THE REPRESENTATION AND ACCESSING OF LINGUISTIC AND PHONETIC KNOWLEDGE

M.A.A. Tatham*

Despite the formulation of explicit algorithms for speech processing, insufficient account is being taken of the theory of speech. A simulation is outlined which models, using a parallel connectionist object oriented paradigm, the mechanisms and cognitive processes of speech production, incorporating phonetic knowledge into the structure itself.

INTRODUCTION

Despite mathematical innovations in speech processing (Moore 1986, Bridle and Moore 1984), the problem remains of exactly what it is that novel techniques in processing and representation are to deal with (Tatham 1986). The techniques, and to a certain extent the underlying philosophy, are showing signs again of changing; perhaps this time in a more radical way.

PHYSICAL MECHANISMS IN SPEECH PRODUCTION

Lately there is a shift in phonology, phonetics and neurophysiology away from the idea of detailed cognitive control of speech production toward a more physical approach (Kelso 1977). Questions are being raised as to the role of peripheral mechanisms which possess intrinsic properties hitherto modelled as cognitively organised. On the one hand cognition is relegated to a gross central role, and on the other it is having to be re-modelled to take account of the requirement to provide a trigger for continuously varying adjustment at the periphery. Operationally, central processing is being freed of an unwieldy computational load in favour of a distributed system of hierarchically managed local co-temporal processing.

MODELLING THE PHYSICAL SYSTEM - THE OBJECT ORIENTED PARADIGM

Distributed systems of this kind are often modelled as a network of message carrying connections communicating among various nodes. The nodes themselves are simple processors with certain attributes including local private memory. A characteristic of the messages is that they communicate what is to be done, rather than how to do it. This is the reverse of the ideas held by the previous theory. Thus intrinsic to each processor, or message receiver, are methods for executing messages. Local nodes are recognised by class membership; that is, as being instances of particular classes. Classes describe the operational properties of their objects.

A distributed processing network is useful in modelling a simulation of human speech production. An object oriented language has been used for implementation because the concepts involved in the model readily fit the object oriented paradigm (Goldberg and Robson 1985).

*Centre for Cognitive Science, Essex University, Colchester UK-C04 3SQ
The speech production simulation accepts the notion of internally structured peripheral mechanisms akin to those called co-ordinative structures in the action theory literature. The model is abstract, so it is of little importance that the literature has not so far presented us with anything but trivial examples of exactly what a co-ordinative structure might be. On one occasion we might be concerned with a grouping of specific muscles, whose co-operative function is to ensure a time-governed unfolding of an articulatory gesture. The muscles themselves can be thought of as objects within the network, where the class *muscle* specifies the actions each is intrinsically capable of. Messages between objects forming the co-ordinative structure trigger appropriate action, both spatially and temporally, of the whole structure. A system of thresholds on connections and regulable sensitivity of objects to messages ensures correct differential contraction of muscles and correct relative contraction timing. A special category of message sending within the co-ordinative structure minimises runaway effects.

In the action theory model, the co-ordinative structure describes muscle grouping, and a further descriptive device, the equation of constraint, characterises the intrinsic spatial and temporal relationships between muscles. Our network of processors could equally be described as a co-ordinative structure, and its internal message sending capability, system of thresholds, etc. can be described by an equation of constraint.

It is important to realise that I am not describing an implementation of the action theory model, but a device which, as a simulation of speech production, is readily described by action theory as being equivalent to the actual human system. Simulation is an attempt to replicate real objects and events; it is not the running of a descriptive model. The real object and its simulation are open to comparison using observations motivated by a single descriptive theory.

At various points within the network hooks are placed which together constitute an interface between this system and its control mechanism. Two classes of hook are incorporated:

(a) a class with a single instance, used for messaging the entire structure, and

(b) a class with several instances, used for messaging interference with the intrinsic operational relationships between muscle-objects within the structure.

The network of processing nodes described above simulates only part of the peripheral production system characterised as a co-ordinative structure with its equation of constraint. The speech production simulation has many such devices nested hierarchically - the entire structure being a single object. Each device below this object has the same properties: local processing power, memory, programs defining actions, programs for system stability control, message sending and receiving capabilities. Running the system consists of sending an initiating message and many control messages assigned to the task of disturbing the system’s internal organisation.

