1 Introduction

MELISSA A. REDFORD

1.1 Speech production: What it is and why it matters

Speech is the principal mode used to convey human language – a complex communication system that creates cohesion (and division) among us; a system that allows us to structure and build knowledge and social-cultural practices through time. Speech is an activity, defined at its core by an acoustic signal that is generated by the speaker and transduced by the listener. Activities unfold through time, and so does speech. When speech is defined in terms of the signal, it is as a time-varying acoustic waveform, amplitude and frequency modulated. The modulations are due to movements of the speech organs (articulators) in service of the message to be conveyed. Since there is no way to move except through time, the generation of speech constrains how the message is structured: the output must be roughly linear, even though the complex thoughts and feelings we want to communicate are not.

The relationship between complexity and the quasi-linearity of action was famously explored by Karl Lashley (1951). Lashley's concern was to explain "the existence of generalized schemata of action which determine the sequence of specific acts, acts which in themselves or in their associations seem to have no temporal valence (122)." To do so, Lashley devoted over half of his presentation, intended for an audience of neuroscientists, to language. He argued, based on the evidence from language, that the control structures he sought to define (i.e., schemata) must be hierarchical in their organization:

> I have devoted so much time to the discussion of the problem of syntax, not only because language is one of the most important products of human cerebral action, but also because the problems raised by the organization of language seem to me to be characteristic of almost all other cerebral activities. There is a series of hierarchies of organization; the order of vocal movements in pronouncing the word, the order of words in the sentence, the order of sentences in the paragraph, the rational order of paragraphs in a discourse.
>
> (1951: 121)

The Handbook of Speech Production, First Edition. Edited by Melissa A. Redford.

Thus, well before Chomsky's (1959) equally famous critique of Skinner's (1957) book, *Verbal Behavior*, Lashley argued that the hierarchical structure of language implied hierarchical representations of language. No doubt that ideas about hierarchical structure and representations were in the intellectual ether of the time, so the issue here is not so much who was first to insist on the importance of these, but instead to note that Lashley linked such representations to action. In so doing, he suggested an interdependency between language and speech that is underappreciated in modern linguistics and psychology.

Insofar as thoughts and feelings are to be externalized and shared between people, they must be structured so that they can be realized in time (i.e., in a serial order). Pinker and Bloom (1990: 713) noted as much when thinking about the evolution of language: "grammars for spoken languages must map propositional structures onto a serial channel." Pinker and Bloom suggest that from this constraint many of the features that Lashley found so intriguing about language follow. Just as Lashley began his paper with reference to "an unpronounceable Cree word" that is analyzed into a "verbal root *tusheka*, 'to remain,' and the various particles which modify it" to produce the meaning "may it remain with you" (1951: 112–113), Pinker and Bloom referred to complex verbal morphology in Native American languages when listing the various devices by which propositions are mapped onto a serial channel: "(v)erb affixes also typically agree with the subject and other arguments, and thus provide another redundant mechanism that can convey predicate-argument relations by itself (e.g., in many Native American languages such as Cherokee and Navajo)" (1990: 713). For Lashley, the fact that particular subsequences within a long Cree word had particular meanings that contributed to a larger one suggested that the "generalized schemata of action" could be conceived of as a hierarchical structure, with smaller units embedded in larger ones. Since the argument was based on language, it is reasonable to conclude that meaning structures action. But what if we were also to emphasize that the units exist to make action possible? With such an emphasis, we might just as well conclude that action structures meaning.

The focus on grammar in the study of language has relegated speech production and perception to the periphery of what is fundamental about language. Speech is merely the dominant mode of communication. Grammar is central. Grammar is structure. But if, as suggested here, the structure that our communication system takes is due in large part to its dominant mode of transmission, then other questions emerge: To what extent are speech and language representations independent? From a developmental perspective it is hard to see exactly where the divide should be drawn since the proximal target of early language acquisition is the ability to produce a sound shape that is linked to some meaning (i.e., a word or construction). Later on, the child must figure out how to link stored units (words) into longer and longer sequences (sentences) that will be "read off" by some speech production mechanism as fluent speech. Shouldn't the speech production mechanism then contribute to how these sequences are chunked for output? Or does linguistic structure (e.g., syntax) wholly determine the chunking, thereby driving the development of the mechanism? And what of the action sequences that are the

realization of fluent speech? Is the plan (generalized schemata) that guides extended action defined by units that are independent of action? If so, then how is sequential action executed?

