

The reality of phonological forms: a rejoinder

Robert F. Port

Department of Linguistics and Cognitive Sciences, Indiana University, Bloomington, IN 47405, USA

I thank Dr. Fowler for her thoughtful comments on my paper (see [Port, 2010](#); [Fowler, 2010](#)). I am especially pleased that she acknowledges several points of agreement between us in her first paragraph. I would like to begin by highlighting one of these points since I believe it is a radical view that she and I may share but which probably contrasts with the views of many linguists and psychologists. Certainly it contrasts with the traditional Chomskyan view.

Fowler says, and I agree, a linguistic description “reflects what works for a language community. It is not what sits inside . . . someone’s mind or brain.” Instead, it is “on average what language users know.” This means a language is a set of phonological, lexical and syntactic categories shared as conventions across some community of speakers and some range of contexts. If this is where language lives, then looking at an individual speaker, one should not expect to always find the same categories specified in the same terms. Each speaker will develop their own speech motor patterns, suitable to their vocal tract, as well as perceptual patterns and categories that are idiosyncratic in detail across speakers but which work well enough to permit successful communication. In such a system, consonants and vowels as well as other linguistic units are only definable as social patterns of behavior. And we do not necessarily know how these patterns are instantiated within any speaker. My conclusion from this is that we no longer need to look for linguistic units like Cs or Vs, distinctive features, phones, words, etc. as identifiable psychological units in speakers’ minds. They are social objects that individuals, through imitation during language acquisition, are able to produce and perceive using idiosyncratically-defined perceptual chunks and motor routines of various types and sizes.

Nevertheless, Fowler still wants to postulate consonants and vowels as real-time tokens “used by members of a community.” In other words, she apparently supports some uniform code of symbols that is the same across the brains of all speakers of a language and used for real-time production or perception of words. She gives several reasons to support the real-time role of phones or phonemes.

The first source of evidence is the research conducted on speech errors since frequently speech errors can be succinctly described as cases of “switching” segments (e.g., the error *sprit blain* for intended *split brain*). But speech error data are not reliable. First, they depend on auditory phonetic transcription (i.e., reduction to written form) thereby forcing the use of a low bit-rate representation of phonetic symbols. Plus the transcription relies on an ill-understood subjective process. But worse, as acknowledged by Fowler, recent data suggest that speech errors may seem to be discrete mainly because they were collected using a discrete phonetic transcription. As soon as one looks at the real-time behavior of someone making a speech error ([Goldstein et al., 2007](#)), then continuous valued parameters, temporal overlap of gestures and other real-time phenomena are found that ruin the simple picture of discrete segment-to-segment speech production. The second difficulty

E-mail address: port@indiana.edu

with traditional speech error data is that, as I read the literature, speech errors do not consistently support segments as the unit of analysis. They actually provide evidence for units at many different scales, from simple phonetic features (e.g., *kick* to *pick*) to whole phrases (such as the case of a *laboratory in our own computer*). I do not see how these data support segments over other size units of speech.

The second argument she presents for a psychological role for segments is the observation that if words were made from discrete particles (e.g., features or phonemes) that are not modified or blended when recombined, then it is easy to understand how new words might be coined. The “particles” could be recombined to make a great many potential new words. It is difficult to disagree with this but several points should be kept in mind.

1. Perfect unblended combination is not required for this purpose. Most phones are, in fact, partially blended since they are pronounced differently depending on their context (think of the [æ] in *bad*, *bat*, *band*, etc.). But only approximate discreteness is required for the “particulate principle” to support creation of new words.
2. This nondiscrete character of phonological particles is also true of genes which were once thought to be discrete little tokens but are now understood to be vastly more complex than previously thought and not nearly so discrete (see [Pearson, 2006](#)).
3. As suggested by Fowler, my hunch is that the particulate character of phonology is something achieved by a language as a social institution. Individual speakers do not necessarily have to represent these particles. It is sufficient for the particles to be implicit in one’s remembered inventory of speech chunks. Thus, for me to coin the new word *glimp*, I need only combine fragments from words like, eg, *glitter*, *lump*, *glimpse*, etc. I do not need to begin with /g/, /l/, /I/, etc.

So, the particulate principle itself does not provide support for simple, low-bit-rate phonemes since the principle can work even on patterns that are merely statistically differentiable in many dimensions. Thus the principle provides no support for a real-time role for discrete segments.

The third argument is that the “success” of alphabetical writing, both with the original “inventors” and with skilled users supports claims for real-time phoneme-like units.

