The intervocalic palatal glide in Cognitive Phonetics

Veno Volenec & Charles Reiss

Concordia University

1.   Introduction

This paper demonstrates the contribution that instrumental phonetic study can make to the development of a formally coherent, feature-based model of phonological representation and computation, as well as to a neurobiologically plausible model of the transduction between mental representations and speech sounds. At the same time, it illustrates how the adoption of an explicit phonological model constrains and guides the formulation of the research questions and the interpretation of the results of phonetic studies. Our discussion is built around a puzzle in Croatian regarding the fate of the underlying palatal glide in intervocalic position. We show that analyzing this particular puzzle from both the phonological and the phonetic side has implications of a more general theoretical nature concerning the phonology-phonetics interface. On the basis of the analyses, we will advance two specific claims:

A.   Surface phonological representations consisting of features are not directly interpretable by the articulatory system. Rather, the interface between phonological competence and the articulatory system is mediated by a transduction component.

B.   The basic units of speech production are transduced phonological features, and not segments (Hickok 2014), syllables (Levelt et al. 1999), or articulatory gestures (Browman & Goldstein 1989, 1992).

We aim to demonstrate in this paper that the two claims presented above lead to interesting, non-trivial implications for the study of a notable phonetic phenomenon, coarticulation, while at the same time preserving the well-supported[1] findings of generative distinctive feature theory.

---

[1] "I would consider the discovery of distinctive features, and the continual refinement of their formulation over some decades, to be a scientific achievement on the order of the discovery and verification of the periodic table in chemistry." (Jackendoff 1994: 60)

It is traditionally claimed that in Croatian there is a phonological process that inserts a palatal glide between two vowels of which at least one is front (Škarić 1991, 2007; Marković 2013). Relevant examples of this alleged phonological process are provided in (1):

(1)   /ʋid-i-o/   →   [ʋidijo]   'he saw'
      /gled-a ix/   →   [gledajix]   'he looks at them'
      /na-it͡ɕi/   →   [najit͡ɕi]   'come along'
      /rad-i-o/   →   [radijo]   'he worked'
      /dio/   →   [dijo]   'part'
      /iako/   →   [ijako]   'even though'
      /ne-obit͡ʃ-a-n/   →   [nejobit͡ʃan]   'unusual'
      /pre-aktiʋ-a-n/ →   [prejaktiʋan]   'too active'

The triggering front vowel can be on either side of the inserted glide as long as the glide is flanked by another vowel on the opposite side. This means that there are two triggering environments for this purported epenthesis: $V\_\_V_{[-BACK]}$ and $V_{[-BACK]}\_\_V$.[2] Since these two triggering environments cannot be collapsed into a single one, two rules are required to capture the pattern displayed by the mappings in (1). Indeed, a prominent phonological description of Croatian (Marković 2013: 76) posits two rules of glide insertion, stated in (2):

(2) $\emptyset \rightarrow [j] / V_{[-BACK]}\_\_V$
      $\emptyset \rightarrow [j] / V\_\_V_{[-BACK]}$

Embedded in this phonological description is the claim that the supposedly epenthesized (i.e., non-underlying) glide receives the same phonetic interpretation as an underlying glide in the same context. Even though minimal pairs that would oppose Ø and /j/ in the relevant intervocalic environments are relatively hard to come by, Croatian does provide some such environments. Thus, it is claimed that the surface representations in (3) and (4), derived from different underlying representations, are realized without significant phonetic differences (Škarić 1991).

(3)   /gled-a ix/   →   [gledajix]   'he looks at them'

(4)   /gled-a-j ix/ →   [gledajix]   'look at them'

A more recent acoustic study, however, disconfirmed both the claim that a speech sound corresponding to a palatal glide is pronounced in cases such as (1) and (3), and the claim that surface representations in (3) and (4)—particularly their putative V(j)V sequences—are pronounced without significant differences (Volenec 2013). The study showed that in the

---

[2] There is some disagreement between sources on Croatian about whether /e/ actually triggers the insertion: Škarić (2007: §236) claims that it does, Marković (2013: 75) claims that it does not. Both sources agree that /i/ triggers the glide insertion. This issue can be put aside, however, since in light of the analysis in this paper both of these accounts will be discarded in favor of a non-phonological explanation.

