Dynamic Articulatory Phonology and the Supervision of Speech Production

Mark Tatham

Reproduced from: Tatham, M. (1995) Dynamic articulatory phonology and the supervision of speech production. In K. Elenius (ed.) *Proceedings of ICPhS 95*, 58-61. Stockholm: Royal Institute of Technology

---

ABSTRACT

Articulatory Phonology unifies the domains of phonetics and phonology, linking utterance planning and execution by common units of control. It links with the **Task Dynamic Model** of speech production, forming a smooth data pathway from the most abstract level to the physical level of articulatory configuration. This paper reviews the need for refinements to the model and proposes task *supervision* to explain some data previously overlooked.

ARTICULATORY PHONOLOGY

Articulatory Phonology was proposed by Browman and Goldstein [1] partly as an attempt to unify phonetics and phonology by treating them as 'low and high dimensional descriptions of a single system' [2]. They come together by the idea that the constraints of the physical system underlie the phonological system, and by making the units of control at the planning level the same as those at the physical level; planning and execution are seen as more closely related than in other theories of phonetics and phonology.

The plan of an utterance is formatted as a *gestural score* (see Figs. 2 and 3 for examples) which provides an input to a physically based model of speech production – the **Task Dynamic Model** [3]. The gestural score graphs locations and degrees of constrictions within the vocal tract, as well as time markers as an utterance progresses. The sequencing of gestures and their durations, and the timing relationships between the various vocal tract variables involved are critical to the score. The tract variables form a parametric framework which is manipulated 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 examples of tract variables.

As an example the gestural score for the utterance of a single [æ] would show that for a certain time the tongue body constriction is to be in the area of the pharynx and wide, with the velar aperture closed to prevent nasalisation, and the glottis closed to promote vocal cord vibration. Other tract variables mayor may not be specified depending on how crucial they are to the utterance.

The gestural score graphs the utterance plan – an abstract representation related to vocal tract movements. Since they are abstract score gestures are correctly represented as discontinuous. Thus they capture the cognitive discreteness of phonological segments while indicating how they are to be organised within the plan.

THE TASK DYNAMIC MODEL

In the **Task Dynamic Model** gestures have a functional goal, called the *task* and executed by *coordinative structures* [4]. Coordinative structures are groups of articulators or their underlying musculature which are able to *internally communicate.* The model derives its phonological perspective from the expression of functionality, and its phonetic perspective from the task specification.

Within the **Task Dynamic Model** the individual tasks are independent of each other, though they are related functionally in the gestural score representation. The model's dynamic perspective is achieved through the control of movement towards the specified physical goals. The **Task Dynamic Model** focuses on the task itself, rather than on the parts of the articulatory system involved in executing it.

PLAN AND EXECUTION

Articulatory Phonology seeks to unify phonetics and phonology though a common framework and a formal statement of low level constraints on cognitive processes. The constraints are *prior* conditions on planning; the planner knows about them in a general sense before undertaking to score a particular utterance. The constraint knowledge base is formally static in nature.

Tatham [5] attempted to show that phonetic constraints fall broadly into two types: those which are obligatory and those which are optionally controllable. The optionality of a physical constraint rests in its ability to be itself limited or enhanced. Constraints which are not optional are not able to be manipulated in this way. The recognition of two major categories of physical constraint on articulation had been proposed much earlier [6].

Some consequences arise from modelling constraints in this way:

1. the planning mechanism must be aware that a class of constraint is manipulable;
2. the manipulation takes place at a phonetic rather than phonological level;
3. the universal set of linguistically usable phonetic possibilities is augmented by the manipulative processes.

Tatham and Morton [7] claimed that the internal behavioural properties of a phonetic object (the **Task Dynamic** *coordinative structure)* could be interfered with *(re-tuned* in **Task Dynamic** terms) dynamically during the course of an utterance. The interference is planned into the utterance.

RE-TUNING THE PHONETIC OBJECT

A phonetic object has internal properties. That is, much of its realisation is internally specified rather than being computed at some higher level. This object oriented approach is a major innovation in speech production theory, proposed by Fowler [8] (**Action Theory**) at the physical level and Tatham [9] (**Cognitive Phonetics**) at the cognitive level.

Tatham's model [10] allows for some dynamic adjustment of the phonetic object's internal properties. Two purposes:

1. phonological inventory enlargement;
2. dynamic contextual variation.

Dynamic contextual variation is the ability of the system to vary the precision of the realisation of a phonetic object dependent on semantic, syntactic and phonological context. The clearest example of this is when the context of a phonetic object significantly affects the probability of perception confusion – in which case its articulatory precision is enhanced. There are many examples of this kind of cognitively determined re-tuning of a coordinative structure [10].

