2019. *Proc Ling Soc Amer* 4. 37:1-8. https://doi.org/10.3765/plsa.v4i1.4541

Subsegments and the emergence of segments

Stephanie S Shih & Sharon Inkelas*

Abstract. Q Theory proposes that the most granular and basic temporal unit of abstract phonological representation is not the segment, as widely assumed in classic generative phonology, but the quantized subsegment. With a more granular quantization of the speech stream, Q Theory provides phonological grammar with the representational capability to model behaviors that affect both the parts and the wholes of segments. In Q Theory, segments are emergent from strings of subsegments and from subsegmental interactions based on the principles of similarity, proximity, and co-occurrence that already underlie phonological operations. Evidence is presented from linguistic typology, and mechanics are drawn from speech segmentation and recognition. Q Theory makes it possible to develop an advanced theory of complex segments.

Keywords. Q Theory; segments; subsegments; complex segments; Autosegmental Phonology; Aperture Theory; Articulatory Phonology; speech segmentation; natural language processing

1. Introduction. The underlying guiding assumption throughout the history of phonological theory is that the speech stream is modeled as a temporal sequence of discrete units. These units can vary in granularity, from the utterance to the phrase, foot, syllable, and mora. The most granular unit is traditionally assumed to be the segment (phone).

However, the indivisibility of the segment as the most granular unit of phonological analysis has been challenged in several ways throughout the history of phonological theory. For one, segments have been decomposed in the vertical dimension, as (distinctive) features that co-exist at the same time (e.g., in feature theory: Chomsky & Halle 1968; in Autosegmental Theory: Leben 1973; Goldsmith 1976; Williams 1976; and in Articulatory Phonology; Browman & Goldstein 1989). This vertical decomposition is illustrated in (1).

(1)
Segment
↓
[feature]
[feature]
[feature]

Segments have also been decomposed in the horizontal, *temporal* dimension, as sets of features that are sequenced in time. This horizontal decomposition is illustrated in (2).

*Thank you to the many individuals who have discussed Q Theory ideas with us, from the participants in 2014 ABC⟷C at UC Berkeley, to the participants in the 2019 "Inside Segments" special session at the LSA meeting, where this paper was presented. We are particularly grateful to comments received over the years from Will Bennett, Louis Goldstein, Dan Jurafsky, Rachel Walker, Jie Zhang, Donca Steriade, Myriam Lapierre, Karee Garvin, and Martha Schwarz. Authors: Stephanie Shih, University of Southern California (shihs@usc.edu) & Sharon Inkelas, University of California, Berkeley (inkelas@berkeley.edu).

$$
\begin{array}{ccc}
 & \textbf{Segment} & \\
\leftarrow & \cdots & \rightarrow \\
\begin{bmatrix}[\text{feature}]\\ [\text{feature}]\\ [\text{feature}]\end{bmatrix} & \begin{bmatrix}[\text{feature}]\\ [\text{feature}]\\ [\text{feature}]\end{bmatrix} & \begin{bmatrix}[\text{feature}]\\ [\text{feature}]\\ [\text{feature}]\end{bmatrix}
\end{array} \tag{2}
$$

Temporal decomposition of segments occurs in some form or flavor in Aperture Theory (Steriade 1993, 1994), Autosegmental Theory, Articulatory Phonology, and Q Theory (e.g., Shih & Inkelas 2019). Despite positing that temporal subsegmental behavior is needed to capture phonological phenomena, many of these theories, including Aperture Theory and Autosegmentalism, still maintain segments as their smallest timing units. This discord between what is needed for phonological analysis and what is available in the phonological representation in the temporal domain has given rise to consternation about what, for example, complex segments such as prenasalized stops or affricates are—one segment? two segments in quick succession? something in between?

In contrast, Q Theory explicitly commits to the primacy of temporally-ordered subsegmental units as the fundamental building blocks of phonological representation.[1] Q Theory proposes that all canonical segments consist of a sequence of subsegments. Expanding upon the concept of segment-internal landmarks introduced in Articulatory Phonology, Q Theory proposes that there are internal phases, or subsegments (*q*), in every canonical segment (*Q*). In Q Theory, the subsegment *q*, and not the segment *Q*, is the feature bearing unit. Each subsegment is a representational unit consisting of a canonical, internally uniform feature bundle. Example (3) below illustrates a LHL-toned segment (a) and a L-toned segment (b): each segment is comprised of three subsegments. As depicted, the features of subsegments are the traditional binary (or privative) features of phonological theory:[2]