realization of both (3) and (4) there are no typical acoustic correlates of an intervocalic palatal glide. These correlates include a temporal segment with lower F1 and higher F2 than adjacent vowels, and a decreased intensity between F1 and F2 (Stevens 1998; Reetz & Jongman 2009: 188). The average durations for the V(j)V sequences from the two different underlying sources did not show a statistically significant difference. A typical pronunciation of surface representations like those in (3) and (4) can be seen on a spectrogram in (5), which clearly shows the absence of acoustic correlates of a palatal glide. In contrast, (6) shows what the spectrogram would look like if a glide indeed were pronounced. It should be noted that (6) does not correspond to actual spontaneous pronunciation of Croatian speakers, but rather to an artificial, hyperarticulated pronunciation.

(5)   *Normal, casual pronunciation of* [gledajix] *by a native Croatian speaker*
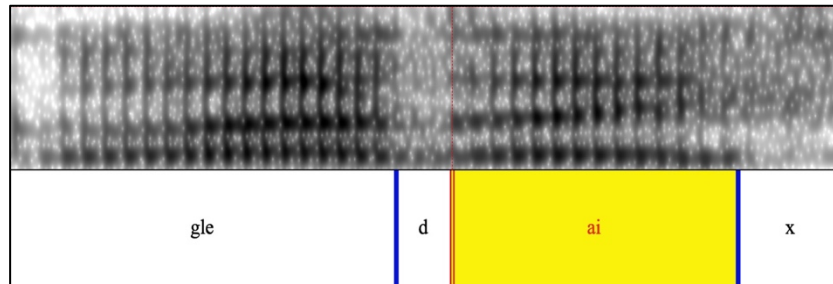
(6)   *Hyperarticulated pronunciation of* [gledajix] *by a native Croatian speaker*
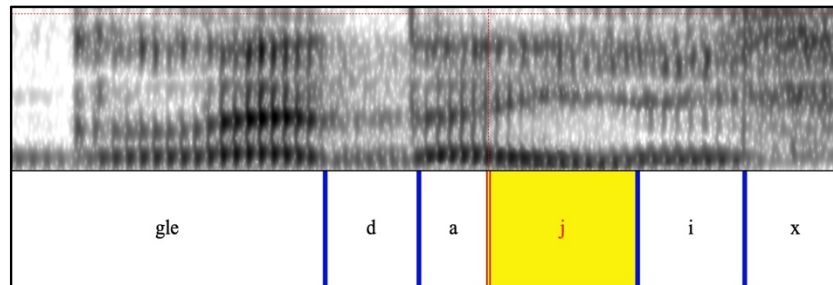
The acoustic study also showed that in words with underlying /j/, such as (4), the vowel preceding the palatal glide had its own F1 significantly lowered, suggesting that the glide exerted anticipatory coarticulatory influence on the vowel, despite the glide lacking its own typical acoustic correlates, including its own temporal window. In words with no underlying /j/, such as (3), this lowering of F1 of the preceding vowel was not present.

The unresolved question is whether any phonological alternation actually takes place—for example, insertion of the glide in forms such as (3) or deletion of the glide from forms such as (4)—or are the differences in the acoustic output of (3) and (4) due to phonetic implementation? We will approach this research question from both the phonological and the phonetic side (§3). On the phonological side, we will assume a Substance Free Generative framework (Hale & Reiss 2008; Reiss 2018). On the phonetic side, will assume the framework of Cognitive Phonetics (Volenec & Reiss 2017). The basic principles of both frameworks are briefly presented in the following section.

2.   Substance Free Generative Phonology and Cognitive Phonetics

Phonology is a component of the language faculty that involves formal computations over discrete symbols such as phonological features. Since phonology is part of linguistic competence, by definition "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 or statistical information; phonological computations treat features and other phonological units as arbitrary symbols (Hale & Reiss 2008: 169). Representational levels are related by ordered phonological rules which serve as the computational aspect of phonology (Bale & Reiss 2018).