The object that is the system, and the objects within the system, know to be what they are and to behave as they do - once triggered to spring to activity as a system by the initial message. They are not, however, a co-ordinative structure unless described as such: that is, the notion co-ordinative structure, like other similar notions, is a descriptive construct of the theory, not a property of the object being observed.
SYSTEM CONTROL

Action theory uses two types of control signals: gross control involving the sending of a command to co-ordinative structures, and tuning control involving signals which modify terms within equations of constraint to alter the actions intrinsic to co-ordinative structures. Equations of constraint are designed such that automatic compensation of values of terms other than those tuned takes place. The model is confused in that equations of constraint are sometimes presented as descriptive devices and sometimes as actual properties of the system.

The simulation provides for entry gateways or hooks into networks which correspond to co-ordinative structures within the nested system. As described earlier the two classes of hook satisfy the descriptive action theory model. Gross control messages entering the system are automatically actioned by objects with matching method attributes. Objects can be individual nodes, or nodes grouped within the hierarchy; the whole network constituting an object which responds to the message signalling readiness for speech.

INCORPORATING LINGUISTIC KNOWLEDGE

Messages calling for gross action are equivalent to extrinsic allophones in descriptive linguistics theory. They correspond to objects concatenated as strings which exit from a performance phonology. Phonology characterises sound patterning in a language (Chomsky and Halle 1968). In transformational generative grammar description is confined to an enumeration of primitives and rules as a knowledge base on which an unspecified algorithm draws to encode sentences as abstract sound segments (the extrinsic allophones), ready to be translated into actual sounds by phonetics. A performance phonology incorporates both the knowledge base of generative grammar and the encoding algorithm.

In the simulation – unlike the specification of segments in phonology – messages controlling the activation of the production network are only minimally parameterised. Processors within the network respond if they are intrinsically able to do so, and not otherwise; that is, if an object's complement of operations includes one which matches a message then firing will occur. In this way, the device could be said to self-parameterise. Put another way, it might appear to the observer that the final behaviour of the device was being controlled parametrically; in fact the device knows to operate parametrically despite non-parametric control.

Fine operations within the device are initiated by the second class of message, corresponding to the tuning of co-ordinative structures. Until recently phonetic theory made no provision for this subcategory of control which falls unambiguously within the cognitive domain (Tatham 1987). For the linguist cognition stops with phonology, but the idea of a cognitive element within phonetics has been gaining ground (Tatham 1980, 1984).

Phonology enumerates a knowledge base of mappings from abstract underlying objects (phonemes), set up to define morphemes minimally, onto surface objects (extrinsic allophones) which embody all information required to trigger an automatic phonetics to produce matching soundwaves. It is sometimes said to produce an idealised requirement for phonetics to execute. Phonetics introduces degradation into the translation process.
ascribing this to inadequacies of the physical system.

COGNITION IN PHONETICS

The theory of cognitive phonetics has been developed based on evidence (Morton 1987) indicating that

(a) one kind of variability in the articulatory and acoustic signals is systematic, yet has no obvious physical explanation, and
(b) successful speech depends on decisions to manipulate what were previously thought of as involuntarily generated encoding errors.

The claim is that cognitive processes are involved in speech which cannot be assigned to phonology. Processes are modelled which represent decisions to generate overlays to phonological requirements to produce systematic variations superimposed on the phonetic output. They consist of evaluations of information from many sources (including phonology, a phonetics knowledge base, a predictive perceptual model, a pragmatic encoding component) to compute a signal to adjust the usual physical phonetic processes. The resultant composite waveform is such that this information can be decoded perceptually.

The simulation network models phenomena described by both phonology and cognitive phonetics (Tatham 1985), with the cognitive phonetics network responsible for generating the fine tuning messages sent to objects within the phonetic network.

CONCLUSION

In adopting the technique described in this paper the mechanism of speech production has been modelled by the network and the knowledge required to produce speech has been incorporated into the mechanism itself. To a large extent this dispenses with the traditionally separate representations of mechanism and knowledge.

REFERENCES