There cannot be layers of abstract linguistic representations all the way down, so to speak, since at some point explicit motor commands are needed to instantiate movements that produce an acoustic signal. The instantiation of planned actions is considered the execution stage of speech production. At this stage, it is generally assumed that the plan is coded as a sequence of realizable motor goals, which are not executed individually during fluent speech, but instead in practiced sequences, that is, as "motor programs" (Keele 1968). Still, the goals themselves are usually considered the critical link between the plan and actual movement. So, what precisely are speech motor goals? There is no current consensus on this question other than to point to those vocal tract configurations that, after centuries of work in phonetics, we have come to associate with particular speech sounds. But whether it is the vocal tract configurations or the sounds themselves that define the goals, depends on your theory. If it is the former, the goal is a motor command (or set of commands) that realize(s) a specific vocal tract constriction (e.g., gesture). If it is the latter, the goal is a specific, acoustically defined set of features. Acoustic goals require an extra step in execution because they must be translated into motor commands with reference to an auditory-motor map set up during development. Nonetheless, most of the contributors to this Handbook who address motor control directly assume acoustic goals since these may allow for speech movements that are more adaptable to the speech context. No matter how the goals are defined, the question that animates research on speech execution is what kinds of feedback and feedforward processes are involved in making sure that the goals are achieved. The stunning complexity of speech action entails that this question be explored from many angles; competing hypotheses regarding the nature of goals ensures that this happens.

The concept of motor goals and their association with particular vocal tract configurations/speech sounds might suggest to the reader that every aspect of speech production is centrally controlled. This is highly unlikely, though, given the evidence that neurotypical adults will spontaneously adapt to certain unanticipated perturbations of the speech articulators (see Chapter 11). Many suggest, including several contributors to this Handbook, that articulatory coordination, which gives rise to particular vocal tract configurations, emerges from dynamical principles. This does not say that the configurations (or their acoustic consequences) can't be goals. It merely says that their achievement cannot depend entirely on central control. But those who advocate a dynamical systems approach to movement coordination are also often wary of goals and the language of control (e.g., Chapter 8). Similar to our earlier question about where to draw the boundary between language and speech representations, these researchers wonder where to draw the boundary between executive function and dynamics. Put another way, we all acknowledge that speech is an intentional activity, but the field has not yet determined where top-down control yields to emergent behavior.

It will hopefully be clear by this point that questions about dynamics, control, and even speech planning cannot be seriously addressed absent a detailed appreciation of speech action. Consider, once again, another extract from Lashley's (1951) paper:

> Pronunciation of the word "right" consists first of retraction and elevation of the tongue, expiration of air and activation of the vocal cords; second, depression of the tongue and jaw; third, elevation of the tongue to touch the dental ridge, stopping of vocalization, and forceful expiration of air with depression of the tongue and jaw. These movements have no intrinsic order of association. Pronunciation of the word "tire" involves the same motor elements in reverse order.
>
> (116)

Leaving aside the oversimplification of articulatory movements described here and the lack of reference to biomechanical linkages within and across time, it is decidedly *not* the case that "the word 'tire' involves the same motor elements [as 'right'] in reverse order." In actual fact, interarticulatory coordination varies substantially by position and context, and especially by position within a syllable (see Chapter 7). There is no sound for which this may be more true than the English *r* that Lashley inadvertently made prominent by having it occur in both syllable-onset and syllable-offset position in his example.

Although Lashley's (1951) true goal was to demonstrate the recombinatorial (i.e., generative) nature of language, his description of vocal action is highlighted here to argue that a misunderstanding of the details of speech has both theoretical and practical consequences. A specific consequence of Lashley's misunderstanding is that it suggests a one-to-one relationship between articulatory configurations and phonemes ("letters" in Lashley's words), which in turn suggests a long-discredited model of execution that proceeds phoneme-by-phoneme. Lashley may have rescued himself from a commitment to the most naive version of such a model by also noting that "letters" are embedded in a "word" and thereby given context. But embedding movements in a context is not quite the same as understanding that interarticulatory coordination is never independent of context. The nuance is important. If taken seriously, it suggests a theory of speech production that is built up from biomechanics, goals, and motor programs. It also suggests *language* representations, such as those proposed in Articulatory Phonology (Browman and Goldstein 1986), that are very different from the familiar atemporal units assumed in modern phonological theory and by Lashley himself. Of course, it is possible to appreciate the details of articulation and reject the types of representation proposed in Articulatory Phonology (a number of contributors to this Handbook do); but, by doing so, one incurs the responsibility of proposing models to bridge the hypothesized divide between speech and language (see Chapter 19 for an example of how this might be achieved).

In summary, Lashley (1951) emphasized that behavior is planned action and argued that the plan for complex behaviors – read serially ordered actions – is best described by hierarchical models, where smaller units of action are embedded in larger ones. I accept this as true, but have also suggested that there are substantial

benefits to looking at the relationship between the plan and the behavior from the other direction: where it is the constraints on behavior – the anatomy of its realization and its fundamental temporality – that define the units of action and so contribute to structuring a plan that must also encode meaning. When viewed in this way, speech production is no longer peripheral to language; it is central to its understanding. The chapters in this Handbook are intended to provide readers with a broad base of knowledge about speech production research, models, and theories from multiple perspectives so that they may draw their own conclusions on the relationship between speech and language.