The conceptualization of phonological features as symbols is facilitated by adopting the cognitive-neuroscience framework proposed by Gallistel & King (2010). There, symbols are "physical entities in a physically realized representational system" (p. 72), where the physical system in the case of phonological symbols, and all other linguistic symbols, is the human brain. Thus, phonological features are symbols realized in the human brain. The symbolic nature of something is that it stands for something else, something that is not the same as the symbol. That for which a symbol stands, that which it represents, is variably called its 'referent', 'correlate', or 'the represented'. Phonological features are symbols which refer to aspects of speech. At this point, it is of utmost importance not to "make the common mistake of confusing the symbol with what it represents" (p. 56). There is a connection between phonological features and speech, but this connection is highly complex and indirect (see below), and features *do not* encode speech-related information in any straightforward way. In linguistics, information related to speech is called 'phonetic substance'. It is the totality of the articulatory, acoustic and auditory properties and processes that constitute speech. For example, properties and processes of speech such as movements of the tongue, values of formants, loudness, duration expressed in milliseconds etc. fall under the rubric of substance. Since features are symbols physically realized in the brain, they cannot contain phonetic substance. In other words, features are substance-free. Failing to distinguish the two and believing that features 'are' substance or that they 'contain' substance is just an instance of the aforementioned mistake of confusing the symbol with what it represents.

The relation between any given feature and its correlate is not random or arbitrary. If it were, then any feature could in principle be realized by any possible human articulation, just as the concept/signified 'DOG' can in principle be assigned to any possible signifier. If such Saussurean arbitrariness were applicable to the realization of features, then it would be possible that [+ROUND] sometimes gets realized as a lowering of the velum, sometimes as a raising of the tongue dorsum, and so on. But that is obviously never the case. Instead, there is a non-arbitrary, lawful relation between features and their correlates. The nature of this relation is described in more detail below, but at this point it is important to emphasize that the lawful relation between features and their correlates is *phonologically* irrelevant. That means that phonological computation treats features as invariant categories, manipulating them in an arbitrary way irrespective of the variability in their realization in speech and irrespective of phonetic substance in general (Hale & Reiss 2008). The relation between features and phonological computation is thus completely independent from the relation between features and phonetic substance, as can be seen in the diagram in (7).
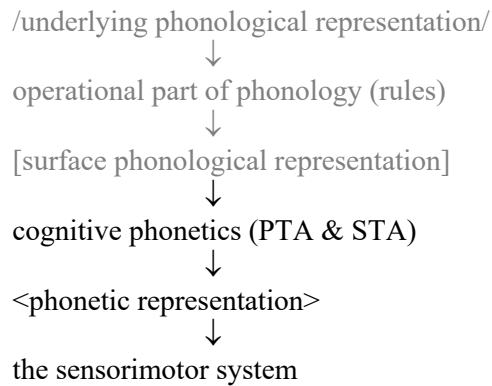
(7)   *Phonology (horizontally) and the phonology-phonetics interface (vertically)*

```
Feature – – – – (arbitrary) – – – – Phonological rule
|
|
(non-arbitrary)
|
|
Phonetic substance
```
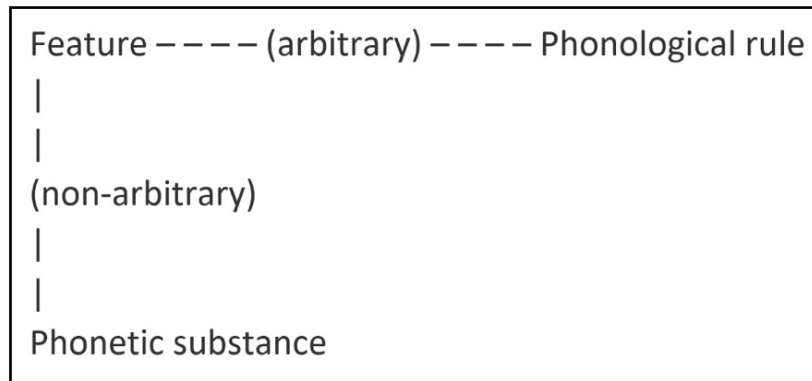
The diagram in (7) also reflects the difference between computation and transduction. Computation—corresponding to the horizontal relationship in (7)—is the formal manipulation (reordering, regrouping, deletion, addition etc.) of representational elements within a single cognitive system, and without a change in the 'representational alphabet' characteristic for that system. Transduction—corresponding to the vertical relationship in (7)—is a process of transmuting an element in one form into a distinct form, that is, a mapping between dissimilar formats.

