universal properties of the human species, as witnessed by the fact that humans, in all non-pathological cases, use them without fail (see Dronkers (1996)

for an example of a pathological case demonstrating a disruption of A_1). However, although the transduction algorithms are biologically universal in humans, CP will still show great output variability due to these two transduction steps being used on language-specific SRs.

The implementational level is concerned with the neurobiological substrate of CP (see *Fig. 5*). How is transduction of features at the PPI instantiated in the human brain? Many mysteries still surround this question and proposed answers are ever-changing. At a relatively gross neuroanatomical level, speech production engages a widely distributed neural network. In a meta-analysis of overt speech production, Eickhoff *et al.* (2009) reported consistent activation in left inferior frontal gyrus (IFG), ventral precentral gyrus (motor and premotor cortex), ventral postcentral gyrus (somatosensory cortex), superior temporal gyrus (STG) (auditory cortex), supplementary motor area (SMA), anterior insula, superior paravermal cerebellum (lobules V and VI), basal ganglia and thalamus. Of particular importance for transduction is the 'dorsal stream', usually stated to have an "auditory-motor integration function" (Hickok & Poeppel 2007: 394) and to be "involved in mapping sound representations onto articulatory-based representations" (Hickok & Poeppel 2004: 72). The dorsal stream is comprised of structures in the posterior frontal lobe and the posterior dorsal-most part of the temporal lobe and parietal operculum. The dorsal stream is strongly left-dominant, which is why production deficits result predominantly from dorsal temporal and frontal lesions. The specifics of these general findings lend support for various aspects of CP.

The articulatory motor programs for executing features are coded in posterior IFG of the left hemisphere, traditionally known as Broca's area. More specifically, Hickok (2012: 138) reports that pars opercularis (BA44) and the ventral-most part of BA6 store articulatory programs needed to reach the auditory targets imposed by features. BA44 and BA6 are thus the most likely candidates for storing articulatory aspects of features (see *Fig. 6*). The anterior insula, a cortical area beneath the frontal and temporal lobes of the left hemisphere, is reported to be involved in preparation of speech, that is, in "translating a phonetic 'concept' obtained from left IFG into articulatory motor patterns" (Blumstein & Baum 2016: 649; Eickhoff *et al.* 2009), roughly corresponding to our A_1. Dronkers (1996) showed that lesions to that part of the brain lead to apraxia of speech, the inability to assign muscular activity to a phonological representation. Dronkers' results are rather robust and show a clear disruption of A_1, since all 25 examined stroke patients suffering from apraxia of speech had the same lesion, while the anterior insula was spared in all 19 healthy participants. By way of the dorsal stream, information from the anterior insula is transmitted to the pre-SMA, often implicated in articulatory initiation and sequencing of neuromuscular activity (Alario *et al.* 2006; Guenther *et al.* 2006; Bohland & Guenther 2006), and then projected to the primary motor cortex. The pre-SMA also receives temporal information from the cerebellum and the basal ganglia (see below). It can therefore be hypothesized that the pre-SMA integrates information from A_1 and A_2, and forms a finalized True Phonetic Representation. From the primary motor cortex, neurons send signals to the brainstem and spinal cord that ultimately result in muscle contractions.

Important structures for the temporal organization of speech (corresponding to A_2) include the cerebellum and basal ganglia. Information from the insula (corresponding to A_1) is directly transmitted to the cerebellum and basal ganglia, structures that are well-

established constituents of cortical-subcortical loops for movements preparation (Jueptner & Krukenberg 2001). More specifically, selection and sequencing of motor programs for articulation is mediated through basal ganglia, and the conversion of the discretely prepared sequences into a fluent, temporally distributed action is carried out by the cerebellum (Eickhoff *et al.* 2009: 2416). Cerebellar dysfunction affects temporal aspects of speech production and results in a dysarthria characterized by improper timing of cognitively discrete elements (such as feature bundles), substantial aberrations in their total and relative duration, disrupted coordination of orofacial and laryngeal movements, slowed/delayed execution of articulatory movements etc. (Ackerman *et al.* 2007). Information from the cerebellum and basal ganglia ties into the pre-SMA, presumably where A_1 and A_2 are integrated to form a True Phonetic Representation directly interpretable by the primary motor cortex (PMC) which sends efferent neuromuscular commands.[9]

Features also have acoustic correlates (see *Fig. 6*) that serve as targets for articulatory movements. There is accumulating evidence and a convergence of opinion that portions of the superior temporal sulcus (STS) — bilaterally but perhaps with a mild leftward bias — are important for encoding acoustic/auditory aspects of phonological representations (Indefrey & Levelt 2004; Buchsbaum *et al*. 2001). In an attempt to pinpoint this region more narrowly, Hickok & Poeppel (2007: 398) "suggest that the crucial portion of the STS that is involved in phonological-level processes is bounded anteriorly by the most anterolateral aspect of Heschl's gyrus and posteriorly by the posterior-most extent of the Sylvian fissure". Mesgarani *et al.* (2014) showed that acoustic phonetic information is represented in the STS and is distributed along 5 distinct areas, each roughly corresponding to a general 'manner of articulation' class of speech sounds. By measuring the responses in implanted electrical cortical grids placed along the superior-most part of the temporal gyrus, they found that their electrode e1 responded selectively to stops, e2 to sibilant fricatives, e3 to low back vowels, e4 to high front vowels and a palatal glide, and e5 to nasals (Mesgarani *et al*. 2014: 1009). Similarly, Bouchard *et al*. (2013) constructed an auditory-based 'place of articulation' cortical map in the STG, confirming labial, coronal and dorsal 'places' with different electrodes, and cutting across various manner classifications. Scharinger *et al*. (2012) found, using magnetoencephalography, neural correlates of three phonologically relevant vowel variables — height, frontness and roundness spelled in terms of first three formants — again localizing them in the superior temporal gyrus.

STS and STG project auditory representations to an area in the Sylvian fissure at the boundary between the parietal and temporal lobes (called 'Spt'), where they are integrated with articulatory representations (Hickok *et al.* 2009; 2011; Gow 2012). Activity in Spt is highly correlated with activity in the pars opercularis (Buchsbaum *et al.* 2001; 2005), the posterior sector of Broca's region implicated in storage of articulatory motor

[9] "The basal ganglia and the cerebellum both forward their information to the PMC which precedes M1 in a serial fashion. The parallel engagement of the subcortical motor loops is thus followed by a sequentially organized common final pathway: the PMC first combines the processed information about selected movement programs and their temporal sequencing provided by the basal ganglia and the cerebellum, respectively, into *a final movement representation*. These are then forwarded to M1 for the generation of the final output to lower motor neurons and hence execution." (Eickhoff *et al.* 2009: 2416; our emphasis)

programs. White matter tracts identified via diffusion tensor imaging suggest that Spt and the pars opercularis are densely connected neuroanatomically (Hickok *et al*. 2009). Spt therefore appears to be involved in sensorimotor integration, that is, in translation between auditory and articulatory correlates of features.

At the beginning of this section, we have stated that the main goal of this paper is to gain a better understanding of how phonological features are related to human neurobiological structures. Let us summarize our findings. Recent neuroscience evidence is consistent with the idea that Cognitive Phonetics transduces abstract features (elements of SRs) into temporally distributed neuromuscular activities (elements of PRs), relating the phonological grammar to the vastly different SM system. This is carried out by assigning each feature a specific set of muscular contractions (A_1) and by ordering them temporally (A_2). Transduction is implemented by a widely distributed neural network which engages the inferior frontal gyrus (stores articulatory correlates), the superior temporal gyrus (stores acoustic correlates), the Spt (sensorimotor integration), the anterior insula (A_1), the cerebellum and basal ganglia (A_2), the supplementary motor area (integrates A_1 and A_2), and the primary motor cortex (sends efferent neural commands to the muscles).

