Introduction

The Second Conference on Laboratory Phonology was organized by the Department of Linguistics at the University of Edinburgh and took place in Edinburgh from 30 June to 3 July 1989. The conference's primary aim was to further the general intellectual agenda set by the ambitiously named First Conference, which brought together researchers in phonological theory and experimental phonetics to discuss the increasing convergence of their interests. An important secondary aim was to bring together researchers from both sides of the Atlantic: whereas the first conference was almost exclusively American, the second had significant delegations from several European countries and Japan as well as the USA. This book is the record of the second conference.

We say "record" rather than "proceedings" because the papers collected here are neither all nor only those presented in Edinburgh. As in the first conference, the main papers were circulated in advance and invited discussants gave prepared comments at the conference; this format is reflected in the organization of this volume as a series of chapters with commentaries. However, all of the main papers were formally refereed after the conference, and have been revised, in the light of both referees' comments and discussion at the conference itself, for publication in their present form. Moreover, for a variety of reasons, quite a number of contributions to the conference (listed at the end of this introduction) do not appear in the volume, and as a result, the volume's organization is rather different from that of the conference program.

We have grouped the chapters into three main sections, which we have called Gesture (on temporal coordination of articulatory gestures), Segment (on the nature and classification of segments), and Prosody (on certain aspects of the prosodic organization of speech). At the beginning of each section we have included a tutorial chapter, presenting a synopsis of recent

theoretical developments that are presupposed in some or all of the papers in the section, which we hope will make the volume accessible to a wider range of readers. This reorganization means that certain contributions appear in a context rather different from that in which they were presented at the conference: for example, the chapter by Cutler was originally prepared as a general commentary on a series of papers dealing with "The Segment," and the commentary by Vogel on Pierrehumbert and Talkin's paper was originally prepared as a general commentary on a series of papers dealing with prosody and prosodic effects on segmental realization. We hope that in imposing the organization we have chosen we have nevertheless succeeded in preserving the sense of productive debate that was present at the conference.

Cutting across the organization into three subject areas we see three major issues: the methodology and design of laboratory research on phonology; the psychological reality of lexical and phonological representations; and the nature of the phonology–phonetics "interface." On the question of methodology and design, the papers collected here seem to show a growing consensus. Fujimura finds Pierrehumbert and Talkin's paper an exemplary case of how laboratory phonology should be carried out, but the rigor exhibited in their experimental descriptions is seen in several of the papers, such as Lahiri and Marslen-Wilson's, Beckman, Edwards, and Fletcher's, Browman and Goldstein's, and Nolan's. We feel that papers such as these set the baseline for future work in laboratory phonology. Fujimura, Vogel, and Cutler all note that if phonology is to be successfully tested in the laboratory, scientific rigor is essential. This applies to the use of terminology as well as to laboratory practice; some specific problems in this area are raised in the contributions by Vogel and by Hewlett and Shockey.

The second theme – the question of psychological reality of lexical representations – is explicitly addressed only in the papers by Lahiri and Marslen-Wilson and by Cutler, but we feel that it is an underlying issue throughout the volume. Lahiri and Marslen-Wilson, in an approach that departs from much recent work in laboratory phonology, use listeners' ability to recognize sounds in a gating-paradigm experiment to argue for abstractness in the lexical representation. Cutler reviews the considerable body of evidence for the psycholinguistic relevance of the segment. Both papers raise a fundamental problem: to what extent is it possible to relate models of phonological and phonetic representation, which are at the center of the debate in the rest of this volume, to what speakers actually do when they are producing an utterance? Lahiri and Marslen-Wilson suggest that psycholinguistic evidence can be used successfully to empirically evaluate the claims of theoretical phonology, whereas Cutler suggests that phonology and psycholinguistics may be fundamentally different exercises, and points to the "orthogonality of existing psycholinguistic research to phonological issues."

Some resolution of this question is obviously crucial for any approach to phonology that looks to the laboratory and to detailed observations of speech production for evidence and evaluation.

The central issue in laboratory phonology, however, is and remains the long-standing problem of the relation between phonetics and phonology. Where does one end and the other begin – or, to be more specific, can phonetics be attributed entirely to neuromotor aspects of the vocal mechanism, or is there a case for inclusion of a phonetic implementation component in the grammar?

