Frontiers and Challenges in Articulatory Phonology

Dani Byrd
Department of Linguistics, University of Southern California, United States
E-mail: dbyrd@usc.edu

ABSTRACT

Browman and Goldstein have outlined an account of the cognitive representation of words in which phonological and phonetic specification is isomorphic [see e.g. 1, 2]. Their *Articulatory Phonology* framework examines the hypothesis that linguistically significant vocal tract constrictions, or *gestures*, are the atomic units of phonological representation. And, using Saltzman's *Task Dynamics* model [3], these gestural units are quantitatively described as dynamical systems. These innovations have led to significant improvements in our understanding of how spoken language is produced and why it exhibits particular qualities and patterns. The success of this approach has ignited an interest in exploring the range of linguistic challenges to which Articulatory Phonology might rise. There are new frontiers to be explored within the approach and known questions to be considered in new ways.

1. INTRODUCTION

The phonological structure of words can be viewed as relying on a finite (and relatively small) set of underlying units that can be combined in language-particular ways. Let's briefly review the nature of the phonological units postulated in Articulatory Phonology. Gestures function simultaneously as combinatorial units of information (i.e. contrast) and as action units in speech production [1]. Consequently, there is no grammatical mediation between the phonological representation and its implementation by the speech production system. Functionally independent constrictor systems of the vocal tract (*tract variables)* form the foundations for phonological contrast. Tract variable *gestures* are abstract and task-defined, not defined in terms of individual articulator movement. The atomic gestural units, as well as the larger gestural molecules composed of them, are modeled as dynamical systems and, as such, exhibit many of the well-known properties of such systems, including stability of form (resistance to perturbation) and lawful warping of motion form with changes in performance dimensions. Finally, gestural units are coordinated, or phased, with one another in a highly structured and interwoven fashion. This patterning causes competition for articulators simultaneously called on by concurrently activated gestures, leading to the coarticulation known to pervade spoken language.

Articulatory Phonology allows contrast to be encoded in three ways: the presence or absence of a gesture, the dynamical parameterization of a gesture (e.g. its point attractor target/equilibrium position), and the relative coordination of gestures. Despite this lexical specification, phoneticians have long pointed out that there is great variability in the realization of words depending on a variety of factors. The dynamical systems approach pursued in Articulatory Phonology provides a way of handling underlying invariance at the level of control (i.e. the specification of phonological units) and variability in performance.

In this paper we will provide a brief tour of some frontiers and challenges for Articulatory Phonology. We will start with an overview of some traditional challenges in phonetics and phonology that have been successfully addressed using this approach. Then we will move to some outstanding issues that will require the elaboration or expansion of the current working Articulatory Phonology model. Finally, we will conclude with a brief discussion of challenges that obtain for any phonological theory and how these questions, while not specific to Articulatory Phonology, might be sharpened by a gestural view of lexical representation.

2. SOME (RELATIVELY) CONQUERED FRONTIERS

The majority of past work within the Articulatory Phonology framework has addressed the nature of phonological representation of words and how variability in the realized form of a word can be satisfactorily captured given these representations. The postulation of abstract, dynamically-specified, and underlyingly stable phonological units constrains the possibilities for accounting for variability in word forms. These constraints have shaped accounts of how gestural overlap and reduction can lead to cross-linguistic variability, variability due to syllable structure, and casual speech variation, such as assimilation and reduction (among many references, see [4]). Allophony related to syllable structure has also received much attention, and an understanding of syllable structure as particular patterns of gestural coordination has been pursued [1, 5]. For example, consonants in onsets and in codas are demonstrably different in their internal timing and in their coordination with respect to a tautosyllabic syllable nucleus (e.g. see [1] [6] and references therein). Distributional patterns like those of light & dark [l]s [1] and of nasalized vowels in English [1, 5] are elegantly understood in terms of gestural relative timing. Two challenges that remain in understanding the importance of relative timing in explaining allophony are: first, to formulate a dynamical (rather than rule-based) description

of relative timing [7, 8, cf. 4] (more on this below), and secondly, to understand why these particular context-specific timing patterns have evolved as they have (e.g. in onsets & codas). While progress has been slower here, important advances are found in [9, 10, 11].