In the next section of this paper, the idea of supervised, rather than automatic, execution of plans is discussed within the framework of **Articulatory Phonology** and the **Task Dynamic Model**.

SUPERVISED PLAN EXECUTION

In this revision of **Articulatory Phonology** speech production planning is concerned with specifying the **Dynamic Speech Scenario**. The variability data which formed the basis of cognitive phonetic theory was inconsistent with the notion that the gestural plan might be carried through from its abstract level to the physical articulatory level. This approach only allows for simple, non-cognitively based co articulatory effects to explain why unexpected variants of gestures arise.

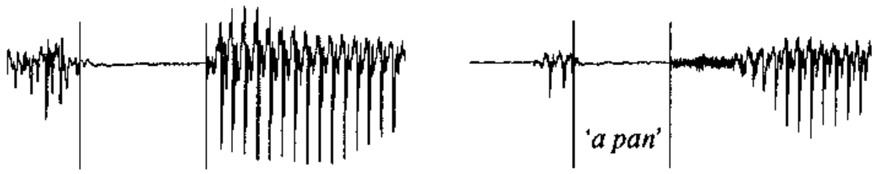

Fig.1. Waveforms of French 'une panne' and English 'a pan'. Note the aspiration in 'a pan'.

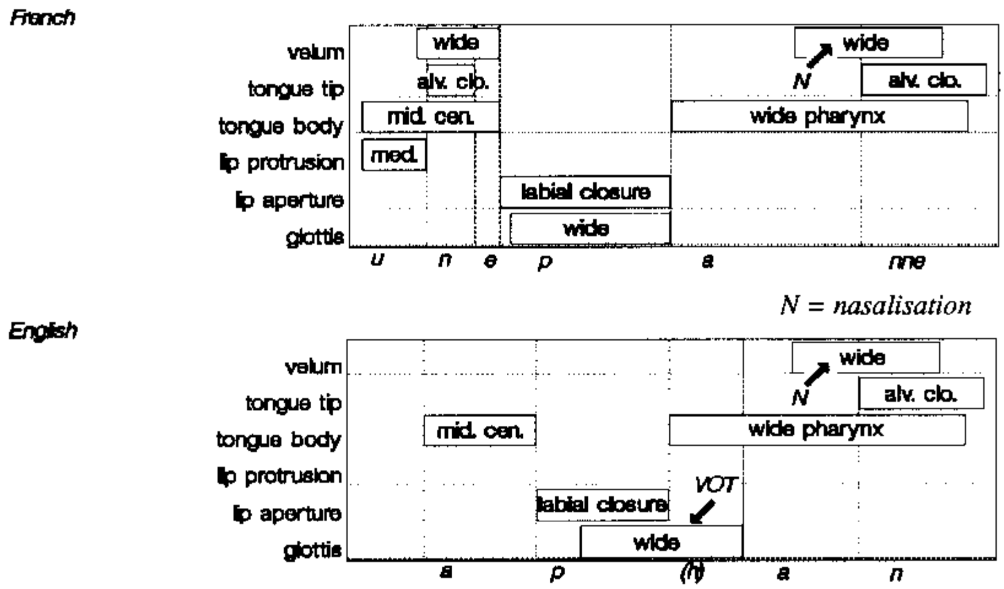

Fig.2. Unsupervised gestural scores for French 'une panne' and English 'a pan'.