(3) a.
$$
\begin{array}{ccc}
\boldsymbol{q} & \boldsymbol{q} & \boldsymbol{q} \\
\begin{bmatrix}\text{L}\\ -\text{cons}\\ -\text{high}\\ +\text{back}\\ \cdots\end{bmatrix} & \begin{bmatrix}\text{H}\\ -\text{cons}\\ -\text{high}\\ +\text{back}\\ \cdots\end{bmatrix} & \begin{bmatrix}\text{L}\\ -\text{cons}\\ -\text{high}\\ +\text{back}\\ \cdots\end{bmatrix}
\end{array} = [\text{ǎ̀}]
$$

b.
$$
\begin{array}{ccc}
\boldsymbol{q} & \boldsymbol{q} & \boldsymbol{q} \\
\begin{bmatrix}\text{L}\\ -\text{cons}\\ -\text{high}\\ +\text{back}\\ \cdots\end{bmatrix} & \begin{bmatrix}\text{L}\\ -\text{cons}\\ -\text{high}\\ +\text{back}\\ \cdots\end{bmatrix} & \begin{bmatrix}\text{L}\\ -\text{cons}\\ -\text{high}\\ +\text{back}\\ \cdots\end{bmatrix}
\end{array} = [\text{à}]
$$

In Q Theory, the most granular level of (temporal) analysis is the subsegment, not the segment.

The remainder of the paper is organized as follows. Section 2 overviews evidence for the subsegmental nature of the speech stream in phonological analysis. Section 3 returns to the seg-

[1] For more on the differences and similarities between Q Theory and other approaches, including Articulatory Phonology, Autosegmentalism, and Aperture Theory, see Shih & Inkelas 2019:151–152.

[2] A promising avenue of future research in Q Theory is to combine the temporal phasing of subsegments in Q Theory with the scalar feature values proposed in e.g., Lionnet 2016.

ment, arguing that, under Q Theory, the canonical segment emerges from the interaction of subsegmental strings. Section 4 discusses ramifications of subsegmental granularity.

2. Evidence for subsegmental granularity. A variety of evidence supports the granularity of the subsegmental evidence, as posited by Q Theory.

2.1. CAPTURING THE RANGE OF POSSIBLE SEGMENT COMPLEXITY. The most immediate argument in favor of Q Theory is that it offers phonological purchase on the range of complex segments. Q Theory predicts that consonants and vowels which have been analyzed as single segments despite showing internal phasing can have as many—but not more than—three distinct phases. Aperture Theory had great success in depicting prenasalized stops with two phases: one nasal and one oral. Q Theory predicts the possibility of circumoralized stops, with three distinct phases, as in (4a) (see e.g. Lapierre 2017 and Garvin et al. 2018, on the Jê language Panãra). Q Theory can readily depict a prenasalized affricate, as in (4b), or a triple tone contours, as in (4c). We have discussed in previous work (e.g., Shih & Inkelas 2014; Inkelas & Shih 2016, 2017; Shih & Inkelas 2019) the challenges of representing such segments in Aperture Theory, which offers only two phases (and at that, only for consonants); we have discussed the challenges of capturing an upper bound on complexity in Autosegmental Theory, which places no inherent limits on the number of sequenced feature values that a segment can contain. Q Theory makes the strong prediction that a canonical segment can have up to three, but no more than three, featurally uniform and distinct phases:

(4) Maximally complex segments in Q Theory

		Segment (Q)		Subsegments (q q q)
a.	Circumoralized nasal stop:	bmb	↔	b m b
b.	Prenasalized affricate:	ndz	↔	n d z
c.	Triple tone contour on monophthongal vowel:	ě̀	↔	è é è
d.	Triphthong:	e̯͡ai̯	↔	e̯ a i̯

2.2. CONTRASTIVE TEMPORAL PHASING WITHIN SEGMENTS. A second argument for phonologically representing segments as strings of three subsegments comes from language-internal evidence of contrasts between segments differing only with respect to where within a complex segment the transition between phases takes place. Remijsen (2013) and Remijsen & Ayoker (2014) report that in Dinka and Shilluk, vowels with HL falling tones fall into two types: those in which the fall from High to Low tone occurs earlier in the vowel (a), and those in which it occurs later in the vowel (b). In Q Theoretic representations, this distinction is captured in terms of which pair of adjacent subsegments differs tonally:

(5) Segment-internal phasing contrasts:

a. Early transition HL contour: (H L L)
b. Late transition HL contour: (H H L)

For a similar case in affricates of contrastive temporal phrasing within a segment, see Pycha 2010, Shih & Inkelas 2019: 150, and Schwarz et al (this volume) for discussion of Hungarian affricates.