3. CHALLENGES CRITICALLY RELATED TO A GESTURAL REPRESENTATION

PROSODICALLY STRUCTURED VARIABILITY

Understanding the representation of phonological structure at and above the level of the word, namely prosodic structure, and how this relates to articulatory performance is one of the primary current challenges for the Articulatory Phonology approach. In contrast to the well-established exploration of the dynamic characterization of atomic gestural units and more elaborated gestural molecules, the dynamical nature of phrasal structure (juncture & intonation), lexical stress, and focus or accent is a relatively new frontier for Articulatory Phonology.

Within this dynamical systems approach, there are specific options available for the realization of prosody. Prosodic structure could *directly* alter gestural parameter values and relative timing. The extent to which this is theoretically appealing depends on the level of commitment to invariant or stable gestural molecules as the basis for lexical representation. That is, if a word can appear with its gestures having, for example, a different target parameter value in each and every prosodic environment, this weakens the assumption of stable canonical word representation. However, if gestural parameter values are constrained to vary only within specific *windows* or ranges of variation [8, 12, 13], then a principled account of the relation of prosodic structure to chosen values within these windows may be facilitated.

A dynamical approach offers, however, other options for the instantiation of prosodic structure in addition to direct influence on gestural parameter values and relative timing. A dynamics at the gestural activation level is also possible, such that gestural activation trajectories unfold in a way sensitive to the prosodic environment. Variation at this level will in turn play out in the on-going gestural parameter values (whose instantiation are controlled by the activation functions) and in gestural relative timing.

Possible foundations for understanding how prosody might alter a gestural score from its underlyingly invariant (or at least stable) form include: (a_1) direct changes in parameter values and/or timing, (a_2) changes in parameter values and/or timing over a particular interval that defines a prosodic event (or prosodic gesture [14]), or (b) changes in the unfolding of gestural activation functions, either (b_1) tied to prosodic position/specification or (b_2) during an interval defining a prosodic event. In addition to influencing ongoing working parameter values, changes of sort (b) are likely also to have consequences for the relative timing among affected gestures [14].

The challenge of determining the best approach to characterizing prosodic variability is an active area of current research. Because of the opacity of the kinematics with respect to their source in underlying control, modeling work will prove requisite. Several studies [14, 15, 16] have argued that options (a_1, a_2), the manipulation of particular gestural parameter values, cannot successfully account for the articulatory kinematic patterns seen at phrase boundaries. Byrd and Saltzman [14] pursue option (b_2) with results that show some promise. They model prosodic events as non-tract variable π-gestures (prosodic gestures with their own intrinsic temporal properties) that, in the case of phrase boundaries, cause a slowing of the central clock that controls the unfolding of activation trajectories [14].

Other aspects of prosodic variability include the realization of word form under accent (i.e. focus) and the realization of lexically stressed versus non-stressed syllables. Kinematic studies of articulatory behavior under differing stress and accent are available (see, among many, [16, 17, 18, 19, 20]) and find durational and relative timing consequences of linguistic prominence. However, just as for phrasal prosody, the dynamic basis for these kinematic changes is not straightforward [18, 16], and control level changes that effect multiple parameters and/or gestural activation seem likely to be at work.

SOUNDS OF THE WORLD'S LANGUAGES

A challenge for Articulatory Phonology, which, while perhaps more mundane, is still formidable, is the appropriate characterization of the sounds of the world's languages. Clearly, additional tract variables will have to be defined for non-pulmonic sounds and for words involving contrasting phonation type and/or lexical & grammatical tone (see e.g. [21]). It's also unclear whether the currently hypothesized dimensions of constriction location, degree, and orientation will need to be extended to account for linguistically important differences in constriction shaping, for example in fricatives or liquids (see e.g. [22]). Finally, an approach to vowel representation elaborated enough to account for various contrasts, such as tense/lax and ATR, is necessary.