The papers on gestural coordination and "task dynamics" in the first section of the volume directly address the general issue here. These papers ask whether a gesture-based phonology couched in terms of task dynamics (the theory of skilled movement developed at Haskins Laboratories and presented here in a tutorial chapter by Hawkins) provide a basis for understanding the phonology–phonetics interface. Browman and Goldstein, and to some extent also Beckman, Edwards, and Fletcher, argue that this is a powerful approach, but notes of caution are sounded by both Fujimura and Kingston in their commentaries on these papers. Kingston, for example, while applauding the descriptive elegance of Browman and Goldstein's gestural phonology, raises serious doubts about its explanatory adequacy; he claims that the model is too powerful and lacks the constraints that would permit it to predict only those things that are found to occur in speech.

More specific aspects of the phonology–phonetics interface question are also raised by the papers in the first section of the volume. Hawkins, in her commentary on Browman and Goldstein, comments on the need to be specific about which aspects of the realization of an utterance can be attributed to the gestural score, and which to the model of motor control. For example, should the gestural score contain all the language-particular phonetic characteristics of an utterance? If so, this would involve incorporating a large amount of (redundant) phonetic detail into the gestural score. If not, it has yet to be demonstrated how such aspects of phonetic realization could arise from a task-dynamics model of motor control. A related issue is the degree of specification associated with the spatial and temporal targets in the gestural score. Browman and Goldstein propose that certain aspects of the gestural score can be left unspecified and the detailed instantiation of any particular target determined by the task-dynamics model. In a somewhat different approach to a comparably complex problem of phonetic variability, the papers by Beckman, Edwards, and Fletcher and by Pierrehumbert and Talkin suggest that a notion of "prosodic modulation" – effects of phrase-level and word-level prosodic structure on the laryngeal component in the production of consonants – may make it unnecessary to posit a large number of separate rules for superficially independent types of phonetic variability.

In the second section of the volume the question of the phonology–phonetics "interface" is attacked on two further fronts. First, papers by Local and Nolan (with associated commentaries by Hayes, Browman, Kohler, and Ohala) deal with assimilation: to what extent is assimilation the result of a phonological "rule" rather than a phenomenon emerging from the organization and coordination of articulator variables in the execution of an utterance? The key data discussed in these papers are instrumental measurements showing that assimilations are commonly no more than partial; the theoretical models which are brought to bear on such findings by Local, Hayes, and Kohler are extremely varied, and themselves make a number of predictions which are open to future empirical investigation. The second question addressed here, in the papers by Ohala and Local (and the comments by Clements, Kohler, and Rossi), is whether it is justified to posit a role for the segment in a phonological representation. This is not a new issue, but as Broe points out in his tutorial chapter, the power of nonsegmental representations in accounting for the sound pattern of languages has been recognized by more and more investigators over the last couple of decades.

Finally, in the third section of the volume, we see instrumental phonetic evidence treated as central in the search for appropriate phonological descriptions – for example, in Arvaniti's paper on the phenomena to be accounted for in any phonological description of Greek stress. We feel that the central role accorded to instrumental data in these papers has somewhat equivocal implications for laboratory phonology in general. It could be seen as pointing the way to a new conception of the phonology–phonetics interface, or it could simply show that segmental and prosodic phenomena really are different. That is, phonological descriptions of prosodic phenomena such as intonation have tended (in the absence of a pretheoretical descriptive construct like the orthographically based segment) to differ from one another in ways that cannot be resolved by reference to common conceptions of phonology and phonetics. Only with the application of instrumental evidence to questions of phonological organization rather than speech production have we begun to approach some consensus; as Beckman and Pierrehumbert note in their commentary, the modeling of high tones (fundamental-frequency peaks) is "one of the success stories of laboratory phonology." It remains to be seen how widely this success can be extended into the realm of gestures and segments.