2.3. COMPUTATIONAL EVIDENCE FOR SUBSEGMENTAL GRANULARITY. A third line of arguments in favor of recognizing discrete, sequenced subsegmental units comes from natural language processing work on speech segmentation, phone recognition, and speech synthesis. In these domains, subsegmental units have regularly been found to be the atomic units of operation. Subsegments—called "triphones" in this literature—capture, roughly, information about the transition into, steady state, and transition out of a segment. These three portions of a segment are crucial to the accurate recognition and prediction of phones: monolithic segments are not optimal (see e.g., Hwang 1993; Pylkkönen 2004).

Triphones parallel the subsegmental representations posited in Q Theory, where approximately three subsegments correspond to a segment. A segment *Q*, then, can be thought of as a matrix of vectors of feature activations that represent the transitional and steady states of a segment: these vectors would be the equivalent of subsegments *q* in Q Theory. Such subdivision approaches to segments have been shown to improve speech recognition and segmentation systems (e.g., Lopes & Perdigão 2011; Maas et al. 2015). In general, windows of about 20–45ms are standard in triphone estimation in speech recognition; in comparison, the average duration of a segment is about 80ms. Using articulatory-based corpora, Zhuang et al. (2008, 2009) also found that there are roughly 3 or 4 relevant articulatory vectors per segment.

While Q Theory is intended to capture phonological generalizations, and is not a precise real-time articulatory or acoustic model of speech production or perception, it is nonetheless striking that the decomposition into transition into, steady state, and release captures both phonological generalizations and improves computational models of speech.

3. How many subsegments? Much of the literature heretofore on Q Theory has focused on providing a rational, usable phonological representation to segments that are standardly recognized as complex (see §2.1: ex. 4). With its more complex internal representations, Q Theory offers new purchase on what the segment inventory of language—or of a particular language—can consist of. This has given rise to questions that are standardly posed for Q Theory:

(6) a. How many subsegments *q* are in each segment *Q*?
Do all segments require decomposition into three subsegments? Can some segments have just one, or just two? Why aren't there more than three subsegments?
b. How different from one another can the subsegments within a segment be?
How chaotic an internal structure does Q Theory allow within a segment?

We argue that, while interesting, these questions are not ultimately that useful, as they presuppose the very assumption that Q Theory challenges, namely that the segment is the key, fundamental unit of phonological analysis.

What Q Theory offers is a different perspective, namely that the subsegment is the fundamental unit of phonological analysis. From this perspective, the interesting and productive question to pursue is this:

(7) Which subsegments *q* behave together with sufficient phonological regularity to achieve status, in a given language, as a constituent unit, identified by the grammar as a type of (segment *Q*) unit that the language deploys?

In Q Theory, segments are emergent: they are strings of subsegments that cohere to one another more than to the subsegments on either side, and which do so with sufficient regularity to be recognized as a unit in grammar. Syllables, which are emergent units consisting of segments, offer a useful analogy. Not all syllables are the same size and shape; but there are significant cross-linguistic and language-internal generalizations about the strings of segments that can be parsed into syllables.

In Q Theory, segments emerge from the regular and frequent co-action of proximal subsegments. These principles of string unithood arise from the similarity- and proximity-based principles of segmental interaction, which can be modeled by frameworks such as Agreement by Correspondence (ABC) (see e.g., Hansson 2001/2010; Zuraw 2002; Rose & Walker 2004; Bennett 2013). ABC was developed as a theory of segment-to-segment correspondence, but has been adapted, under the name ABC+Q, to the subsegmental representations of Q Theory (Shih & Inkelas 2014, 2019; Inkelas & Shih 2016, 2017).

In standard ABC terms, potentially stringent CORR(espondence) constraints make correspondence between similar and/or proximal subsegments more likely than correspondence between dissimilar and/or non-proximal subsegments. The example below in (8), from Rose & Walker (2004), depicts a set of correspondence constraints. Ranked appropriately with respect to offsetting faithfulness constraints, these correspondence constraints can produce correspondence (and resulting phonological assimilation or dissimilation effects) between identical stops only, or between stops that share place of articulation, or between stops that only share voicing, or between any pair of stops. Given this set of constraints, segments which are identical are more likely to correspond, and interact, than segments that share only one feature.