As symbolic units, features do not contain information on the temporal coordination of muscle contractions, on the spectral configuration of the acoustic target to be reached, and so on. Yet without this information, the respiratory, phonatory and articulatory systems cannot produce speech. The sensorimotor (SM) system for speech production requires information about substance and time in order to arrange the articulatory score, therefore this information has to be integrated into a representation before being fed to the SM system. We therefore posit a transduction component that connects phonological competence with the vastly different SM system. That component is called 'Cognitive Phonetics' (CP; Volenec & Reiss 2017).

CP proposes that the phonology-phonetics interface consists of two transduction procedures that convert the substance-free output of phonology into a representational format that contains substantive information required by the SM system to externalize language through speech. The inputs to CP are the outputs of phonology, that is, surface phonological representations (SRs). SRs are strings of segments, each of which is a set of features. Each feature is transduced and subsequently receives interpretation by the SM system. This transduction is carried out by two algorithms. The 'paradigmatic transduction algorithm' (PTA) takes a feature—a symbol in the brain—and relates it to a motor program which specifies the muscles that need to be contracted in order to produce an appropriate acoustic effect. The 'syntagmatic transduction algorithm' (STA) determines the temporal organization of the neuromuscular activity specified by the PTA. In simpler terms, PTA assigns muscle activity to each feature, STA distributes that activity temporally. These transduction algorithms yield an output representation of CP, which then directly feeds the SM system. The output of CP is called the 'phonetic representation' (PR), and it can be defined as a complex array of temporally coordinated neuromuscular commands that activate muscles involved in speech production. The standard schema of phonological competence can thus be expanded to accommodate transduction performed by CP:

/underlying phonological representation/
↓
operational part of phonology (rules)
↓
[surface phonological representation]
↓
cognitive phonetics (PTA & STA)
↓
<phonetic representation>
↓
the sensorimotor system

The gray parts of the schema represent phonological competence, while the black parts correspond to the initial *phonetic* steps in speech production. That is, the difference in shading parallels the competence/performance dichotomy: phonology is competence, cognitive phonetics is performance.

To clarify the effects of PTA and STA, we can explore in some detail the transduction of a few hypothetical SRs. Interestingly, as will be shown below and in §3, PTA and STA have considerable implications: They open the possibility of elegantly accounting for subtle yet systematic interactions of two kinds of coarticulatory effects, which is only possible if we assume that the basic units of speech production are transduced phonological features. Suppose that a hypothetical language contains SRs [lok] and [luk]. Each segment is a set of features, and vowels [o] and [u] both contain the valued feature [+ROUND]. It should be noticed that [o] and [u] are different in terms of height: [o] is [–HIGH], [u] is [+HIGH]. The PTA takes a segment, scans its feature composition and determines the required muscular activity for the realization of every feature. Roughly, for [+ROUND] the PTA activates at least four muscles— *orbicularis oris*, *buccinator*, *mentalis*, and *levator labii superioris* (Seikel et al. 2009: 719– 720)—which leads to lip rounding. The difference in PTA's effect on [–HIGH] and [+HIGH] is that for the latter the algorithm creates an instruction to raise the tongue body and the jaw, while it does not for the former. While transducing [+ROUND], the PTA takes into account the specification for [HIGH] and assigns a slightly different lip rounding configuration for [o] than for [u]. Let us refer to a 'transduced feature', which we take to be the basic unit of speech production, as '$PR_{[F]}$', where 'PR' stands for 'phonetic representation' and 'F' stands for an individual valued feature. So, $PR_{[+ROUND]}$ is the transduced feature [+ROUND]. We can now say that $PR_{[+ROUND]}$ will be different for [o] because of its interaction with $PR_{[–HIGH]}$ than for [u] because of its interaction with $PR_{[+HIGH]}$. Since both of these interactions occur between transduced features from the same segment, we can refer to these effects as '**intra**segmental coarticulation'. The PTA accounts for intrasegmental coarticulation by assigning a specific neuromuscular schema depending on the specification of features from the same segment.

