Articulatory Phonology and Computational Adequacy

Mark Tatham

Reproduced from *Proceedings of the Institute of Acoustics 1996,* pp. 375-382. St Albans: Institute of Acoustics

This paper discusses **articulatory phonology** and **task dynamics** as potentially computationally adequate models which, together, might characterise speech production. The idea is introduced that, particularly at the task dynamic level, the object oriented computational paradigm is appropriate — this is a novel approach in speech production modelling. The paper concludes that **articulatory phonology** and **task dynamics** are a step toward computational adequacy, but that that goal is not quite reached.

THE BASIC THEORY

Articulatory phonology was proposed by Browman and Goldstein (1986) a decade or so ago as an attempt to move towards the unification of phonetic and phonological descriptions of speech production. They identified theoretical discrepancies between the then two distinct models, and differences of approach by theorists in the two areas. They proposed unifying the two by treating them as *low and high dimensional descriptions of a single system* (Browman and Goldstein, 1993).

In Browman's and Goldstein's view the high dimensional description is concerned with utterance planning and the low dimensional description with utterance execution — that is, execution of the plan. Unification, they proposed, can be achieved by incorporating into a single model the idea that the physical system (identified with *phonetics*) constrains the underlying abstract system (identified with *phonology*), making the units of control at the abstract planning level the same as those at the physical level.

For Browman and Goldstein planning and execution are seen as more closely related than in other theories of speech production.

The *plan* of an utterance is formatted as a gestural score (see Fig. l), which provides the input to a physically based model of speech production — the **task dynamic model** (Saltzman 1986). The gestural score graphs locations within the vocal tract where constriction can occur, indicating the planned or target degree of constriction.

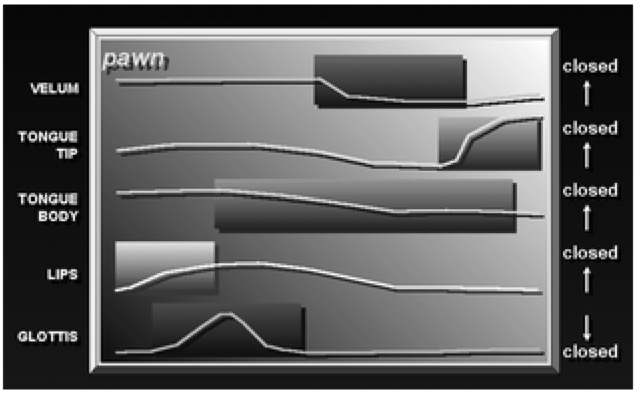

Fig. 1 An example of a gestural score. Time runs from left to right; the tracks define various vocal tract variables and their degree of constriction. Blocks indicate planned events, continuous lines are computed executions of these events.

The sequencing of gestures and their durations, and the timing relationships between the various vocal tract variables involved are critical to the score and how it unfolds. The tract variables form a *parametric framework* which is manipulated later in the **task dynamic model**. Lip aperture, location and degree of tongue tip constriction, location and degree of tongue body constriction, velar aperture and glottal aperture are all examples of tract variables, though the proponents of **articulatory phonology** have not yet published a complete definitive set.

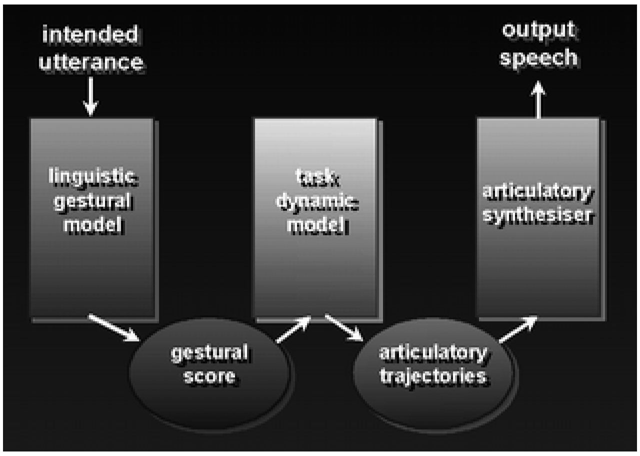

Fig.2 The formal relationship between **articulatory phonology** (the linguistic gestural model) and **the task dynamic model**. Here the two are shown as providing an input to a speech synthesis system designed to assist in testing the models.