5. Implications

We have stressed the importance of adhering to **phonological** facts in **phonetic** theorizing because decisions made on phonological grounds will have considerable impact on phonetic analysis. In particular, this means that we take serious consideration of the following notions: (1) the most basic unit of phonology is the distinctive feature; (2) features are abstract, cognitive, substance-free units; and (3) features are transduced at the phonology-phonetics interface (PPI) by being converted into temporally coordinated muscular activity. Several theoretical and empirical implications follow from Cognitive Phonetics (CP), our theory of this interface.

5.1 Coarticulation

The concept of coarticulation, such as the lip rounding during production of [s] before the rounded vowel of *soon*, rests upon two premises: (a) that discrete units, segments, underlie the continuous, gradient speech signal (Hammarberg 1976: 357),[10] and (b) that these segments are converted into articulatory gestures (Farnetani & Recasens 2013: 317f). The temporal overlapping of articulatory gestures pertaining to different linearly ordered segments can thus be dubbed 'intersegmental coarticulation'. However, if premise (a) is modified to be in line with much of modern phonology (see §3.1), that is, if the

[10] Even Carol Fowler, who disagreed with Hammarberg on many issues related to coarticulation (see Fowler 1983) and who later argued for a gesture-based account of coarticulation (see Fowler & Saltzman 1993), stated that "an intuitive concept of 'segment' underlies our recognition that there is a phenomenon of coarticulation requiring explanation." (Fowler 1980: 114)

phonological feature is taken as the atomic underlying unit, it follows that (c) features are converted into something more basic than segment-bound articulatory gestures (see §4), and (d) that interaction in realization of features *within* a single segment is also possible, leading to what we will call 'intrasegmental coarticulation'. Here we will briefly sketch the consequence of approaching coarticulation from the framework of CP, assuming (c) and (d) instead of (just) the usual (a) and (b).

CP performs the mapping SR → PR, or, in terms of individual features, [F] → $PR_{[F]}$. We will therefore take transduced features (in a general format $PR_{[F]}$, where [F] stands for an individual feature) to be the basic units that enter speech production. To illustrate intrasegmental coarticulation, consider the interaction of $PR_{[HIGH]}$ and $PR_{[NASAL]}$ observed, for example, in Lakhota (Boas & Deloria 1941), Yoruba (Ogunbowale 1970) and Koyra Chiini (Heath 1999), with sketches in *Fig. 7* based on Beddor (1983) and Ladefoged & Johnson (2010).

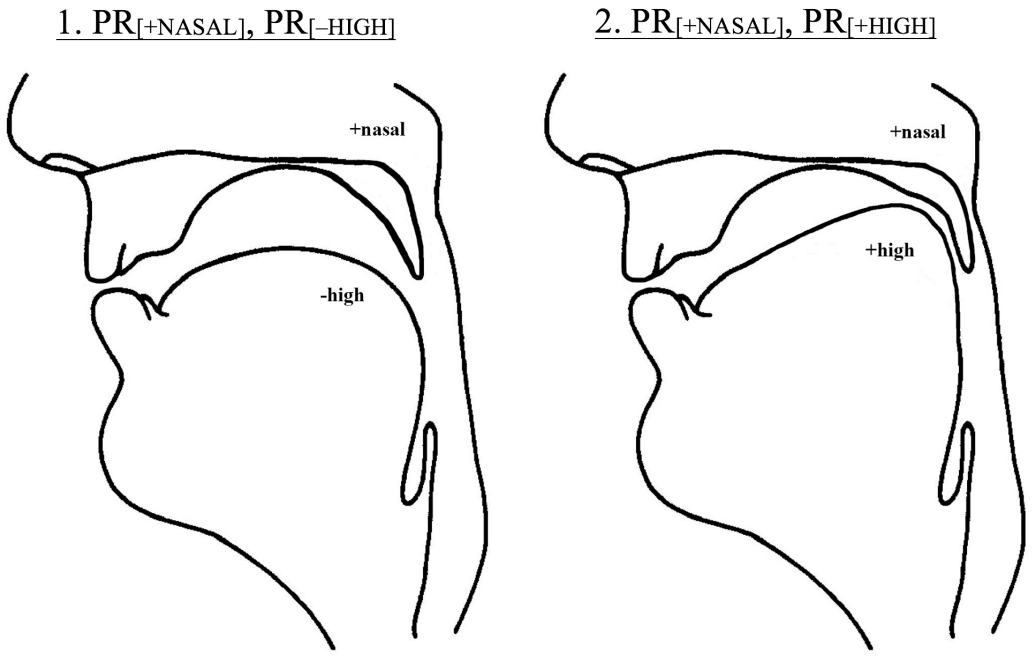

Figure 7. Intrasegmental coarticulation based on the interaction of $PR_{[NASAL]}$ and $PR_{[HIGH]}$.

In principle, $PR_{[+NASAL]}$ entails the opening of the velar port and $PR_{[+HIGH]}$ raising of the tongue dorsum. In sketch (1) $PR_{[+NASAL]}$ can be observed in a 'default', non-coarticulated state, that is, with a substantial degree of velum lowering. The tongue dorsum is not raised due to $PR_{[–HIGH]}$, leaving more space in the oral cavity for the velum port to open. In (2) $PR_{[+HIGH]}$ pushes the tongue dorsum upward, leaving less space for the velum to lower. The velar port is still opened as the realization of $PR_{[+NASAL]}$, but to a substantially lesser extent than in (1). In other words, $PR_{[NASAL]}$ is coarticulated with $PR_{[HIGH]}$ and shows variation depending on the specification (+ or –) of $PR_{[HIGH]}$. This effect can be observed by comparing how features are transduced *within* different segments; $PR_{[NASAL]}$ and $PR_{[HIGH]}$ interact differently within, say, [ã] than within [ũ]. The variation in how individual features

are transduced within the same bundle depending on their specifications is 'intrasegmental coarticulation', as illustrated in *Fig. 7*, while variation in transduction of features due to influence of features from other bundles is 'intersegmental coarticulation'.

CP uses minimal theoretical assumptions needed to account for these two well-attested speech-related phenomena, namely, for contextual variation in realization of features dependent on other features in the *same* bundle, and for contextual variation in realization of features dependent on features of a *different* bundle. In CP, *intra*segmental coarticulation results from the workings of A_1, while *inter*segmental coarticulation arises from the effects of A_2. As defined in §4, A_1 takes a feature from the phonological SR and converts it into a neuromuscular pattern. For each feature, this pattern is partially determined by specifications of other features within the same bundle, as shown in *Fig. 7*. Therefore, A_1 will assign a different neuromuscular pattern to [+NASAL] depending on how the feature [HIGH] is specified. If one imagines a certain SR (say, [dɒg]) as a feature matrix where columns stand for segments and rows for features, then A_1 takes all columns (that were loaded into CP) at once, determines the specification of each feature in each column, and generates a full set of corresponding $PR_{[F]}$s. Intrasegmental coarticulation, that is, contextual variation in transduction of features, arises when different features in the same column impose conflicting demands on A_1. Information from A_1, transmitted via a pathway connecting anterior insula to cerebellum and basal ganglia, is further manipulated by A_2. A_2 arranges $PR_{[F]}$s created by A_1 temporally, but more importantly for this discussion, A_2 extends certain $PR_{[F]}$s over boundaries of their original column. This leads to *inter*segmental coarticulation. A familiar example is labial intersegmental coarticulation in English, where A_2 takes $PR_{[+ROUND]}$, typically originating from a rounded vowel, and overextends it in the regressive (anticipatory) direction. This can be observed in the word *soon*, where $PR_{[+ROUND]}$ from the vowel is overextended to produce a labialized fricative. A_2 can also overextend $PR_{[F]}$s in the progressive (perseverative) direction. This can be observed in the word *seek*, where the $PR_{[-BACK]}$ of [i] is overextended to influence the following [k], yielding ♪si:k̟♪,[11] with a somewhat fronted velar stop. Neurobiological studies suggest (see §4) that the results of A_1 and A_2 are integrated into a finalized true phonetic representation in a region of the supplementary motor cortex at its boundary with the primary motor cortex, from which efferent commands are issued to the musculature of speech organs. However, it would seem that further experimentation is needed in order to establish whether A_1 precedes A_2 or whether there is some, or perhaps total, overlapping in their real-time neural implementation.

