Phonological accounts of microvariation in pronunciation

Mark Hale & Charles Reiss
Concordia University, Montréal

This paper argues from an apparently paradoxical perspective. On the one hand, we develop the theory of 'substance-free' phonology proposed by Hale & Reiss (2000ab); and on the other hand, we argue for a phonological account of the language specific details of pronunciation.

We begin by comparing various definitions of phonetics and phonology. All scholars agree in calling the mappping from underlying representation to surface representation (1a↔b) 'phonology'. Keating (1988) is one who explicitly addresses the issue and we follow her in our use of terms. Other authors, at least since Hammarberg (1976), define any aspect of pronunciation that involves cognition to be part of 'phonology', and thus they include the mapping from surface representation to, say, gestural score (1b↔c), to also be part of phonology. In discussing the nature of phonetics and phonology, we must be clear that we are not using the term 'phonology' in this inclusive sense.

(1) Phonetics vs. phonology

- a↔b is phonology
- Hammarberg 1976: b↔c involves cognition, therefore it is phonology
- Keating 1988: b↔c does not map features to features, therefore it is not phonology

$\boxed{\text{a. UR in features}} \longleftrightarrow \boxed{\text{b. SR in features}} \longleftrightarrow \boxed{\text{c. Score or Percept}} \longleftrightarrow \boxed{\text{d. Physics}}$

Thus, for us, the mappings between the output of the grammar and the articulatory/perceptual systems (b↔c) are not part of phonology, by definition. Instead we propose that they are to be accounted for by a pair of *transducers*. Unlike the computations of the phonology, whose inputs and outputs are both described in the same representational alphabet, a transducer coverts between different representational formats, or even between physical/neurological states and symbolic representations. The relevant mappings are shown in (2):

(2) Relevant mappings (I and O are featural, symbolic representations):

- Grammar (Phonology): $I \leftrightarrow O$
- Transducer$_{auditory}$: $O \leftrightarrow AuditoryPercept$
- Transducer$_{articulatory}$: $O \leftrightarrow GesturalScore$

Note that this model obeys strict modularity—the Grammar can't 'see' inside the Transducers, and the Transducers can't see inside the Grammar or each other. Only the *output* of one module may be fed to another. This view allows us to take a new perspective on an old question: Are features articulatory or acoustic? The answer is that they are neither—features are just the symbolic arguments of the two transducers under consideration. Once we realize that the actual acoustics or articulation of speech can hardly be stored (there are no sound waves stored in our brains!) this conclusion is inevitable.

Generating the tremendous diversity of pronunciation in the world's languages, including, for example, microvariation in vowel realization between dialects, is a daunting task. One approach (e.g. Kingston and Diehl, 1994) has been to posit a module of language specific phonetic knowledge or rules of 'phonetic implementation'.

In this paper we propose that the null hypothesis in accounting for 'low level phonetic' variation should be that it is actually phonological, that is, encoded symbolically in representations: if language L_1 has a slightly higher [i] than L_2, then the vowels are in fact featurally distinct.

We assume a universal phonological feature set. We also assume that the transducers discussed above are universal, and thus invariant across speakers. We adopt the following points from Keating (1988):

(3) Underspecification in phonetic representations (output of derivation)

- "Each feature value is related somehow to physical targets. Targets can be context-sensitive, so it is not the case that a given feature value corresponds to one and only one numerical value along a physical scale" (280-1).
- If features correspond to targets (acoustic or articulatory) then phonetic underspecification implies that some segments have no value for some articulatory target
- Segments (even in surface representations) need not be fully specified feature bundles

(4) Context rules vs. phonetic rules

- Context rules assign a feature value to an underspecified segment. They are categorical, thus phonological.
- Transduction rules are transitional. They are best stated as gradient, thus phonetic.
- "if a segment acquires a feature value from an adjacent segment, it will share a phonetic property with that segment across most or all of its duration; if a contour is built through a segment it will have a more or less continuously changing, transitional, quality from beginning to end that will depend on context on either side" [285].
- Russian /x/: the following are not mirror images:
  - /axi/ → fully fronted fricative; context rule filled in [-back]
  - /ixa/ → transient fricative; fricative remained underspecified

The possibility of phonetic underspecification thus vastly increases the possibilities for variation in pronunciation.

Another potential increase in phonologically encoded variation comes from increasing the number of possible featural distinctions. Now, the number of segments definable with a given set of n binary features is approximately 2^n (some combinations like [+high +low] must be excluded). If the number of actual features provided by UG is even slightly higher than what we currently believe it to be, the number of possible *phonological* distinctions is vastly greater than what we currently recognize. For example, with five features for vowels we can define, in theory, 2^5 vowels, but if there turn out to be just two more features available, we get 2^7, or four times as many possibilities.

With the microvariation generated by transduction of underspecified segments and the recognition that we may (or probably) have not identified the full set of features, positing language specific phonetic implementation rules is unjustified. It is incumbent upon the proponents of such models to show that the microvariation they are attempting to model *cannot* be derived from independently needed components of the theory.