Let us suppose further that, while determining the durational properties of transduced features, the STA temporally extends $PR_{[+ROUND]}$ from the vowel onto the preceding consonant, that is, in the anticipatory direction. This amounts to the more familiar '**inter**segmental coarticulation', where transduced features from different segments interact. Returning to SRs [lok] and [luk], it is now apparent that: (a) $PR_{[+ROUND]}$ is different for [o] than for [u] due to its intrasegmental coarticulation with $PR_{[HIGH]}$; (b) [l]'s inherent $PR_{[–ROUND]}$ is now temporally overlapping with the $PR_{[+ROUND]}$ from the adjacent vowels because of intersegmental

coarticulation. It is important to note that the difference in PR[+ROUND] from [o] and PR[+ROUND] from [u] will be reflected on the preceding consonant: [l] in [lok] will be articulated differently with respect to lip rounding than [l] in [luk]. Thus, [l] simultaneously bears the effect of both intra- and intersegmental coarticulation. CP allows us to account for such subtle yet systematic phonetic variations in an explicit and straightforward way—they follow automatically from PTA and STA which are independently motivated by the need for transduction.

3.   The phonology and phonetics of the Croatian intervocalic palatal glide

With the preceding theoretical assumptions in mind, we can now return to the main research question of this paper, namely whether any phonological alternations take place in the Croatian forms in (3) and (4) or if the differences in the acoustic output of (3) and (4) should be attributed to phonetic implementation. In line with Cognitive Phonetics, we will argue that no phonological alternations take place in any of these cases and that the documented effect is cognitive phonetic, with concomitant articulatory and acoustic consequences.

        The traditional account, which employs the two phonological rules in (2) that lead to glide insertion in cases like (1), is implausible because, as shown by Volenec (2013), there is in fact no evidence that any phonetically realized segment was inserted between the vowels of forms in (1). In casual pronunciation of the words or phrases in (1) by native speakers, there are no acoustic correlates that would suggest the existence of a glide between the two vowels.

        Despite the fact that neither [gledajix] derived from /gled-a ix/ nor [gledajix] derived from /gled-a-j ix/, when pronounced, contain a time-span clearly corresponding to [j], it also cannot be claimed that there is a phonological process of glide *deletion*. That is, it is also not the case that the mappings in (3) and (4) should be modified so as to reflect deletion as opposed to insertion, as shown in (8) and (9).

(8)   /gled-a ix/   →   [gledaix]   'he looks at them'

(9)   /gled-a-j ix/   →   [gledaix]   'look at them'

The main reason for why (8) and (9) are implausible phonological mappings for Croatian is that even though the realizations of SRs in (8) and (9) do not contain clear glide-like acoustic correlates, the one in (9) does in fact contain more subtle evidence for the presence of [j] at the surface level. That evidence is the systematic lowering of F1 of the preceding vowel, which suggests that [j] exerts a coarticulatory effect on that vowel. If the correct SRs were indeed those in (8) and (9), then it would be impossible to explain why [a] in (9) systematically has a lower F1 than [a] in (8).

        We can therefore conclude that the are no phonological alternations taking place here and that the correct mappings are those in (10) and (11). That also means that the mappings in (1) and all other analogous cases should be reanalyzed as *not* entailing glide insertion. In still other words, the rules in (2) do not exist in Croatian phonology.

(10)   /gled-a ix/   →   [gledaix]   'he looks at them'

(11)   /gled-a-j ix/   →   [gledajix]   'look at them'

What remains to be explained is the coarticulatory effect in the realization of the SRs like in (11), that is, where there is a glide both at the underlying and the surface level. This is an effect that takes place at the phonology-phonetics interface and is, we propose, governed by the transduction algorithms proposed by Cognitive Phonetics.