A great deal of variation in the execution of $PR_{[F]}$s is of course to be expected among speakers, especially given that after transduction, various other non-linguistic and non-phonetic factors influence the actual acoustic output of the human body. The output of CP is dependent on utterance-specific SRs that feed it and on the neurophysiological structures that serve as its physical implementation. Various other situational factors are introduced after transduction, which we have put aside due to their irrelevance for the general nature of CP, but it is important to keep in mind that, if not somehow separated, these factors *will*

[11] The symbol ♪ represents the actual acoustic output of the human body.

'contaminate' all experimental results (of neural imaging techniques, for example), thus leading to even greater variation in what can be observed.

The architecture of CP opens the possibility of simultaneously exploring coarticulation along two dimensions instead of just one, which leads to interesting empirical consequences, rarely explored in phonetic literature. Here we will merely state a hypothetical situation to illustrate CP's potential empirical coverage.

Let us suppose that in some language we have detected that $PR_{[+ROUND]}$ is different in [u] than in [o]. In other words, A_1 assigns a slightly different configuration to [+ROUND] depending on whether it has to take into account [+HIGH] or [–HIGH] within the same bundle. This kind of intrasegmental coarticulation can clearly be observed in *Fig. 8*.

Figure 8. Intrasegmental labial coarticulation. Notice the difference in lip rounding corresponding to [u] on the left, and to [o] on the right.

Suppose further that A_2 temporally overextends $PR_{[+ROUND]}$ across the segmental boundary in the anticipatory direction (from 'right' to 'left'). *Intra*segmental and *inter*segmental coarticulation of the same $PR_{[F]}$ is now in effect. Consider, for example, the tokens [lu:k] and [lo:k]. The [l^w] of the former token and the [l^w] of the latter token will *systematically* differ, since $PR_{[+ROUND]}$ of the former will carry with it the effect of intrasegmental coarticulation due to A_1, namely, the effect of $PR_{[+HIGH]}$, while the latter will carry the effect of $PR_{[–HIGH]}$. To reiterate, the effect of *inter*segmental coarticulation reflects the effect of *intra*segmental coarticulation. If we consider only SRs, then there can be no explanation for a systematic difference in the realization of the rounding on [l], since in both cases [l] precedes [+ROUND]. CP allows us to account for these subtle phonetic variations in an explicit and straightforward way — they follow naturally from its transduction algorithms. Thus, A_1 and A_2 are not just mechanisms that transduce features into information directly interpretable by the SM system, they are also mechanisms from which both types of coarticulation follow automatically, simply by adhering to the the minimal architecture of CP.

Our discussion has focused on the variable neuromuscular realization of a given property, such as the rounding of the vowels [u] and [o]. It is worth remembering that such a discussion of *phonetic* variability is predicated upon acceptance of the existence of a logically prior *phonological* category of vowels containing the feature [+ROUND] — it only makes sense to talk about variable realizations of *x* once we accept that *x* is a category.[12] Why do we accept the existence of such a category? Because the two segments [o] and [u] behave alike with respect to linguistic phenomena. For example, in Turkish, a process called 'vowel harmony' generates different suffix vowels depending on the preceding root vowel. As we see in (1), the vowels [u] and [o] both trigger a suffix form with [u], whereas the corresponding [–ROUND] vowels [ɨ] and [a] trigger a suffix form with [ɨ] (see Isac & Reiss 2013: §6.4 for a more comprehensive analysis).

(1) Turkish Vowel Harmony

NOM.	GEN.	gloss
uç	uçun	'edge'
son	sonun	'end'
kɨl	kɨlɨn	'body hair'
sap	sapɨn	'stalk'

As the photographs (of a Turkish speaker) in *Fig. 8* show, the lip rounding on two vowels is realized differently, but we treat the vowels as members of a category [+ROUND] because of their phonological behavior. Such considerations explain why we must recognize a distinction between phonetics and phonology. Since the two domains are different but interact with each other, there must be a transduction between them. That transduction is CP.

We fully recognize that the properties of CP outlined in this paper are too general to serve immediately as a full model of coarticulation. Not only the properties of the two component transduction algorithms, A_1 and A_2, but also the basic inventory of distinctive features must be made more explicit if CP is to be an empirically testable model. In principle, however, CP offers a theoretically coherent way to account for both intra- and inter-segmental coarticulation, and their complex interactions, while maintaining theoretical and empirical insights of generative phonology.

5.2 The (Illusory) Naturalness of Phonological Processes

The nature of the PPI as understood in CP shows the need to strictly distinguish between phonology and phonetics. This has implications for the idea of 'naturalness' in phonology. Naturalness is an elusive notion, but it usually entails explaining linguistic phenomena in

[12] This is an extension to the feature level of Hammarberg's (1976) argument for phonological segments as "logically and epistemologically prior" to their phonetic correlates.

terms of directly observable empirical facts grounded in acoustics, articulation, statistics, behavior, communication etc. Donegan & Stampe (1979: 126), proponents of Natural Phonology, suggest that the same notion of naturalness plays a role in explaining synchronic phonological patterns, diachronic phonology, as well as patterns of speech development in children:

> Natural Phonology is a modern development of the oldest explanatory theory of phonology. (...) Its basic thesis is that the living sound patterns of language, in their development in each individual as well as in their evolution over the centuries, are governed by forces implicit in human vocalization and perception.

We follow Hale (2007: §11.1) in denying any significance to apparent parallels among synchronic, diachronic and developmental 'sound patterns', therefore we will restrict our discussion to the 'naturalness' of synchronic phonology, as determined by phonetic facts. It is not difficult to find, on superficial inspection, phonological processes that seem natural in this sense. Why does [s] assimilate in voicing before adjacent [b] in a language *L*? Because it is easier for the human vocal system to maintain, and not to rapidly change the laryngeal configuration. Since voicing assimilation is indubitably a well-attested phonological process, and since this process receives an explanation from the efficient workings of "human vocalization" (*ibid.*), naturalness must obviously be a part of phonology. However, this reasoning suffers from a failure of separating 'what' from 'why'. The 'what' and the 'why' do not have the same status in linguistic theory. If the goal of linguistics, phonology included, is to explicitly model the speaker's knowledge of language, that is, to model linguistic competence, then linguistics, phonology (again) included, is to be concerned with the 'what' questions: 'What is it that a speaker knows when she or he is said to know phonology?' and 'What are the rules and representation of particular phonological grammars?' The 'why' question — 'Why is phonology (or some aspect of it) the way it is?' — does not enter into discussion at this level of inquiry (but see below). Simply put, 'what' is part of competence, but 'why' is not.

Donegan & Stampe (1979), and many other phonologists more recently, proposed to offer phonetic explanations for phonological phenomena, but despite ongoing efforts in a variety of phonological frameworks (for example, see Hayes *et al.* (2004) for attempts within Optimality Theory), this enterprise has not been convincing:

> The attempts by those who are interested in psychological phonological grammars and in finding ways to represent phonological processes (...) in phonetically natural ways have been abysmal failures (...). One possible solution to this is not to put more phonetic sophistication into psychological grammars but rather to abandon phonetic naturalness as a necessary feature of them. (Ohala 2003: 685)

Ohala's perspective (see also Ohala 1990) is not only that efforts to build naturalness into phonology have failed, but also that we would not want them to succeed, on grounds of scientific elegance. If certain recurrent phonological phenomena have a perfectly good phonetic explanation, then we do not get a better theory by duplicating the explanation

inside phonological grammar — in science, it is not better to have two explanations than one. If naturalness (e.g., the prevalence of voicing assimilation) receives a perfectly fine phonetic explanation, then it is not better to posit another, quasi-phonological explanation, especially not if the latter explanation offers no new insight.