My position is that a language, as a set of speech practices, typically tends to build an inventory of linguistic patterns (eg, words, morphemes) that reuse the same components in a number of distinct patterns. This is the phonology. These reused components can be feature-sized, segment-sized or consonant clusters, syllable codas, pitch patterns, rhythm patterns, etc. – of any size, in principle. It turns out that many of these patterns in various sizes can be distinguished from each other fairly well using a small set of letters laid out in a single row. The Greek-Phoenician alphabet engineers, around 1300 BCE, discovered that a single row of letters could specify the phonological patterns well enough to use for literacy. Letters are an engineering trick that fits many phonologies well enough that we find it worthwhile to train ourselves and our children to translate easily between letters and phonological patterns ([Port, 2006](#)). So languages have phonological structures and these structures can be differentiated from each other into various semi-discrete categories. But once we have learned to interpret letters fairly well, then the graphically discrete letters seem to be the inherent units of language. Letters are a technology that encourages us to think that phonology is only about the letter-based descriptions. Phonemes are really the hypothesized psychological counterparts of orthographic letters.

As Fowler points out, research on reading has shown very clearly that essentially all skillful reading of modern orthographies, whether in English, Russian or Chinese, works by activating the phonological and motor structures of the reader’s brain. One might have guessed that a skilled reader of Chinese would go straight from the graphic shapes of Chinese characters to meaning, but they do not. They activate pronunciations along the way (see [Frost, 1998](#)). Fowler appeared to consider this an argument in support of traditional segmental representations. But this cannot be so since the neuro-motor representations during silent reading must surely run in continuous time and would have no use for an abstract, segmental, consonant-and-vowel representation (as far as I can tell).

So I certainly would *not* say that our impression that speech has segment-sized units is due solely to alphabet training. Languages do have some segment-sized units, along with many other sized units as well. What is true, however, is that alphabet skills do tend to *highlight* the segment-like (i.e., letter-like) properties of pho-

nologies for us. Our literacy skills are thus only part of why phonemes seem so natural to us. (Formal linguistics, unfortunately, takes the letter-like properties of phonology as an *a priori* truth that we can build on.)

We need to discover the real, unbiased structures of each phonology in auditory terms. In my opinion, it will not do to say, as Fowler does, that phones and features live in the space of motor gestures. Speech is fundamentally auditory. We only need to hear language spoken to understand it. And as speech perceivers, we do not have direct access to the speaker's articulation. The auditory patterns, in frequency \times time, may emulate aspects of the articulatory space (i.e., exhibit the logic of speech articulation) yet the patterns that are important for auditory speech perception must logically remain auditory.

In several comments, Fowler seems concerned that I may reject any abstractions and employ only concrete exemplars. She asks “what sameness listeners detect” when they call two (different) things the same? But I assume that exemplar-like episodic memories include not only detailed sensory information but also many categories that events or objects in the episode might belong to by cultural convention. Grammars, like the rest of a culture, consist very largely of conventional categories: lexical ones, phonological ones, etc. Items belong to a category if, for whatever reason, community members agree that they do. Thus, syllables like /di/ and /du/ begin “the same” both (a) because the speakers think they do (due to familiarity with their written form) and (b) because the language gains benefit by beginning both syllables with similar gestures. In the same way, the glottal stop of [k^haʔn] (ie, *cotton*) is a member of the category of /t/ in my dialect of English. So aspects of exemplars may be marked as belonging to various linguistic and cultural categories. Thus two very different things, like the [d] of *Dee* and the [d] of *dew*, as well as the [ʔ] of *cotton* and the [th] of *Tom*, may be categorized the same.

Dr. Fowler and I seem to agree that a language and grammar are not structures to be found within individual speakers and should be looked for in the speech corpus. But she still offers arguments for why we should continue to trust our intuitions that speakers use phoneme-like representations of speech in real time. I have tried to show that the arguments she has raised are weak and unpersuasive. This makes one suspect that the primary reason for believing the words have a phonological spelling in memory is that they have an orthographic spelling on paper – and we academics have spent our lives sharpening our skills in the use of our orthography.

References

- Fowler, Carol, 2010. The reality of phonological forms: a reply to Port. *Language Sciences* 32, 56–59.
- Frost, Ram, 1998. Toward a strong phonological theory of visual word recognition: true issues and false trails. *Psychological Bulletin* 123, 71–99.
- Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., Byrd, D., 2007. Dynamic action units slip in speech production errors. *Cognition: International Journal of Cognitive Science* 103, 386–412.
- Pearson, Helen, 2006. What is a gene? *Nature* 441, 398–401.
- Port, Robert, 2006. Graphical basis of phones and phonemes. In: Murray, Munro, Ocke-Schwen, Bohn (Eds.), *Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production*. John Benjamins, Amsterdam, pp. 349–365.
- Port, Robert, 2010. Rich memory and distributed phonology. *Language Sciences* 32, 43–55.