INTONATION

If the appropriate task space were understood for lexical tone—and it is an open question whether such a task space would be in the articulatory or auditory domain—there would exist the remaining question of whether that same task space would be at work in capturing intonational gestures compatible, for example, with descriptions such as that of Pierrehumbert and Beckman [23] (see also [21]).

INTERPLAY OF BIOMECHANICS AND GESTURAL CONTROL

In addition to continuing to improve our understanding of the phonological primitives that gives rise to articulation, we cannot ignore the fact that the biomechanical plant that effects this representation plays a role in creating the kinematic and acoustic behavior we observe [24]. Much progress has been made in characterizing the plant, but an interactive, 3-dimensional vocal tract model (including tongue & larynx) interfaced with both Browman and

Goldstein's linguistic gestural model and with an articulatory synthesizer necessary for model testing still remains a challenge for the future. Conversely, what has previously been viewed as an attribute of the plant —namely, ease of articulation evaluated in terms of metabolic cost (e.g. for distance traveled)—can now be addressed from an abstract dynamical perspective, in which articulatory ease can be seen as a function of dynamical efficiency at the task and systems levels (see Pouplier, this volume, & [25]).

ACTIVATION DYNAMICS

The ultimate challenge for a dynamical systems approach to speech production is the description of a sequential dynamics on the activation functions of the atomic gestural units. Such a sequential dynamics would eliminate the use of rules for stating phasing relations among gestures (c.f. [4]) by allowing the overlapping activation patterns to be created flexibly on the fly. Such an activation dynamics would capture linguistic structure such as syllabic, stress, and prosodic structure by reflecting it in the behavior of the network(s) connecting the gestural structures that are the individual words and the collection of words into larger phrases. Activation dynamics and the consequences of activation trajectory shaping have only begun to be addressed within this framework, see for example, [14, 26].

4. CHALLENGES: BROAD ISSUES IN PHONOLOGY

Articulatory Phonology provides a description of phonological representation in which time and timing are critical elements. Adopting such a representation does not obviate the need to explain phonological patterns in a principled way, but it can sharpen phonological questions. Imagine a chemist trying to explain the nature of particular chemical reactions without a correct account of the underlying atoms and molecules and their properties—a hard task indeed. Improvements in our understanding of the nature of atomic phonological units and their dynamical characteristics offered by Articulatory Phonology can clarify the way in which certain Optimality Theoretic constraints should be stated, and indeed may affect decisions as to which aspects of speech patterning should be captured in a constraint-based grammar and which can fall out from an account of the nature of lexical entries, underlying forms, and/or their articulatory realization. So while we indicate below some general questions and challenges faced by any theory of phonology, this is done with an eye to how questions might be productively reframed given dynamical phonological units based on linguistically significant vocal tract constrictions.

CONNECTIONS AMONG LEXICAL ENTRIES

Any phonology, regardless of representational primitives, will need to address the challenge of how words can be related to one another—for example, how the patterning of phonological units within and across morphological paradigms can influence the structuring of individual words. A gestural account of lexical representation will be helpful within to the extent that it allows for an insightful characterization of the structure of lexical entries so as to better state relations among them. Still, the issue of how patterns across related word forms are established and characterized is a general phonological question. In some approaches, metrics of similarity among word forms play a role. Given a gestural representation, in which time is an inherent aspect of a unit, it may be that (abstract) duration (e.g. the stiffness parameter value) could play a role in the calculation of similarity in a way parallel to similarity in constriction location and degree (R. Walker, p.c.).