We suggest that phonological naturalness is an illusion that arises when inspecting *phonetic* data with the purpose of understanding *phonological* processes. In other words, 'naturalness' is introduced into data in the process of externalization (and internalization in speech perception). Since we cannot have direct access to phonological representations and computations, all of our observations are of phonetic data, that is, data from actual utterances resulting from language use, which reflects many different factors. As we argued in §3 and §4, CP is the first step in externalization, so understanding CP can hopefully provide insight into what is mistakenly taken as phonological naturalness. Attaining such an insight removes the need for attributing naturalness to the phonological grammar, leading to a more parsimonious and elegant phonological theory.

Once we remove the traditional 'why' questions of Natural Phonology and its derivatives from the purview of *phonology*, we will be better prepared to answer the proper 'why' questions related to the phonological domain. At this level of inquiry we will be uncovering the biological foundations, not of speech, but of *language*, the study of which is Universal Grammar. The 'why' questions of the phonological grammar are answerable only in terms of the neurobiological substrate of the phonological faculty.

5.3. Gradience

Phonology is computation over discrete, categorical symbols. At the lowest taxonomic level, these symbols are features. However, the phonological literature is full of case studies showing the graded nature of 'phonological' units and processes (see Ernestus (2011) for an informative survey). We believe that the rejection of discreteness in phonology reflects a failure to distinguish the object of study from the data used to draw inferences about that object.

The following is a fairly standard definition of 'categoricality' vs. 'gradience', and by emphasizing certain words in it, we wish to draw the reader's attention to the conceptual level at which the definition is given:

> [C]ategorical *sounds* (...) are stable and represent clear distinct phonological categories (e.g. *sounds* showing all characteristics of voiced segments *throughout* their *realizations*) (...); gradient *sounds* (...) may change *during* their *realization* and may simultaneously represent different phonological categories (e.g. *sounds* that *start* as voiced and *end* as voiceless). (Ernestus 2011: 2115)

While we have no objection to such a characterization of categoricality vs. gradience, from the emphasized words it is obvious that the definition is immersed in the domain of the substance-laden and temporal, that is, speech (performance), not grammar (competence). The problem arises when phonetic data is used to make inferences about phonology directly

and reflexively, as if every idiosyncratic datum recorded in speech or found in a corpus is relevant for phonology, without acknowledging the distance between competence and performance. Consider another passage from Ernestus (2011: 2118):

> Ellis and Hardcastle (2002) found [by using electropalatography and electromagnetic articulography — vv & cr] that four of their eight English speakers showed categorical place assimilation of /n/ to following velars in all tokens, two speakers showed either no or categorical assimilation, and two speakers showed gradient assimilation. Together, the data show that place assimilation processes (...) may be gradient in nature. These processes cannot simply be accounted for by the categorical spreading of a phonological feature from one segment to another.

What is to be inferred from these findings that is relevant for phonology? In our view, very little (see below). The cited results, showing inter- and intra-speaker variation, as well as both discrete and gradient effects, may constitute a salient illustration of the ubiquitous lack of uniformity in the behavior of members of a speech community, but it is not in the purview of phonology to provide an explanation of such phenomena. The fact that such variation "cannot simply be accounted for by the categorical spreading of a phonological feature from one segment to another" (*ibid.*), a claim most certainly true, does not automatically mean there is something wrong with phonology conceived as categorical symbol manipulation. It is important to clearly distinguish between the object of study of phonology and the sources of evidence for that study. The object of phonological study is the human knowledge of externalizable aspects of I-language and the cognitive capacity required to construct that knowledge on exposure to limited experience. One of the sources of evidence, perhaps the primary one, bearing upon that object of inquiry are spoken utterances. Therefore, to a certain degree, it can be said that both phonology and phonetics draw from the same pool of evidence, namely, the analysis of speech. The point is merely that not all data from that pool is relevant for phonology, and a phonologist *qua* cognitive scientist needs to peel off the various complications that were introduced in the process of externalization from the underlying system of linguistic knowledge she or he is studying.

As understood here, gradience is introduced by CP's A_2, which is responsible for the temporal coordination of muscular activity specified by A_1; that is, gradience is not a phonological phenomenon. Notice the references to time highlighted in the above quote from Ernestus (2011: 2118), for example, "during" and "start as... end as". Gradience involves change over time. If we think of human phonology as involving a representational system (features and the like) that encodes the phonological portion of morphemes stored in the lexicon, and a computational system that can be thought of as a complex function of, say, composed rules (see Bale & Reiss 2018), then there is no temporal aspect to phonology. In this way phonology mirrors other competence modules, for the same reasons discussed at length by Chomsky (1980; 1986; 1988; 2000a), Anderson & Lightfoot (2002) and others. A fundamental property of the human language faculty is that on all analytical levels it fractionates language-related aspects of an analog signal into discrete elements to

which formal operations apply.[13] Even vastly different, mostly incompatible linguistic theories have acknowledged the discreteness as a defining property of language: it can be found in Martinet's (1949: 30) notion of 'Double Articulation', Hockett's (1959: 32) 'Duality of Patterning', Chomsky's (2016: 4) 'Basic Property'. Adopting such a position not only preserves a clear distinction between competence and performance, a necessity on many different grounds, but it also facilitates disentangling phonological conclusions from phonetic conclusions even though both are drawn from the same data. The only kind of conclusion a phonologist can draw from the Ellis & Hardcastle experiment cited by Ernestus is that the I-language of (some) English speakers contains a following rule: [+NASAL, CORONAL] → [+NASAL, DORSAL] / __ [DORSAL]. Phonologists can draw only this kind of conclusions because their theory both provides and determines the limits of their descriptive vocabulary. Phonological theory does not provide us with the vocabulary to describe a nasal consonant as 'kind of dorsal'. We pointed out above (§5.1) that [o] and [u] behave phonologically the same, and that both must be analyzed as [+ROUND] vowels, despite the involvement of different muscles in realizing this feature, due to intrasegmental coarticulation with [–HIGH] and [+HIGH], respectively. Again, phonologists do not have, and do not want, the vocabulary to describe a segment as 'kind of round'. [14]

If a featural assimilation rule correctly models a part of the implicit phonological knowledge of a speaker, a phonetician can then posit hypotheses as to why such a pattern exists, why there is variability in externalization of this knowledge, what are the limits of its variation, whether the variation is purely biomechanical or partly/mostly/solely cognitive, and so on. For example, the first of these questions might be explained by arguing that the demands of the $PR_{[DORSAL]}$ override the demands of the $PR_{[CORONAL]}$ because of the robustness and mechanical inertness of the relatively massive dorsal part of the tongue compared to less constrained, more mobile coronal part.[15] Therefore, the velar exerts its coarticulatory influence over the nasal. Taken this way, the relationship between assimilation and coarticulation is parallel to that of phonology and phonetics in general, that is, the former is a discretely and abstractly constructed mental *representation of* or an implicit *knowledge of* the latter.

In brief, the data most often used in inferring about phonology comes from spoken utterances. But spoken utterances are not the object of phonological study. Therefore, it does not follow that gradience of phonetic objects automatically translates to gradience of phonological objects.

---

13 "Our mind structures the linguistic input in a digital form (as opposed to an analog form), and we call this property of language *discreteness*." (Boeckx 2009: 57)

14 The idea that one's theoretical apparatus determines the range of possible observations that can be made is an old idea in the philosophy of science, discussed in particular reference to the domains of phonetics and phonology by Hammarberg (1976) and Bale & Reiss (2018).

15 This is the main idea behind the 'degree of articulatory constraint' (DAC) model of lingual coarticulation (Recasens et al. 1997), which states that the degree of coarticulatory influence and resistance of a phonetic unit rises in proportion to the degree of tongue dorsum involvement in the production of that unit.