In order to explain this coarticulatory effect, it is important to first explicitly state the phonological difference between [j] and [i]. Several ways of distinguishing between them have been proposed in the literature. According to Odden (2013: 48), the difference is in terms of features: [j] is [–SYLLABIC], [i] is [+SYLLABIC], all other features are the same. According to Gussenhoven & Jacobs (2011: 86), the difference is also featural, but [j] is [+APPROXIMANT], while [i] is [–APPROXIMANT]. Ladefoged & Johnson (2010: 232–234) claim that [i] and [j] are featurally identical and that the only difference is in their positioning within syllables: [j] is non-nuclear, [i] is nuclear. For our present purpose, it is not necessary to resolve this question because the following analysis can be expressed in a manner compatible with all of them. For the sake of exposition, we will arbitrarily choose Odden's (2013) approach. We can now state partial feature specifications of the segments in the relevant [aji] sequence from [gledajix] (12).

(12)

	a	j	i
SYLLABIC	+	–	+
SONORANT	+	+	+
CONSONANTAL	–	–	–
HIGH	–	+	+
LOW	+	–	–

…

The surface representation [gledajix], containing the sequence of feature matrices in (12), is interpreted by CP's transduction algorithms at the phonology-phonetics interface. The PTA transduces each feature from (12) into a corresponding $PR_{[F]}$, as displayed in (13). Note that the palatal glide and the front high vowel share all but one $PR_{[F]}$, namely $PR_{[–SYLLABIC]}$. An articulatory correlate of [–SYLLABIC] is a narrowing in the oral cavity which leads to lower sonority and thus to a lesser propensity to be included in the syllabic nucleus Odden (2013: 48). Since the glide's primary constriction is in the palatal region, that sound's $PR_{[–SYL]}$ leads to a more significant palatal constriction compared to that of the high front vowel. In other words, the PTA transduces [j]'s [+HIGH] differently than [i]'s [+HIGH] because these two segments have different specifications for [SYL]. The PTA thus leads to **intra**segmental coarticulation, as indicated by the vertical connection in (13).

(13)

a	j	i
$PR_{[+SYL]}$	$PR_{[–SYL]}$	$PR_{[+SYL]}$
$PR_{[+SON]}$	$PR_{[+SON]}$	$PR_{[+SON]}$
$PR_{[–CONS]}$	$PR_{[–CONS]}$	$PR_{[–CONS]}$
$PR_{[–HIGH]}$	$PR_{[+HIGH]}$	$PR_{[+HIGH]}$
$PR_{[+LOW]}$	$PR_{[–LOW]}$	$PR_{[–LOW]}$

PTA / STA

…

The STA then temporally extends [j]'s already coarticulated $PR_{[+HIGH]}$ in the anticipatory direction (from 'right' to 'left'), where it influences [a]'s $PR_{[–HIGH]}$. Due to this coarticulatory effect, [a] is now less low than it would have been without this effect, as is the case in the realization of the SR in (10). The effect of the STA thus gives rise to anticipatory **inter**segmental coarticulation, as indicated by the horizontal connection in (13). A well-known acoustic result of this tongue dorsum raising is the lowering of F1 (Raphael et al. 2011: 106), which is in line with the findings of the acoustic study by Volenec (2013). The fact that [j]'s coarticulatory influence on [a] is systematically different from [i]'s coarticulatory influence on [a] is explicitly captured by intrasegmental coarticulation resulting from the PTA: [j]'s $PR_{[+HIGH]}$ is coarticulated with $PR_{[–SYL]}$, while [i]'s $PR_{[+HIGH]}$ is coarticulated with $PR_{[+SYL]}$. Therefore, the $PR_{[+HIGH]}$ which the STA extends onto the preceding vowel is different when coming from [j] than when coming from [i].

4.   Conclusion

To summarize, previous literature claimed that Croatian contains a phonological alternation in which the palatal glide [j] is inserted between two vowels of which at least one is front. We have argued that this cannot be the case because there is no phonetic evidence for this insertion: had the glide been inserted, the realizations of underlying representations such as /gled-a ix/ would have to feature some glide-related phonetic correlates, but there are none. The subtle but systematic anticipatory coarticulatory effect in [gledajix] was explained from the perspective of Cognitive Phonetics (Volenec & Reiss 2017). Its paradigmatic transduction algorithm accounts for the intrasegmental coarticulation between transduced features $PR_{[–SYL]}$ and $PR_{[+HIGH]}$ from [j]. Its syntagmatic transduction algorithm accounts for the intersegmental coarticulation between the transduced feature $PR_{[+HIGH]}$ from [j] and the $PR_{[–HIGH]}$ from [a].