COMPUTATIONAL ADEQUACY

It is considered unarguable that to be of any real use models of speech production and perception must be computationally adequate. Moore (1995), however, proposes a narrower concept: that moving toward more computationally adequate models in speech production and perception should be about *the exploitation of the theoretical and practical tools and techniques from speech technology for the creation of more advanced theories of speech perception and production (by humans and machines).* I find it difficult to see why models of speech production would have necessarily anything to do with speech technology, unless the idea is that by involving speech technology the model is guaranteed to be computationally adequate and complete.

Aside from obviously being more explicit than a discursive model, a computational model lends itself to

- rigorous testing, and
- transparent application.

Rigorous testing is a *sine qua non* for any theory, as is, for me, the idea that theories should be designed with some explicit application in mind. For Moore testing and application are to be in the field of speech technology — though clearly this could not be the only possibility. It is true, though, that automatic speech generation and recognition are areas of topical interest and are themselves quite rigorous; as such they form a good and useful testbed for phonetic theory. This is however less true for phonological theory, since the phonological parts of speech technology — particularly the phonology found in knowledge based automatic speech recognition — are fairly *ad hoc*, not very principled and far from rigorous in the sense that they do not adhere coherently to any established linguistic theory.

It becomes essential when considering adequacy of a computational model to distinguish between areas of speech production and perception which are best modelled as static (linguistic knowledge is an example), and which are best modelled as dynamic (motor control in production is an example) (Tatham, 1995).

One reason for this is that different approaches are optimised by the use of different computational paradigms. Thus, for example, some self-contained descriptive details in static phonology might best be expressed using a declarative paradigm. The reason for this is that static phonology (the archetypal example is **generative phonology**) is much more concerned with logical relationships between its primitives than with any dynamic phonetic realisation of those primitives, and this is precisely what the *declarative paradigm* is designed to express.

On the other hand, an algorithm for calculating fundamental frequency changes to align with planned or abstract phonological prosodic contours might be best expressed using a procedural paradigm. The reason for this is that in this situation we are concerned with a formulaic approach to step by step computation, which is what the *procedural paradigm* does best. Furthermore *the object oriented paradigm* may be optimal for computationally modelling the dynamics of speech production — this is my preferred approach to tackling a computationally adequate model of speech production dynamics.

SPEECH PRODUCTION DYNAMICS

In **action theory**, originally proposed by Fowler and described in Fowler *et al.* (1980), it was persuasively argued that earlier speech production models (called by Fowler **translation models**), such as **co-articulation theory**, had assigned too many computationally intensive procedures to phonetics and phonology (Tatham 1979). Fowler re-assigned these unrealistic procedures to a much lower level. More importantly from our point of view she modelled them as *self-organising systems*. These systems, called by Fowler **coordinative structures**, embody the knowledge of how they are to behave dynamically under a range of externally determined conditions.

Fowler endowed **coordinative structures** with hooks, enabling mid- and long-term tuning of the internal structural *knowledge*. Tatham (1995a, b) used them for short-term on the fly dynamic tuning during the utterance. The computational technique involves setting up candidate methods within the object coordinative structure and a system of *parameter passing* as the utterance unfolds.

MOTOR OBJECTS

In my preferred computational paradigm, and perhaps in more modern terms, **coordinative structures** are *motor objects,* internally arranged to respond to simple *control messaging* from outside. Modelling here falls self-evidently into the **object oriented paradigm** — each motor object is described in the model in terms of its *internal static structure* and its *dynamic response to messages.*

Thus, a coordinative structure is a motor object. The internal and private static structure of the object is a set of descriptors and a set of procedures or methods defining the object's response to externally sourced messaging. Messages directed at a motor object may bring with them parameters to be passed to the motor object to enable short-term *tuning* of the object's internally defined behaviour. I have referred elsewhere to such short-term tuning as **supervision**, and it is characterised in the **theory of cognitive phonetics** (Tatham 1990). Computationally, motor objects are arranged *class-wise* on an *inheritance basis,* thus capturing relationship generalisations between them.