5.4. Speech Planning and the Case of the Intervocalic /j/ in Croatian

Anticipatory coarticulation is widely adduced as proof that coarticulation is not merely a reflection of biomechanical properties (e.g., inertness) of speech organs (Farnetani & Recasens 2013). In order for a coarticulatory effect of, say, labialization ([^{w}]) to influence a unit *preceding* a rounded vowel from which the effect derives, it is necessary that some cognitive planning is involved. As we see it, phonology provides the knowledge about the discretely constructed form about to be loaded into the speech production mechanism, and CP the means to plan the coarticulatory effect. An example may be drawn from findings presented by Volenec (2013).

The purpose of that study was to see whether there is a statistically significant difference between the acoustic properties of a Croatian intervocalic palatal glide [j] present in the underlying representation, as in /pijem/ → [pijem] 'I am drinking', and a (supposedly) epenthesized palatal glide that is not present underlyingly, as in /vidio/ → [vidijo] '(I) saw'. In the latter case, the glide is supposed to surface only when adjacent to a front vowel, therefore only intervocalic environments consisting of at least one front vowel were compared. The first result was that in both cases none of the typical acoustic correlates of palatal glides (lowering of F1 and heightening of F2 compared to adjacent vowels, lowering of the intensity between F1 and F2) were found in the intervocalic position. This would suggest that the correct derivations are actually /pijem/ → [piem] and /vidio/ → [vidio], that is, with deletion, not epenthesis intervocalically. However, the second result was rather surprising. In words with underlying /j/, vowels preceding the palatal glide had their F1 significantly lowered, suggesting that the glide exerted anticipatory coarticulatory influence on the vowel, despite not being otherwise present in the acoustic signal. In words with no underlying /j/, this lowering of F1 of the preceding vowel was not present.

We argue that this case shows a dissociation between three levels of analysis: phonological, cognitive phonetic, and articulatory phonetic. Since there is no incontrovertible evidence of discrete phonological alternations in any of these cases, the most plausible derivations are /pijem/ → [pijem] and /vidio/ → [vidio], despite the fact that the spectrogram corresponding to [pijem] contains no span that independently corresponds to a segment [j]. Note that segments are abbreviations for feature bundles. The A_1 of CP receives features and transduces them into $PR_{[F]}$s. Identical adjacent $PR_{[F]}$s are thus fused to make a continuum; note that the palatal glide and front vowels share many distinctive features, and therefore many $PR_{[F]}$s. CP's A_2 temporally overextends the only $PR_{[F]}$ discriminating between the glide and front vowels — the neuromuscular command responsible for the narrowing of the palatal constriction, which results in the lowering of F1 — to serve as an acoustic cue for the glide. The articulatory system then produces something like ♪piem♪, but with ♪i♪'s F1 lowered (as compared to a 'normal' /i/ that is not in the context of an underlying /j/). The hearer usually picks up this cue, which explains why native Croatian speakers consistently report vaguely hearing some sort of [j] in these cases.

Two conclusions can be drawn from this. First, what enters the articulatory system is *not* the output of phonology ([pijem]); if it were, we would expect to find at least some independent glide-like acoustic properties between the vowels, but there are none. Therefore, a cognitive phonetic stage, distinct from both phonology and articulatory phonetics, is needed for transduction and planning. Second, the phonetic transformations that CP introduces are of a finer level of granularity than segments. The phenomenon presented here makes sense only if the input to CP consists of features, and not indivisible segments; and if the output of CP does not consist of segment-bound articulatory gestures, but $PR_{[F]}$s. This suggests that neither articulatory gestures nor segments, but transduced features ($PR_{[F]}$s) are the basic units of speech production. The apparent necessity of units at this intervening level justifies our CP model.

6. Conclusion

In this paper, we have argued that the interface between phonology and phonetics (PPI) consists of a transduction process that converts elementary units of phonological computation, features, into temporally specified neuromuscular patterns, which are directly interpretable by the motor system of speech production. Our inquiry is inspired by Lenneberg's magisterial book *Biological Foundations of Language* (1967), in which he discussed the transformation of phones into neuromuscular schemata. Our view of the PPI is constrained by substance-free generative phonological assumptions, on the one hand (§3.1), and by insights gained from psycholinguistic and phonetic models of speech production (§3.2), on the other. To distinguish transduction of abstract phonological units into planned neuromuscular patterns, arguably the very first step in speech production, from the biomechanics of speech production usually associated with physiological (or more narrowly, articulatory) phonetics, we have termed our theory 'Cognitive Phonetics' (CP). The inner workings of CP (§4) are described in terms of Marr's (1982/2010) tri-level approach, which we used to construct a 'bridge' from a formal phonological model to activity one might plausibly find in a human nervous system. In order to connect the substance-free and timeless (phonology) with the substance-laden and temporally coordinated (the SM system used in speech), CP takes features of phonological SRs and infuses them with neuromuscular activity (A_1) and arranges that activity temporally (A_2), thus generating an array of information (in a format which we call 'True Phonetic Representation') directly interpretable by the SM system. We have also presented some potential neurobiological correlates of various parts of CP. Finally, we have explored some of the implications of CP (§5), showing how such an approach might inform the study of certain phonetic phenomena, most notably coarticulation, and suggesting that some phenomena often considered to be phonological receive better explanations within CP.

Further development of CP as an explanatory model of coarticulation and other PPI phenomena will require sharpening the details of both components of the transduction algorithm (A_1 and A_2) and of CP's output units ($PF_{[F]}$). We posit CP as a model intervening between phonology (grammar) and physiological phonetics, and it is not surprising that such ideas have implications for the nature of the adjacent systems. On the phonological

side, CP calls for a reassessment of distinctive features theory in a strict biolinguistic manner. Also, the transduction of other aspects of phonological structure (e.g., prosody) should be explored. Ideally, these further developments of CP should be driven by theoretically sound models of phonological representation and computation on the one hand, and should be grounded in neurobiological findings on the other, thus reducing the conceptual distance between formal linguistics and cognitive neuroscience.

References

Ackermann, Hermann, Klaus Mathiak & Axel Riecker. 2007. The contribution of the cerebellum to speech production and speech perception: clinical and functional imaging data. *The Cerebellum* 6/3, 202–213.

Alario, Xavier, Hanna Chainay, Stéphane Lehericy & Laurent Cohen. 2006. The role of the supplementary motor area (SMA) in word production. *Brain research* 1076/1, 129–143.

Anderson, Stephen R. & David Lightfoot. 2002. *The Language Organ. Linguistics as Cognitive Physiology*. Cambridge: Cambridge University Press.

Bale, Alan & Charles Reiss. 2018. *Phonology: A formal introduction*. Cambridge: MIT Press.

Beddor, Patrice S. 1983. *Phonological and Phonetic Effects of Nasalization on Vowel Height*. Bloomington: Indiana University Press.

Benguerel, André-Pierre & Helen Cowan. 1974. Coarticulation of upper lip protrusion in French. *Phonetica* 30, 41–55.

Binder, Jeffrey R., Julie A. Frost, Thomas A. Hammeke, Patrick S. F. Bellgowan, Jane A. Springer, Jackie N. Kaufman & Edward T. Possing. 2000. Human temporal lobe activation by speech and nonspeech sounds. *Cerebral cortex* 10/5, 512–528.

Blumstein, Sheila E. & Shari R. Baum. 2016. Neurobiology of Speech Production: Perspective from Neuropsychology and Neurolinguistics. In Hickok, Gregory & Steven L. Small (eds.). 2016. *Neurobiology of Language*. 689–699. London: Elsevier.

Boas, Franz & Deloria, Ella. 1941. Dakota Grammar. *Memoirs of the National Academy of Sciences*, 23. Washington: U.S. Government Printing Office.

Boeckx, Cedric. 2009. *Language in Cognition. Uncovering Mental Structures and the Rules Behind Them.* Malden, MA: Wiley-Blackwell.

Boeckx, Cedric, Anna Martinez-Alvarez & Evelina Leivada. 2015. The functional neuroanatomy of serial order in language. *Journal of Neurolinguistics* 32, 1–15.