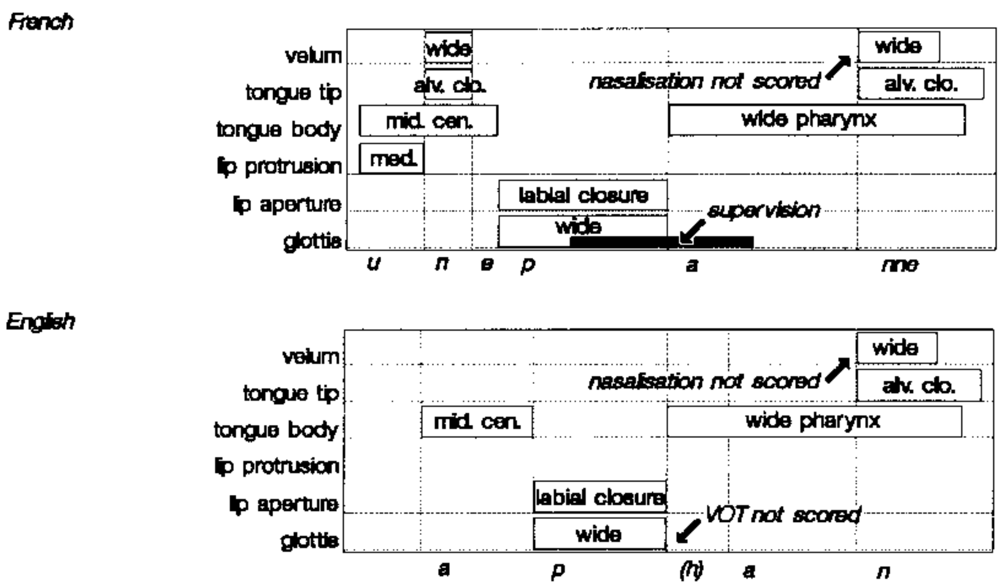

Fig.3. Supervised gestural scores for French 'une panne' and English 'a pan'.

Although the Browman and Goldstein theory implies that a carry through *is* possible, it does not adequately allow for a basis to explain the observed articulatory and acoustic facts. Because **Task Dynamics** is not able to dynamically modify its procedures, the burden of explanation rests with **Articulatory Phonology** or with an additional external component. The **Task Dynamic Model** performs better if, in addition to an underlying gestural plan, it receives an input from an external component with a *supervisory* role. The supervisory component is responsible for overseeing the **Dynamic Speech Scenario** which will unfold under the control of the model.

Tatham and Morton [7] argued this point strongly in the context of modelling the causes of observed variations in articulatory precision. The phonological gestural score cannot, on its own, enable the explanation of why precision of articulation varies during the course of utterances. And a-linguistic coarticulatory phenomena offer no explanation. The *co articulation supervisor* was introduced to allow for predictions derived from a model of perception running contemporaneously to determine areas of an utterance requiring increased articulatory precision.

Using an example from the data presented in [5] we note that in English word-initial [p] is aspirated (as in *a pan)* whereas in French word-initial [p] is not aspirated (as in *une panne).* Waveforms of these two utterances are shown in Fig. 1. **Articulatory Phonology** would account for these two utterances using the gestural scores shown in Fig. 2.

But such an account resorts to explaining the long voice onset time following English initial [p] as a deliberate and *planned* event. Many researchers however have attributed this aspiration to an involuntary coarticulatory effect. But if the effect is involuntary it cannot be planned in or planned out at a cognitive level – unless the effect falls into our earlier optional class of low level constraint. And if this the way we might choose to model we observe then we can only say that the low level constraint exists as a *universal,* but that somehow its effect is partially negated in French.

The proposal is that the basic gestural plan for both English and French should be identical in the case of *a pan* and *une panne* in the relevant parameter (glottal constriction), but that the French plan be executed *under supervision* to allow for the optional limiting of the coarticulatory constraint (Fig. 3).

Notice that it is no longer necessary to show the aspiration in the English. In the same examples we see that likewise it is no longer necessary to show nasalisation on the score. A result of introducing articulatory supervision is that we can leave phenomena such as aspiration and nasalisation to the **Task Dynamic** level – only when these phenomena are to be manipulated for the purposes discussed above do we need to deal with them in the gestural score. But because they are of a special nature (in traditional terms, not properly phonological) it is necessary to model them distinctly.

CONCLUSION

Articulatory Phonology and the **Task Dynamic Model** of speech production constitute a formidable advance in speech theory which is able to explain much data previously ignored. They do not handle well, though, subtle dynamic manipulations at the physical level during execution. This paper has argued that there is something to be gained by adding **a cognitive supervisory component** to the planning and physical components of the model.

REFERENCES

[1] 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.

[2] Browman, C.P. and Goldstein, L. (1993), 'Dynamics and articulatory phonology', *Status Reports on Speech Research,* SR-l13, New Haven: Haskins Laboratories, pp. 51-62.

[3] Saltzman, E. (1986), 'Task dynamic coordination 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.

[4] Fowler, C.A., Rubin, P. Remez, RE. 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.

[5] Tatham, M.A.A. (1995), '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

[6] Tatham, M.A.A. (1971), 'Classifying allophones', *Language and Speech*, Vol. 14, pp. 140-145.

[7] Tatham, M.A.A. and Morton, K. (1980), 'Precision', *Occasional Papers*, Vol. 23, University of Essex: Linguistics Dept., pp. 107-116.

[8] Fowler, C.A. (1977), *Timing Control in Speech Production*, Bloomington: Indiana University Linguistics Club.

[9] Tatham, M.A.A. (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.

[10] Morton, K. (1987), 'Cognitive phonetics – some of the evidence', in R Channon and L. Shockey (eds.), *In Honor of Ilse Lehiste*, Dordrecht: Foris, pp. 191-194.