PHONOTACTICS

Whatever the nature of representation (e.g., featural or gestural), any theory of phonology is going to require an understanding of the phonotactic constraints of a language. For a gestural approach, the permissible, language-specific combinations of gestures in words must be identified. This would include a description not only of which gestures and gestural molecules are permitted in which particular word and syllable positions but also an account of the acceptable temporal pattering of those combinations. Browman and Goldstein have identified how some phonotactic statements previously relying on features can be made in a more illuminating way given a gestural representation of the underlying phonological units. For example, they describe aspiration patterns in syllable onsets showing that a restatement of the phonotactic restrictions on aspiration as sanctioning only a single laryngeal opening gesture in an onset (in English) can explain the lack of aspiration in [sC] onsets in an elegant way (e.g. [1], cf. [3]).

CORRESPONDENCE AND ALIGNMENT

Phonological description also seems to require an account of the formal relations (e.g. correspondence), both local and long-distance, among phonological units within a word. This might be called on, for example, to state descriptions of harmony (e.g., corresponding elements in a single word form) or reduplication (e.g., correspondence between elements in a base and reduplicant). Exploration of these relations in terms of gestural units is in its infancy. It may be that an understanding of activation dynamics will be relevant. For example, Gafos [27] has approached consonant harmony patterns in terms compatible with an account that allows a continuous (but possibly waxing & waning) activation of a gestural point attractor.

The relative coordination of higher and lower levels of phonological structure, as well as the relative organization of morphological and phonological structure, are currently characterized by ALIGN constraints in Optimality Theoretic approaches to phonological grammars. How to understand alignment (in its formal sense) in the context of *intrinsically temporal* phonological units, such as gestures, is an outstanding challenge. Gafos [28] offers an approach to alignment that highlights the phonological importance of articulatorily significant landmarks in the production of a gesture.

Another tool available within a dynamical systems

approach that may prove relevant for capturing correspondence and/or alignment relations is system coupling. If two systems are coupled, the behavior of one can affect, at least under certain circumstances, the behavior of the other. The characterization of coupling relations among linguistic gestures has been examined in a preliminary way in [7, 10, 25] (though coupling relations among action units in other skilled motor tasks has long been a topic of active research.) In order to describe relations among phonological units, not only will intergestural coupling need to be more thoroughly examined, but coupling among larger gestural structures—gestural molecules—will have to be broached. Such an approach could potentially provide a quantitative basis for correspondence, long-distance dependencies, and alignment that is amenable to testing and modeling.

5. CONCLUSION

The *Articulatory Phonology* framework has defined and evaluated the hypothesis that the atomic cognitive units of word representation are simultaneously phonological and phonetic. Linguistic successes within this approach have fueled an interest in further exploring a range of phonological challenges from a dynamical perspective.

ACKNOWLEDGMENTS

The author gratefully acknowledges the support of NIH grant DC03172 and thanks Louis Goldstein, Peter Ladefoged, Elliot Saltzman, & Rachel Walker.

REFERENCES

[1] C. P. Browman & L. Goldstein, "Articulatory phonology: An overview," *Phonetica*, 49, 155-180, 1992.

[2] C. P. Browman & L. Goldstein, "Dynamics and articulatory phonology," in *Mind as Motion: Dynamics, Behavior, and Cognition*, R. Port, & T. van Gelder, Eds., Cambridge MA: MIT Press, 1995.

[3] E. L. Saltzman & K. G. Munhall, "A dynamical approach to gestural patterning in speech production," *Ecological Psychology*, 1, 333-382, 1989.

[4] C. P. Browman & L. Goldstein, "Tiers in articulatory phonology, with some implications for casual speech," in *Papers in Laboratory Phonology: I. Between the Grammar and the Physics of Speech,* J. Kingston & M. E. Beckman, Eds., pp. 341-338. Cambridge UK: Cambridge Univ. Press, 1990.

[5] R. Krakow, "Physiological organization of syllables: a review," *J. of Phonetics*, 27, 23-54, 1999.

[6] D. Byrd, "Influences on articulatory timing in consonant sequences," *J. of Phonetics*, 24(2), 209-244, 1996.