Bohland, Jason W. & Frank H. Guenther. 2006. An fMRI investigation of syllable sequence production. Neuroimage 32/2: 821–841.

Bouchard, Kristofer E., Nima Mesgarani, Keith Johnson & Edward F. Chang. 2013. Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495(7441), 327–332.

Brentari, Diane. Sign language phonology. In Goldsmith, John A., Jason Riggle & Alan C. L. Yu (eds.), *The Handbook of Phonological Theory*. 691–721. Oxford: Wiley-Blackwell.

Browman, Catherine P. & Louis M. Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3, 219–252.

Browman, Catherine P. & Louis M. Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6, 201– 251.

Browman, Catherine P. & Louis M. Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49/3-4, 155–180.

Buchsbaum, Bradley R., Gregory Hickok & Colin Humphries. 2001. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science* 25/5, 663–678.

Buchsbaum, Bradley R., Rosanna K. Olsen, Paul Koch & Karen Faith Berman. 2005. Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* 48/4, 687–697.

Buchwald, Adam & Michele Miozzo. 2011. Finding levels of abstraction in speech production: Evidence from sound-production impairment. *Psychological Science* 22/9, 1113–1119.

Buchwald, Adam & Michele Miozzo. 2012. Phonological and motor errors in individuals with acquired sound production impairment. *Journal of Speech, Language, and Hearing Research* 55/5: 1573–1586.

Cairns, Charles & Eric Raimy. 2011. Precedence Relations in Phonology. In Oostendorp, Marc van, Colin J. Ewen, Elizabeth V. Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*. 798–823. Oxford: Blackwell Publishing Ltd.

Carr, Philip (1993) *Phonology*. London: The MacMillian Press.

Catford, John C. 1982. *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.

Chait, Maria, Steven Greenberg, Takayuki Arai, Jonathan Z. Simon & David Poeppel. 2015. Multi-time resolution analysis of speech: evidence from psychophysics. *Frontiers in neuroscience* 9, 1–10.

Chomsky, Noam. 1980. *Rules and Representations*. Oxford: Basil Blackwell.

Chomsky, Noam. 1984. *Modular Approaches to the Study of the Mind*. Sand Diego, CA: Sand Diego State University Press.

Chomsky, Noam. 1986. *Knowledge of Language. Its Nature, Origins, and Use*. New York: Praeger.

Chomsky, Noam. 1988. *Language and Problems of Knowledge: The Managua Lectures.* Cambridge: MIT Press.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge: MIT Press.

Chomsky, Noam. 2000a. *New Horizons in the Study of Mind and Language*. Cambridge: Cambridge University Press.

Chomsky, Noam. 2000b. *The Architecture of Language*. New Delhi: Oxford University Press & Replika Press.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic inquiry* 36/1, 1–22.

Chomsky, Noam. 2007. Approaching UG from below. In Sauerland, Uli & Hans-Martin Gärtner (eds.), *Interfaces + recursion = language?*: Chomsky's minimalism and the view from syntax-semantics. 1–24. Berlin: Mouton de Gruyter.

Chomsky, Noam. 2012. *The Science of Language*. Cambridge: Cambridge University Press.

Chomsky, Noam. 2013. Problems of Projection. *Lingua* 130, 33–49.

Chomsky, Noam. 2016. *What Kind of Creatures Are We*? New York: Columbia University Press.

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Clements, George N. 1985. The geometry of phonological features. *Phonology* 2/1, 225–252.

Clements, George N. & Elizabeth V. Hume. 1995. The internal organization of speech sounds. In Goldsmith, John (ed.), *The Handbook of Phonological Theory*. 245–306. Oxford: Blackwell.

Cooper, William E. & Marc R. Lauritsen. 1974. Feature processing in the perception and production of speech. Nature 252(5479), 121–123.

Curtiss, Susan. 2013. Revisiting Modularity: Using Language as a Window to the Mind. In Piattelli-Palmarini, Massimo & Robert C. Berwick (eds.), *Rich Languages from Poor Inputs*. 68–90. Oxford: Oxford University Press.

Donegan, Patricia & David Stampe. 1979. The Study of Natural Phonology. In Dinnsen, Daniel A. (ed.), *Current Approaches to Phonological Theory*. 126–173. Bloomington: Indiana University Press.

Dronkers, Nina F. 1996. A new brain region for coordinating speech articulation. *Nature* 384(6605), 159–161.

Eickhoff, Simon B., Stefan Heim, Karl Zilles & Katrin Amunts. 2009. A systems perspective on the effective connectivity of overt speech production. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 367(1896) 2399–2421.

Ellis, Lucy & William J. Hardcastle. 2002. Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics* 30, 373–396.

Embick, David & David Poeppel. 2015. Towards a computational(ist) neurobiology of language: correlational, integrated and explanatory neurolinguistics. *Language, cognition and neuroscience* 30/4, 357–366.

Ernestus, Mirjam. 2011. Gradience and Categoricality in Phonological Theory. In Oostendorp, Marc van, Colin J. Ewen, Elizabeth V. Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*. 2114–2136. Oxford: Blackwell Publishing Ltd.

Everaert, Martin, Marinus Huybregts, Noam Chomsky, Robert C. Berwick, Johan J. Bolhuis. 2015. Structures, Not String: Linguistics as Part of the Cognitive Sciences. *Trends in Cognitive Sciences* 19/12, 729–742.

Fadiga, Luciano, Laila Craighero, Giovanni Buccino & Giacomo Rizzolatti. 2002. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience* 15/2, 399–402.

Fant, Gunnar. 1960. *Acoustic theory of speech production*. The Hague: Mouton.

Farnetani, Edda & Recasens, Daniel. 2013. Coarticulation and connected speech processes. In Hardcastle, William J., John Laver, Fiona E. Gibbon (eds.), *The Handbook of Phonetic Sciences*. 316–351. Oxford: Wiley-Blackwell.

Fowler, Carol. 1980. Coarticulation and theories of extrinsic timing. *Journal of Phonetics* 8, 113–133.

Fowler, Carol 1983. Realism and unrealism: A reply. *Journal of Phonetics* 11, 303–322.

Fowler, Carol & Elliot Saltzman. 1993. Coordination and coarticulation in speech production. *Language and Speech* 36, 171–195.

Gallistel, Charles R. & Adam Philip King. 2010. *Memory and the computational brain: Why cognitive science will transform neuroscience*. Malden: Wiley-Blackwell.

Gow, David W. 2012. The cortical organization of lexical knowledge: a dual lexicon model of spoken language processing. *Brain and language* 121/3, 273–288.

Guenther, Frank H. 1995a. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological review* 102/3, 594–621.

Guenther, Frank H. 1995b. A modeling framework for speech motor development and kinematic articulator control. *Proceedings of the XIIIth International Congress of Phonetic Sciences.* Vol. 2. 92–99. Stokholm: KTH and Stockholm University.

Guenther, Frank H., Michelle Hampson & Dave Johnson. 1998. A theoretical investigation of reference frames for the planning of speech movements. *Psychological review* 105/4, 611–633.

Guenther, Frank H., Satrajit S. Ghosh & Jason A. Tourville. 2006. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and language* 96/3: 280–301.

Guenther, Frank H. & Tony Vladusich. 2012. A neural theory of speech acquisition and production. *Journal of neurolinguistics* 25/5, 408–422.

Guenther, Frank H. & Gregory Hickok. 2016. Neural Models of Motor Speech Control. In Hickok, Gregory & Steven L. Small (eds.). 2016. *Neurobiology of Language*. 725–740. London: Elsevier.

Gussenhoven, Carlos & Haike Jacobs. 2011. *Understanding Phonology*. Third edition. London: Hodder Arnold & Hachette Livre.

Hale, Mark. 2007. *Historical Linguistics. Theory and Method*. Oxford: Blackwell Publishing.

Hale, Mark & Charles Reiss. 2000a. 'Substance abuse' and 'dysfunctionalism': Current trends in phonology. *Linguistic inquiry* 31/1, 157–169.