Conference contributions not included in this volume

Papers

Lou Boves, "Why phonology and speech technology are different"
Jean-Marie Hombert, "Nasal consonants and the development of vowel nasalization"
Brian Pickering and John Kelly, "Tracking long-term resonance effects"
Elisabeth Selkirk and Koichi Tateishi, "Syntax, phrasing, and prominence in the intonation of Japanese"

Commentaries

Stephen R. Anderson
Gösta Bruce
Stephen D. Isard
Björn Lindblom
Joan Mascaró

Posters

Anne Cutler
Janet Fletcher
Sarah Hawkins
Jill House
Daniel Recasens
Jacques Terken
Ian Watson

Section A
Gesture

1
An introduction to task dynamics

SARAH HAWKINS

1.1 Motivation and overview

The aim of this paper is to describe for the nonspecialist the main features of task dynamics so that research that uses it can be understood and evaluated more easily.* Necessarily, there are some omissions and simplifications. More complete accounts can be found in the references cited in the text, especially Saltzman (1986) and Saltzman and Munhall (1989); Browman and Goldstein (1989, 1990) offer clear descriptions that focus more on the phonologically relevant aspects than the mathematical details. The task-dynamic model is being developed at the same time as it is being used as a research tool. Consistent with this paper's purpose as a general introduction rather than a detailed critique, it mainly describes the current model, and tends not to discuss intentions for how the model should ultimately work or theoretical differences among investigators.

Task dynamics is a general model of skilled movement control that was developed originally to explain nonspeech tasks such as reaching and standing upright, and has more recently been applied to speech. It is based on general biological and physical principles of coordinated movement, but is couched in dynamical rather than anatomical or physiological terms. It involves a relatively radical approach that is more abstract than many more traditional systems, and has proved to be a particularly useful way of analyzing speech production, partly because it breaks complex movements down into a set of functionally independent tasks.

Task dynamics describes movement in terms of the tasks to be done, and the dynamics involved in doing them. A single skilled movement may involve

*I thank Thomas Baer, Catherine Browman, and Elliot Saltzman for helpful comments on earlier versions of this paper.

several discrete, abstract tasks in this model. Speech requires a succession of skilled movements, each of which is modeled as a number of tasks. For example, to produce English [ʃ], the precise action of the tongue is critical, the lips are somewhat rounded and protruded, and the vocal folds are moved well apart to ensure voicelessness and a strong flow of air. In addition to controlling respiratory activity, therefore, there may be at least five distinct tasks involved in saying an isolated [ʃ]: keeping the velopharyngeal port closed, and controlling the degree of tongue constriction and its location, the degree of lip protrusion, the size of the lip aperture, and the size of the glottal aperture. The same articulator may be involved in more than one task. In this example, the jaw contributes to producing both the correct lip aperture and the correct tongue constriction. To add to the complexity, when the [ʃ] is spoken as part of a normal utterance, each of its tasks must be carried out while the same articulators are finishing or starting tasks required for nearby sounds.

The sort of complicated tasks we habitually do with little conscious effort – reaching for an object, lifting a cup to the mouth without spilling its contents, speaking – are difficult to explain using physiological models in which the angles of the joints involved in the movement are controlled directly. These models work well enough when there is only one joint involved (a "single-degree-of-freedom" task), because such movements are arc-shaped. But even simple tasks usually involve more than one joint. To bring a cup to the mouth, for example, the shoulder, elbow and wrist joints are used, and the resultant movement trajectory is not an arc, but a more-or-less straight line. Although describing how a straight-line trajectory could be controlled sounds like a simple problem, it turns out to be quite complicated. Task dynamics offers one way of modeling quasi-straight lines (Saltzman and Kelso 1987).

A further complication is the fact that in these skilled "multi-degree-of-freedom" tasks (involving more than one joint) the movement trajectory usually has similar characteristics no matter where it is located in space. When reaching outwards for an object, the hand tends to move in a straight line, regardless of where the target is located with respect to the body before the movement begins, and hence regardless of whether the arm starts by reaching away from the body, across the body, or straight ahead. This ability to do the same task using quite different joint angles and muscle contractions is a fundamental characteristic of skilled movement, and has been called "motor equivalence" (Lashley 1930; Hebb 1949).