(8) Similarity-based correspondence hierarchy:

CORR-T↔T	»	CORR-T↔D	»	CORR-K↔T	»	CORR-K↔T
'identical stops'		'same place'		'same voicing'		'any oral stops'

Insofar as membership in the same *Q* is a property of similarity and proximity among subsegments, Q Theory predicts that contiguous strings of subsegments that are more similar to one another than to preceding or following subsegments are ideal candidates for parsing into a *Q* unit.

Logic from segmentation properties also lends credence to the Q-theoretic view of segment emergence. In segmentation, similar and proximal subsegments are much more likely to be parsed as a string unit. For example, consider the string of six subsegments, [à á à n d z], in (9):

(9)
$$\begin{array}{cccccc} \boldsymbol{q} & \boldsymbol{q} & \boldsymbol{q} & \boldsymbol{q} & \boldsymbol{q} & \boldsymbol{q} \\ \begin{bmatrix} \text{L} \\ -\text{cons} \\ -\text{high} \\ +\text{back} \\ \dots \end{bmatrix} & \begin{bmatrix} \text{H} \\ -\text{cons} \\ -\text{high} \\ +\text{back} \\ \dots \end{bmatrix} & \begin{bmatrix} \text{L} \\ -\text{cons} \\ -\text{high} \\ +\text{back} \\ \dots \end{bmatrix} & \begin{bmatrix} +\text{cons} \\ \text{cor} \\ -\text{cont} \\ \text{nas} \\ \dots \end{bmatrix} & \begin{bmatrix} +\text{cons} \\ \text{cor} \\ -\text{cont} \\ +\text{voi} \\ \dots \end{bmatrix} & \begin{bmatrix} +\text{cons} \\ \text{cor} \\ +\text{cont} \\ +\text{voi} \\ \dots \end{bmatrix} \end{array}$$

While no two adjacent subsegments are featurally identical, the first string of three *q*'s [à á à] and last string of three *q*'s [n d z] are more internally similar than any of the other contiguous sequences within the overall string (e.g., [á n d]). In segmentation terms, the point of lowest predictability of feature similarity is between the first three and last three *q*'s in the sequence in (9): low predictability correlates with an increased likelihood of string boundary. Standard segmentation principles (e.g., see Brent 1999 for an overview) would posit a unit boundary—corresponding to a canonical segment *Q*, roughly—between the two most internally similar strings of subsegments that occur with the most regularity in a sequence in learning. If attested with sufficient regularity, each common subsegment string could potentially come to be stored as a known segment type. Developing a formal theory of subsegmental correspondence that can derive a segmental parse of a string of subsegments is the next big challenge for Q Theory; its seeds are present in the subsegmental correspondence constraints of ABC+Q (Shih & Inkelas 2019).

4. Ramifications. Beyond the ability to represent known complex segments in a manner that is legible to phonological grammar, Q Theory provides the opportunity to think creatively about segmentation problems. This includes, for example, hitherto difficult decisions around ambisyllabicity, which involve choices between "cluster" and "complex" segments in representation, such as the boundary between (off- or on-)glides and vowels, as well as the distinction between complex consonants and consonant clusters. Consonant clusters in which certain gestures are shared clearly pattern more like single segments than completely heterorganic clusters do. Q Theory has the potential to treat every so-called consonant cluster as sequences of subsegments: *st* cluster as [s s s t t t], or *kl* cluster as [k k k l l l].

Q Theory also has the potential to represent a speech chunk that is traditionally transcribed using two different IPA symbols as a sequence of subsegments, and differences in the number of subsegments may capture differences in the tighter or looser integration of certain clusters and their patterning with what we consider "single" segments. For example, for a cluster like *st* in English, the symbiosis between the unaspirated [t] and the preceding [s] is not captured by representing the cluster as a sequence of [s] and [t]. Treating the cluster as two independent segments misses the dependency between the two parts. Q Theory offers the potential to capture the close relationship directly, as in (10). Similarly, the considerable gestural overlap between lingual consonants and /l/ in clusters like *kl* and *sl* also offers the intriguing possibility of treating them as a string of fewer than six subsegments, somewhere in between one and two canonical consonants.