Hale, Mark & Charles Reiss. 2000b. Phonology as cognition. In Burton-Roberts, Noel, Philip Carr & Gerard Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*. 161–184. Oxford: Oxford University Press.

Hale, Mark & Madelyn Kissock. 2007. The Phonetics-Phonology Interface and the Acquisition of Perseverant Underspecification. In Ramchand, Gillian & Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*. 81–101. Oxford: Oxford University Press.

Hale, Mark & Charles Reiss. 2008. *The Phonological Enterprise*. Oxford: Oxford University Press.

Halle, Morris (1983/2002) On Distinctive Features and their Articulatory Implementation. In Halle, Morris (2002) *From Memory to Speech and Back*. 105–121. Berlin & New York: Mouton de Gruyter.

Halle, Morris & George N. Clements 1983. Problem book in phonology. Cambridge, Mass.: MIT Press.

Hammarberg, Robert. 1976. The metaphysics of coarticulation. *Journal of Phonetics* 4, 353–363.

Hardcastle, William J., Hewlett, Nigel (eds.) 1999. *Coarticulation: theory, data and techniques.* Cambridge: Cambridge University Press.

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.). 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press.

Heath, Jeffrey. 1999. A Grammar of Koyra Chiini: The Songhay of Timbuktu. Berlin: Mouton.

Hickok, Gregory. 2012. Computational neuroanatomy of speech production. *Nature Reviews. Neuroscience* 13/2, 135–145.

Hickok, Gregory & David Poeppel. 2000a. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92/1, 67–99.

Hickok, Gregory & David Poeppel. 2000b. Towards a functional neuroanatomy of speech perception. *Trends in cognitive sciences* 4/4, 131–138.

Hickok, Gregory, and David Poeppel. 2007. The cortical organization of speech processing. *Nature reviews. Neuroscience* 8/5, 393–402.

Hickok, Gregory, Kayoko Okada & John T. Serences. 2009. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *Journal of Neurophysiology* 101/5: 2725–2732.

Hickok, Gregory, John Houde & Feng Rong. 2011. Sensorimotor integration in speech processing: computational basis and neural organization. Neuron 69/3: 407–422.

Hickok, Gregory & David Poeppel. 2016. Neural Basis of Speech Perception. In Hickok, Gregory & Steven L. Small (eds.). 2016. *Neurobiology of Language*. 299–310. London: Elsevier.

Hickok, Gregory & Steven L. Small (eds.). 2016. *Neurobiology of Language*. London: Elsevier.

Hockett, Charles F. 1955. *A Manual of Phonology*. Baltimore: Waverly Press.

Hockett, Charles F. 1959. Animal 'languages' and human language. *Human Biology* 31/1, 32–39.
Houde, John F. & Michael I. Jordan. 1998. Sensorimotor adaptation in speech production. *Science* 279(5354), 1213–1216.
Idsardi, William J. & Eric Raimy. 2013. Three types of linearization and the temporal aspects of speech. In Biberauer, M. T. & I. Roberts (eds.), *Challenges to linearization*. 31–56. Berlin: Mouton de Gruyter.
Idsardi, William J. & Philip Monahan. 2016. Phonology. In Hickok, Gregory & Steven L. Small (eds.). 2016. *Neurobiology of Language*. 141–151. London: Elsevier.
Indefrey, Peter & Willem J. M. Levelt. 2004. The spatial and temporal signatures of word production components. Cognition 92/1, 101–144.
Isac, Daniela & Charles Reiss. 2013. *I-Language. An Introduction to Linguistics as Cognitive Science*. 2nd edition. Oxford: Oxford University Press.
Jackendoff, Ray. 1994. *Patterns in the Mind. Language and Human Nature*. New York: BasicBooks.
Jakobson, Roman. 1939. Observations sur le classement phonologique des consonnes. *Proceedings of the 3rd International Congress of Phonetic Sciences*: 34–41.
Jakobson, Roman. 1971. *Selected writings*. *Vol. 1: Phonological studies*. Second edition. The Hague: Mouton.
Jakobson, Roman, Gunnar Fant & Morris Halle. 1952. *Preliminaries to Speech Analysis.* Cambridge: MIT Press.
Jakobson, Roman & Morris Halle. 1956. *Fundamentals of Language*. The Hague: Mouton.
Jueptner, Markus & Michael Krukenberg. 2001. Motor system: cortex, basal ganglia, and cerebellum. Neuroimaging Clinics of North America 11/2, 203–219.
Katamba, Francis. 1989. *An Introduction to Phonology*. London & New York: Longman.
Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Oxford: Blackwell.
Kenstowicz, Michael & Charles Kisseberth. 1979. *Generative Phonology. Description and Theory.* New York: Academic Press.
Kirchner, Robert. 2001. Phonological contrast and articulatory effort. In Lombardi, Linda (ed.), *Segmental phonology in Optimality Theory: constraints and representations.* 79–117. Cambridge: Cambridge University Press.
Ladefoged, Peter & Johnson, Keith. 2010. *A Course in Phonetics*. Sixth Edition. Boston, MA: Wadsworth, Cengage Learning.
Larson, Charles R., Theresa A. Burnett, Jay J. Bauer, Swathi Kiran & Timothy C. Hain. 2001. Comparison of voice F0 responses to pitch-shift onset and offset conditions. *The Journal of the Acoustical Society of America* 110/6, 2845–2848.
Lashley, Karl S. 1951. *The problem of serial order in behavior*. Pasadena, CA: California Institute of Techonology.
Lass, Roger. 1984. *Phonology. An Introduction to Basic Concepts.* Cambridge: Cambridge University Press.
Laver, John. 1994. *Principles of Phonetics*. Cambridge: Cambridge University Press.
Lenneberg, Eric. 1967. *Biological Foundations of Language*. New York: Wiley.
Levelt, Willem J. M., Ardi Roelofs & Antje S. Meyer. 1999. A theory of lexical access in speech production. Behavioral and brain sciences 22/1, 1–38.
Liberman, Alvin M. 1957. Some results of research on speech perception. *The Journal of the Acoustical Society of America* 29/1, 117–123.
Liberman, Alvin M., Franklin S. Cooper, Donald P. Shankweiler & Michael Studdert-Kennedy. 1967. Perception of the speech code. *Psychological review* 74/6, 431–461.
Liberman, Alvin M. & Ignatius G. Mattingly. 1985. The motor theory of speech perception revised.