One important property related to motor equivalence is immediate compensation, whereby if a particular movement is blocked, the muscles adjust so that the movement trajectory continues without major disruption to attain the final goal. Immediate compensation has been demonstrated in

"perturbation" experiments, in which a moving articulator – usually an elbow, hand, lip, or jaw – is briefly tugged in an unpredictable way so that the movement is disrupted (e.g. Folkins and Abbs 1975; Kelso *et al.* 1984). The articulators involved in the movement immediately compensate for the tug and tend to return the movement to its original trajectory by adjusting the behavior of the untugged as well as the tugged articulators. Thus, if the jaw is tugged downwards during an upward movement to make a [b] closure, the lips will compensate by moving more than usual, so that the closure occurs at about the same time in both tug and no-tug conditions. These adjustments take place more or less immediately (15–30 msec. after the perturbation begins), suggesting an automatic type of reorganization rather than one that is under voluntary, attentional control. Since speaking is a voluntary action, however, we have a reflexive type of behavior within a clearly nonreflexive organization. These reflexive yet flexible types of behavior are called "functional reflexes."

Task dynamics addresses itself directly to explaining both the observed quasi-straight-line movement trajectories and immediate compensation of skilled limb movements. Both properties can be reasonably simply explained using a model that shapes the trajectories implicitly as a function of the underlying dynamics, using as input only the end goal and a few parameters such as "stiffness," which are discussed below. The model automatically takes into account the conditions at the start of the movement. The resulting model is elegant and uses simple constructs like masses and springs. However, it is hard to understand at first because it uses these constructs in highly abstract ways, and requires sophisticated mathematics to translate the general dynamical principles into movements of individual parts of the body. Descriptions typically involve a number of technical terms which can be difficult to understand because the concepts they denote are not those that are most tangible when we think about movement. These terms may seem to be mere jargon at first sight, but they have in fact been carefully chosen to reflect the concepts they describe. In this paper, I try to explain each term when I first use it, and I tend to use it undefined afterwards. New terms are printed in bold face when they are explained.

1.2 Basic constructs

The essence of task dynamics, making it distinct from other systems, is implicit in its name. It describes movement in terms of the tasks to be done, using dynamics that are specific to the task but not to the parts of the body that are doing the task. A **task** is generally a gesture involving the control of a single **object**, or abstract representation of the actual thing being controlled. For example, the object represents in abstract terms the position of a hand

relative to a target in a reaching task, or an articulatory constriction in a speech task. Such "objects" are called **task variables**; they describe the type of task required. In order to realize the movement, the abstract disembodied task must be converted into a set of parameters appropriate for the part of the body that will perform the task, and finally into movements of the actual articulators. Currently, only one type of task (described below) is modeled for speech, so most recent discussions make no effective distinction between task variables, which are associated with the disembodied task, and the variables associated with specific body parts, which are called **tract variables**, as in vocal-tract variables (sometimes, local tract variables). In current formulations, tract variables are the dimensions allowing particular vocal-tract constrictions to be specified.

In a reaching task, it is the position of the target that is being reached for that is most important, so the task is defined on a system of coordinates with the task target at the common origin. The position of the abstract hand with respect to the target is the variable that is being controlled. In speech, similarly, the task is defined in terms of the location (place) and cross-sectional area (degree) of an ideal constriction. This nonspecific task is then transformed appropriately for a specific vocal-tract constriction. For example, the degree of openness suitable for some lip configuration is regarded in the task-dynamic model as a requirement to achieve a particular lip aperture, not as a requirement for lips and jaw to be in a certain position. The lips, in this case, are the **terminal devices**, or **end effectors**, since their position directly defines the lip aperture; the upper and lower lips and the jaw together form the **effector system** – the set of organs that includes the terminal device and gets it to the right place at the right time. Lip aperture itself is a tract variable.

Thus, whereas earlier models use coordinates of either **body space** or **articulator space** to specify the mathematics of movement, and so describe movement in terms of the position of a physical object with respect to the body or in terms of the angles of joints respectively, task dynamics begins by defining movement in terms of an abstract **task space**, using spatial coordinates and equations of motion that are natural for the task, rather than for the particular parts of the body that are performing the task. Coordinating the different parts of the body to produce the movement is done entirely within the model.

In summary, to carry out the transformations between task, body, and articulator spaces, the task-dynamic model translates each task-specific equation into equivalent equations that apply to motion of particular parts of the body. The first transformation is from the abstract task space to a more specific (but still relatively abstract) body space. The task-space equation is defined solely in terms of the type of task: the place and degree of