GESTURES

In **articulatory phonology** terms gestures as represented in the gestural score characterise the prior planning of motor objects. They too lend themselves to computational modelling using the object oriented paradigm. It is easy to capture the internally assigned properties of a gesture as a statement of the *methods* (procedures or sub-routines) to unfold as particular messages arrive. Mid- and long-term tuning here works using the same mechanism as for the motor objects in the **task dynamic model**.

COMPUTATIONAL ADEQUACY

In this paper I have discussed how computational modelling in speech production is not a novel concept and that it exists distinctly from the requirements of modelling for speech technology. However, the fact that computational modelling is possible and that it is pursued by researchers concerned with being maximally explicit does not guarantee that it *is* computationally adequate.

Computational adequacy occurs when a computational model achieves certain criteria. Trivial among these are that

- the model should *compute* — that is, when properly programmed the program should run and conclude in an orderly fashion with nothing unexpected occurring;
- the results should adequately *reflect the phenomena* being modelled.
- Less trivially, a computational model of speech production should of itself
- *generate hypotheses* concerning its application — for Moore, in speech technology, but clearly also in the psychology and neurology of speech;
- incorporate the *means for testing*;
- indicate transparently how it might be *refuted*.

In these latter requirements the combination of **articulatory phonology** and **task dynamics** falls short of true computational adequacy. It would not be difficult, however, to arrange for these requirements to be met.

But there is one area where the model falls badly short — and this was the very area Browman and Goldstein sought to address when then conceived **articulatory phonology**. For all that the proponents recognise the fundamental difference between planning and execution, and for all that they seek to unify respectively phonology and phonetics their graphically based model and my object oriented computational version do not in the end do anything but provide a very rickety bridge between the two.

Browman and Goldstein's bridge is the use of a common, graphically oriented, mathematics; Tatham's is the use of a common computational paradigm. There is elsewhere a precise parallel in the use of neural networks to characterise, at one and the same time, related psychological and neurological phenomena — once again the bridge is a common mathematics.

For some, this is enough; it is certainly enough for us to proceed now in a formal and explicit way — something speech production and perception theory so clearly lacked in the past. [Moore’s criticism of these *earlier* models is quite right.] For the philosopher of science, though — and in particular for a dualist — there is a long way to go. The consolation is that phonetics shares this problem *with every other science concerned with characterising any aspect of human behaviour!* We should at least be pleased that it now has the potential to lag behind none of them.

REFERENCES

Browman, C.P. and Goldstein, L. (1986) Towards an articulatory phonology. In C. Ewan and J. Anderson (eds.) *Phonology Yearbook 3*. Cambridge: Cambridge University Press, pp. 219-252.

Browman, C.P. and Goldstein, L. (1993) Dynamics and articulatory phonology. *Status Reports on Speech Research*, SR-l 13. New Haven: Haskins Laboratories, pp. 51-62.

Fowler, C.A., Rubin, P. Remez, R.E. and Turvey, M.T. (1980) Implications for speech production of a general theory of action. In B. Butterworth (ed.) *Language Production*. New York, NY: Academic Press, pp. 373-420.

Moore, R. (1995) Computational Phonetics. *Proceedings of the XIIth International Congress of Phonetic Sciences*, Vol.4. Stockholm: KTH, pp. 68-71.

Saltzman, E. (1986) Task dynamic co-ordination of the speech articulators: a preliminary model. In H. Heuer and C. Fromm (eds.) *Generation and Modulation of Action Patterns*. Berlin: Springer-Verlag, pp. 129-144.

Tatham, M. (1979) Some problems in phonetic theory. In H. and P. Hollien (eds.) *Amsterdam Studies in the Theory and History of Linguistic Science IV: Current Issues in Linguistic Theory*, Vol. 9 — Current Issues in the Phonetic Sciences. Amsterdam: John Benjamins B.V., pp. 93-106.

Tatham, M. (1990) Cognitive phonetics. In W.A. Ainsworth (ed.*) Advances in Speech, Hearing and Language Processing*, Vol. 1. London: JAI Press, pp. 193-218.

Tatham, M. (1995a) The supervision of speech production. In C. Sorin, J. Mariani, H. Meloni and J. Schoentgen (eds.) *Levels in Speech Communication — Relations and Interactions*. Amsterdam: Elsevier, pp. 115-125.

Tatham, M. (1995b) Dynamic articulatory phonology and the supervision of speech production. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Vol.1. Stockholm, pp. 58-61 .