Cognition 21/1, 1–36.
Lisker, Leigh. 1978. Segment duration, voicing, and the syllable. In Bell, A. & J. B. Hopper (eds.) *Syllables and Segments*. 133–142. Amsterdam: North-Holland.
Lorenz, Konrad & Nikolaas Tinbergen. 1957. Taxis and instinct. In Schiller, Claire H. (ed.), *Instinctive behavior: the development of a modern concept.* New York: International Universities Press.
Lorenz, Konrad & Nikolaas Tinbergen. 1970. Taxis and instinctive behaviour pattern in egg-rolling by the Greylag goose. In *Studies in animal and human behavior*. Vol. 1. Cambridge, Mass.: Harvard University Press.
Lubker, J. F., R. McAllister & P. Carlson. 1975. Labial co-articulation in Swedish: a preliminary report. In Fant, Gunnar (ed.) *Proceedings of the Speech Communication Seminar*. 55–64. Stockholm: Almqvist and Wiksell.
Maran, La Raw. 1973. In Kenstowicz, M & Kisseberth, C. (eds.), *Issues in Phonological Theory. Proceedings of the Urbana Conference on Phonology*. 61–74. The Hague: Mouton.
Marr, David. 1982/2010. *Vision. A computational investigation into the human representation and processing of visual information.* Cambridge, Mass.: MIT Press.
Marshall, Chloë R. 2011. Sign language phonology. In Kula, Nancy C., Bert Botma & Kuniya Nasukawa (eds.), *The Continuum Companion to Phonology*. 254–277. London: Continuum.
Martinet, André. 1949. La double articulation linguistique. *Travaux du Cercle linguistique de Copenhague* 5, 30–37.
Mausfeld, Rainer. 2012. On some unwarranted tacit assumptions in cognitive neuroscience. *Frontiers in psychology* 3, 1–13.
Mesgarani, Nima, Stephen V. David, Jonathan B. Fritz & Shihab A. Shamma. 2008. Phoneme representation and classification in primary auditory cortex. *The Journal of the Acoustical Society of America* 123/2, 899–909.
Mesgarani, Nima, Connie Cheung, Keith Johnson & Edward F. Chang. 2014. Phonetic feature encoding in human superior temporal gyrus. Science 343(6174), 1006–1010.
Monahan, Philip J., Ellen F. Lau & William J. Idsardi. 2103. Computational primitives in phonology and their neural correlates. In Boeckx, Cedric & Kleanthes K. Grohmann (eds.), *The Cambridge Handbook of Biolinguistics*. 233–256. Cambridge: Cambridge University Press.
Morton, Katherine. 1987. Cognitive Phonetics–some of the evidence. In Channon, Robert & Linda Shockey (eds.), *In Honor of Ilse Lehiste: Ilse Lehiste Pühendusteos*. 191–194. Dordrecht: Foris Publications.
Obleser, Jonas, Aditi Lahiri & Carsten Eulitz. 2004. Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience* 16/1, 31–39.
Odden, David. 2013. *Introducing Phonology*. Second edition. Cambridge: Cambridge University Press.
Ogunbowale, P. O. 1970. *The Essentials of the Yoruba Language*. London: University of London Press.
Ohala, John J. 1990. The Phonetics and Phonology of Aspects of Assimilation. In J. Kingston & M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. 258–275. Cambridge: Cambridge University Press.
Ohala, John J. 2003. Phonetics and Historical Phonology. In Joseph, Brian & Richard Janda (eds.), *The Handbook of Historical Linguistics*. 667–686. Oxford: Blackwell Publishing.
Perkell, Joseph S., Frank H. Guenther, Harlan Lane, Melanie L. Matthies, Pascal Perrier, Jennell Vick, Reiner Wilhelms-Tricarico & Majid Zandipour. 2000. A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing

loss. *Journal of Phonetics* 28/3, 233–272.
Phillips, Colin, Thomas Pellathy, Alec Marantz, Elron Yellin, Kenneth Wexler, David Poeppel, Martha McGinnis & Timothy Roberts. 2000. Auditory cortex accesses phonological categories: an MEG mismatch study. *Journal of Cognitive Neuroscience* 12/6, 1038–1055.
Poeppel, David. 2012. The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language. *Cognitive neuropsychology* 29/1-2, 34–55.
Poeppel, David & David Embick. 2005. Defining the relation between linguistics and neuroscience. In Cutler, Anne (ed.), *Twenty-first century psycholinguistics: Four cornerstones.* 103–118. London: Psychology Press.
Poeppel, David & Martin Hackl. 2008. The functional architecture of speech perception. In James R. Pomerantz (ed.) *Topics in Integrative Neuroscience: From Cells to Cognition* 154-180. Cambridge: Cambridge University Press.
Poeppel, David & Philip J. Monahan. 2008. Speech perception: Cognitive foundations and cortical implementation. *Current Directions in Psychological Science* 17/2, 80–85.
Poeppel, David, William J. Idsardi & Virginie Van Wassenhove 2008. Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 363(1493), 1071–1086.
Poeppel, David & William J. Idsardi. 2011. Recognizing words from speech: the perception-action-memory loop. In Gaskell, Gareth & Pienie Zwitserlood (eds.), *Lexical Representation: A Multidisciplinary Approach.* 171–196. New York, NY: Mouton de Gruyter.
Postal, Paul. 1968. *Aspects of Phonological Theory*. New York: Harper & Row.
Purcell, David W. & Kevin G. Munhall. 2006. Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America* 119/4, 2288–2297.
Pylyshyn, Zenon W. 1984. Computation and cognition: Toward a foundation for cognitive science. Cambridge, Mass.: MIT Press.
Recasens, Daniel, Maria Dolors Pallarès & Jordi Fontdevila. 1997. A model of lingual coarticulation based on articulatory constraints. *The Journal of the Acoustical Society of America* 102/1, 544–561.
Reiss, Charles. 2007. Modularity in the 'sound' domain: Implications for the purview of universal grammar. In Ramchand, Gillian & Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*. 53–79. Oxford: Oxford University Press.
Reiss, Charles. 2017. Substance Free Phonology. In Hannahs, S. J. & A. R. K. Bosch (eds.) (forthcoming), *Handbook of Phonological Theory*. London: Routledge.
Sandler, Wendy. The phonological organization of sign languages. *Language and linguistics compass* 6/3: 162–182.
Scharinger, Mathias, Philip J. Monahan & William J. Idsardi. 2012. Asymmetries in the processing of vowel height. *Journal of Speech, Language, and Hearing Research* 55/3, 903–918.
Stetson, Raymond H. 1951. *Motor phonetics. A study of speech movements in action*. Amsterdam: Oberlin College Press.
Stevens, Kenneth. 1998. *Acoustic phonetics*. Cambridge, Mass.: MIT Press.
Tatham, Mark. 1984. Towards a cognitive phonetics. *Journal of Phonetics* 12/1, 37–47.
Tatham, Mark. 1987. Cognitive Phonetics–some of the theory. In Channon, Robert & Linda Shockey (eds.), *In Honor of Ilse Lehiste: Ilse Lehiste Pühendusteos*. 271–276. Dordrecht: Foris Publications.
Tatham, Mark. 1990. Cognitive phonetics. *Advances in speech, hearing and language processing* 1, 193–218.
Tourville, Jason A. & Frank H. Guenther. 2011. The DIVA model: A neural theory of speech

acquisition and production. *Language and cognitive processes* 26/7, 952–981.
Tremblay, Pascale, Isabelle Deschamps & Vincent L. Gracoo. 2016. Neurobiology of Speech Production: A Motor Control Perspective. In Hickok, Gregory & Steven L. Small (eds.), *Neurobiology of Language*. 741–750. London: Elsevier.
Trubetzkoy, Nikolaj Sergejevič. 1939/1969. *Principles of Phonology*. Los Angeles: University of California Press.
Vaux, Bert. 2008. Why the phonological component must be serial and rule-based. In Vaux, Bert & Andrew Nevins (eds.), *Rules, constraints, and phonological phenomena*. 20–60. Oxford: Oxford University Press.
Vaux, Bert. 2009. The role of features in a symbolic theory of phonology. In Raimy, Eric & Charles E. Cairns (eds.), Conte*mporary views on architecture and representations in phonology*. 75–97. Cambridge, Mass.: MIT Press.
Volenec, Veno. 2013. Acoustic analysis of the intervocalic *J* in Croatian speech. *Speech* 30/2, 117–151.
Volenec, Veno. 2015. Coarticulation. In Davis, Jasmine (ed.), *Phonetics: Fundamentals, Potential Applications and Role in Communicative Disorders*. 47–86. New York: Nova Science Publishers.
Waldstein, Robin S. 1990. Effects of postlingual deafness on speech production: implications for the role of auditory feedback. *The Journal of the Acoustical Society of America* 88/5: 2099–2114.
Yates, Aubrey J. 1963. Delayed auditory feedback. *Psychological bulletin* 60, 213–251.
Zsiga, Elizabeth C. 2013. *The Sounds of Language. An Introduction to Phonetics and Phonology*. Oxford: Wiley–Blackwell.

Veno Volenec
University of Zagreb
Faculty of Humanities and Social Sciences
Ulica Ivana Lučića 3
10000 Zagreb
Croatia
venovolenec@gmail.com

Charles Reiss
Concordia University
Linguistics Program
FB 801.23
1250 GUY STREET
Montreal H3G 1M8
Canada
charles.reiss@concordia.ca