Two conclusions may be drawn from the analysis presented in this paper. First, what enters the articulatory system is *not* the same as the output of phonology. To take a concrete example, the output of Croatian phonology [gledajix] is not what is directly interpreted by the articulatory system. If it were, we would expect to find independent phonetic correlates between the two vowels corresponding to the realization of [j], but there are no such correlates. We also *wouldn't* expect [a] to be realized systematically differently in [gledajix] than in cases where it doesn't precede [j], since this effect cannot be captured solely by phonological features (and it should not be captured/capturable by features since the effect is not phonological). Therefore, features alone cannot account for the phonology-phonetics interface, and a cognitive phonetic stage, distinct from both phonology and articulatory phonetics, is needed for the transduction of features and for the planning of anticipatory coarticulation. In other words, the data and the analysis presented in this paper provide validation for Cognitive Phonetics by showing the necessity of a transduction system mediating between phonological competence and phonetic implementation.

Second, since phonological features are the input to CP, the PTA and the STA produce a data structure of a finer lever of granularity than segments, syllables, or articulatory gestures. Such a data structure—a transduced feature ($PR_{[F]}$)—is suitable for explicitly capturing the intricate interaction between intra- and inter-segmental coarticulation in (13), as well as numerous other comparable interactions. This suggests that it is worthwhile to entertain the

assumption that transduced features are the basic units of speech production. Speech sounds, syllables and/or articulatory gestures can then be derived from these more basic units.

References

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

Browman, Catherine & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6: 201–252.

Browman, Catherine & Louis Goldstein. 1992. Articulatory Phonology: An Overview. *Haskins Laboratory Status Report on Speech Research* 111/112: 23–42.

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

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

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

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

Hickok, Gregory. 2014. The architecture of speech production and the role of the phoneme in speech processing. *Language and Cognitive Processes* 29/1: 2–20.

Jackendoff, Ray. 1994. *Patterns in the Mind. Language and Human Nature*. New York, NY: BasicBooks.

Ladefoged, Peter & Johnson, Keith. 2010. *A Course in Phonetics*. 6th edition. Boston, MA: Wadsworth Cengage Learning.

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.

Marković, Ivan. 2013. *Hrvatska morfonologija*. Zagreb: Disput.

Odden, David. 2013. *Introducing Phonology*. Second edition. Cambridge: Cambridge University Press.

Raphael, Lawrence J., Gloria J. Borden & Katherine S. Harris. 2011. *Speech Science Primer*. 6th edition. Philadelphia: Lippincott Williams & Wilkins.

Reetz, Henning & Allard Jongman. 2009. *Phonetics. Transcription, Production, Acoustics, and Perception*. Oxford: Wiley-Blackwell.

Seikel, Anthony J., Douglas W. King & David G. Drumright. 2009. *Anatomy & Physiology for Speech, Language, and Hearing*. 4th edition. New York: Delmar, Cengage Learning.

Škarić, Ivo. 1991. Fonetika hrvatskoga književnog jezika. In Radoslav Katičić (ed.), *Povijesni pregled, glasovi i oblici hrvatskoga književnog jezika*. 61–376. Zagreb: HAZU – Nakladni zavod Globus.

Škarić, Ivo. 2007. Fonetika hrvatskoga književnoga jezika. In Katičić, Radoslav (ed.), *Glasovi i oblici hrvatskoga književnoga jezika*. 17–160. Zagreb: Nakladni zavod Globus.

Stevens, Kenneth. 1998. *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Volenec, Veno. 2013. Acoustic analysis of the intervocalic J in Croatian speech. *Speech* 30/2: 117–151.

Volenec, Veno & Charles Reiss. 2017. Cognitive Phonetics: The Transduction of Distinctive Features at the Phonology-Phonetics Interface. *Biolinguistics* 11: 251–294.