(10)	**IPA transcription**	**Possible *q* representations**		
		Canonical cluster		Complex segment
a.	/st/	[s s s t t h]	*or*	[s s t]
b.	/kl/	[k k k l l l]	*or*	[k l̥ l]
c.	/sl/	[s s s l l l]	*or*	[s l̥ l]

5. Conclusion. By leveraging existing string-based principles in segmentation and segmental interaction at the subsegmental level, Q Theory opens the door to a new inventory of segment types and to the possibility of viewing segments as dynamic and emergent, rather than atomic elements in a static inventory. This new vantage point on segments offers the opportunity to reconsider the role of segments in phonological decomposition and to state generalizations at a new and more nuanced level.

References

Bennett, William G. 2013. *Dissimilation, consonant harmony, and surface correspondence*. New Brunswick, NJ: Rutgers, The State University of New Jersey dissertation.

Brent, Michael R. 1999. Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Sciences* 3.294–301.

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

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. Harper & Row.

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

Hansson, Gunnar Ólafur. 2001. *Theoretical and typological issues in consonant harmony*. University of California, Berkeley dissertation.

Hansson, Gunnar Ólafur. 2010. *Consonant harmony: Long-distance interaction in phonology*. UC Publications in Linguistics. Berkeley, CA: University of California Press. Hwang,

Mei-Yuh. 1993. *Subphonetic acoustic modeling for speaker-independent continuous speech recognition*. Pittsburgh, PA: Carnegie Mellon University dissertation.

Inkelas, Sharon, & Stephanie S. Shih. 2016. Re-representing phonology: Consequences of Q Theory. In Christopher Hammerly & Brandon Prickett (eds.), *Proceedings of NELS 46.* 2:161–174. Amherst, MA: GLSA Publications.

Inkelas, Sharon, & Stephanie S. Shih. 2017. Looking into segments. In Karen Jesney, Charlie O'Hara, Caitlin Smith, & Rachel Walker (eds.), *Proceedings of the 2016 Annual Meeting on Phonology*. Washington, DC: Linguistic Society of America. https://doi.org/10.3765/amp.v4i0.3996.

Leben, William R. 1973. *Suprasegmental phonology*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Lopes, Carla & Fernando Perdigão. 2011. Phoneme recognition on the TIMIT database. In Ivo Ipsic (ed.), *Speech technologies.* 285–302. Croatia: InTech.

Maas, Andrew L., Ziang Xie, Dan Jurafsky, & Andrew Y. Ng. 2015. Lexicon-free conversational speech recognition with neural networks. *Proceedings of NAACL*.

Pycha, Anne. 2010. A test case for the phonetics-phonology interface: gemination restrictions in Hungarian. *Phonology* 27. 119–152.

Pylkkönen, Janne. 2004. *Phone duration modeling techniques in continuous speech recognition*. Helsinki, Finland: Helsinki University of Technology Master's thesis.

Remijsen, Bert. 2013. Tonal alignment is contrastive in falling contours in Dinka. *Language* 89.297–327.

Remijsen, Bert & Otto Gwado Ayoker. 2014. Contrastive tonal alignment in falling contours in Shilluk. *Phonology* 31. 435–462.

Rose, Sharon & Rachel Walker. 2004. A Typology of Consonant Agreement by Correspondence. *Language* 80.475–531.

Shih, Stephanie S & Sharon Inkelas. 2014. A subsegmental correspondence approach to contour tone (dis)harmony patterns. *Proceedings of Phonology 2013*. University of Massachusetts, Amherst. https://doi.org/10.3765/amp.v1i1.22.

Shih, Stephanie S, & Sharon Inkelas. 2019. Autosegmental aims in surface optimizing phonology. *Linguistic Inquiry* 50. 137–196.
Steriade, Donca. 1993. Closure, release and nasal contours. In M. Huffman & R. Krakow (eds.), *Nasality*. San Diego: Academic Press.
Steriade, Donca. 1994. Complex onsets as single segments. In Jennifer Cole & Charles Kisseberth (eds.), *Perspectives in phonology.* 202–291. Stanford: CSLI.
Williams, Edwin. 1976. Underlying tone in Margi and Igbo. *Linguistic Inquiry* 7. 463–484.
Zhuang, Xiaodan, Hosung Nam, Mark Hasegawa-Johnson, Louis Goldstein & Elliot Saltzman. 2008. The entropy of the articulatory phonological code: Recognizing gestures from tract variables. *Proceedings of Interspeech 2008*.
Zhuang, Xiaodan, Hosung Nam, Mark Hasegawa-Johnson, Louis Goldstein & Elliot Saltzman. 2009. Articulatory phonological code for word classification. *Proceedings of Interspeech 2009*.
Zuraw, Kie. 2002. Aggressive reduplication. *Phonology* 19.395–439.