[7] E. Saltzman, & D. Byrd, "Task-dynamics of gestural timing: Phase windows and multifrequency rhythms," *Human Movement Science*, 19(4), 499-526, 2000

[8] D. Byrd, "A phase window framework for articulatory timing," *Phonology,* 13(2), 139-169, 1996.

[9] I. Mattingly, "Phonetic representation and speech synthesis by rule. In *The Cognitive Representation of Speech, Advances in Psychology 7*, T. Myers, J. Laver, & J. Anderson, Eds., North-Holland Publishing Company, 415-420, 1981.

[10] C. P. Browman & L. Goldstein, "Competing constraints on intergestural coordination and self-organization of phonological structures," *Bulletin de la Communication Parlée*, 5, 25-34, 2000.

[11] D. Silverman, *Phasing and recoverability.* New York: Garland, 1997.

[12] P. A. Keating, "The window model of coarticulation: Articulatory evidence." in *Papers in Laboratory Phonology I Between the Grammar and Physics of Speech*, J. Kingston & M. E. Beckman, Eds., Cambridge: Cambridge University Press, 451-470, 1990.

[13] G. Docherty, *The Timing of Voicing in British English Obstruents.* Berlin: Foris, 1992.

[14] D. Byrd & E. Saltzman, "The elastic phrase: Dynamics of boundary-adjacent lengthening," *J. of Phonetics*, in press.

[15] J. Edwards, M. E. Beckman, & J. Fletcher, "The articulatory kinematics of final lengthening," *JASA*, 89(1), 369-382, 1991.

[16] T. Cho, "Manifestation of prosodic structure in articulatory variation: Evidence from lip movement kinematics in English." *Laboratory Phonology VIII*, subm.

[17] C. Mooshammer & S. Fuchs, "Stress distinction in German: simulating kinematic parameters of tongue-tip gestures," *J. of Phonetics*, 28, 2002.

[18] K. de Jong, "The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation. *JASA*, 97, 491-504, 1995.

[19] M. Beckman & K. Cohen, "Modeling the articulatory dynamics of two levels of stress contrasts," in *Prosody: Theory and Experiment—Studies Presented to Gösta Bruce*, M. Horne, Ed., pp. 169-200. Dordrecht: Kluwer Academic Publishers, 2000.

[20] J. Harrington, J. Fletcher & M. Beckman, "Manner and place conflicts in the articulation of accent in Australian English," in M. B. Broe & J. Pierrehumbert, Eds. *Papers in Laboratory Phonology V*, pp. 40-51 Cambridge UK: Cambridge University Press, 2000.

[21] R. S. McGowan & E. L. Saltzman, "Incorporating aerodynamic and laryngeal components into task dynamics," *J. of Phonetics*, 23 255-269, 1995.

[22] S. Narayanan, D. Byrd, & A. Kaun, "Geometry, kinematics, and acoustics of Tamil liquid consonants," *JASA*, 106(4), 1993-2007, 1999.

[23] J. B. Pierrehumbert, M. E. Beckman, *Japanese Tone Structure,* Cambridge, MA: MIT Press [Linguistic Inquiry Monograph Series, 15], 1998.

[24] Special Issue *Bulletin de la Communication Parée*, 4, 1998.

[25] L. Goldstein, M. Pouplier, L. Chen, E. Saltzman, & D. Byrd, "Gestural action units slip in speech production errors." (submitted).

[26] E. Saltzman, A. Löfqvist & S. Mitra, S. ""Clocks" and "glue"—Global timing and intergestural cohesion," In *Papers in Laboratory Phonology V,* M. B. Broe & J. B. Pierrehumbert, Eds., pp. 88-101. Cambridge UK: Cambridge University Press, 2000.

[27] D. Gafos, "A grammar of gestural coordination," *NLLT* 20, 269-337, 2002

[28] D. Gafos, "Eliminating long-distance consonantal spreading," *NLLT*, 16, 223-278.