1.2 Organization of the Handbook

This Handbook is organized from the most concrete aspect of speech production to its most abstract; from an emphasis on the physical to an emphasis on the mental. Between these poles we consider the organization and control of speech behavior with reference to dynamical principles and underlying neural structures. All of the chapters engage with behavior; many focus on kinematics, some adopt a computational approach, and some a cross-linguistic one. Because speech production is a skill that takes over a decade to acquire and is easily disrupted by injury or disease, many contributions were solicited from researchers who would engage with a particular topic in speech production from a developmental and/or clinical perspective. Gary Weismer and Jordan Green (Chapter 14), referencing Bernstein and Weismer (2000), argue that "speech production and perception models/theories should have the capacity to predict and/or explain data from *any* speaker or listener, regardless of his or her status as 'normal' or communicatively-impaired." They worry explicitly about the practice of refining speech production models and theories "for 'normal' speakers, with minimal attention paid to speakers with communicative disorders" and, I would add, to development. In addition to the inherent explanatory weakness that results from such practice, models and theories that are perfectly tuned to the typical adult speaker also subvert an important function of basic science: to build a foundation for applied scientific advances. Simply put, individuals with disordered speech and children with immature speech skills provide important data on speech behavior that models and theories of speech production should incorporate for intellectual reasons as well as for practical ones. The organization of the Handbook accommodates this point of view by interleaving chapters focused on disorder and/or development with chapters focused on typical adult behavior.

Whether from a clinical, developmental, or typical adult perspective, each chapter in this Handbook addresses some important aspect of speech production. Many contributions focus on theory, and either suggest revisions to dominant frameworks or extend existing ones. Many contributions also make clear the applied consequences of basic research on speech production; several others focus on questions related to the speech–language divide. The following

overview of the chapters in each Part is provided to better orient the reader to the specific content covered in the Handbook.

1.2.1 The speech mechanism

Our anatomy, physiology, and resultant biomechanics define the action that is used to create speech. Phonation is dependent on the constant airstream supplied by the lungs. The waveform generated by vibrations of the vocal folds is further modulated by pharyngeal constrictions, by the movement of the tongue and lips which are biomechanically linked to the jaw, and by virtue of acoustic coupling (or not) with the nasal cavity. The five chapters in Part I provide the reader with a detailed understanding of the action of all of these articulators. But each of these chapters does much more than describe the many muscles involved in speech movement. Pascal van Leishout (Chapter 5) adopts a comparative perspective to present the anatomy and physiology of the lips and jaw in the context not only of speech, but also of the other oral-motor functions to which they are adapted. He concurs with MacNeilage (Chapter 16) and others that "the way we use oral anatomical structures in our communications has been adapted from their original primary use, namely to support feeding and breathing." Brad Story (Chapter 3) and Kiyoshi Honda (Chapter 4) make explicit connections to acoustic theory in their respective chapters on voice production and on the tongue and pharynx. Their chapters are also aimed at updating our understanding of the mechanism: Story provides us with a modern view of the vocal folds as a self-oscillating system, describing computational models of phonation that formalize this view; Honda invites us to jettison our simple tube-model understanding of vocal tract resonances and to consider the contribution of hypopharyngeal cavities to the acoustics of speech (and singing), providing us with compelling 3D MRI images to make his point. Jessica Huber and Elaine Stathopoulos (Chapter 2) and David Zajac (Chapter 6) connect us to language – utterances and oral versus nasal sounds, respectively – while also providing us with information about speech breathing and the velopharyngeal port/nasal cavity: Huber and Stathopoulos document important changes in lung capacity and breath control that occur across the lifespan and in elderly speakers with Parkinson's disease; Zajac documents structural changes in the development of the upper vocal tract and velopharyngeal function in child and adult speakers with typical morphology as well as in those with cleft palate.

1.2.2 Coordination and multimodal speech

The articulators come together, moving into and out of the configurations we associate with specific speech sounds, over and over again through time. Individuals come together to exchange speech, first as the perceiver then as the generator, over and over again through time. This coordination of articulatory movement within and across individuals has consequences for our understanding of speech production processes, as the contributors to Part II of this Handbook

make clear. Philip Hoole and Marianne Pouplier (Chapter 7), Fred Cummins (Chapter 8), Eric Vatikiotis-Bateson and Kevin Munhall (Chapter 9) consider coordination at different levels of analysis from the perspective of dynamical systems. Hoole and Pouplier focus on interarticulatory coordination at the level of the segment and across segments, showing how timing patterns vary systematically by language and by syllable position within a language. Moreover, they embed their discussion of these phenomena within an Articulatory Phonology framework, providing the reader with a sense of the theory; its primitives, emergent units, and the coupling dynamics referenced to account for positional effects. Cummins explores rhythm in speech and language; a phenomenon that binds movement through time and speakers in dialogue. He reviews the various historical attempts to test the rhythm class hypothesis, and argues that "the vigorous pursuit of a classificatory scheme for languages on rhythmic grounds alone has probably enjoyed an undue amount of attention, with little success." He advocates that we consider studying phenomena that relate more intuitively to what we might identify as having high degrees of temporal structure, including choral speaking and dyadic interactions. Vatikiotis-Bateson and Munhall consider the speaker–listener dyad in more detail. They review results from behavioral and computational work to show that articulatory movement "simultaneously shapes the acoustic resonances of the speech signal and visibly deforms the face," that speech intelligibility increases if visual information about speech is provided, but that fairly low quality information is sufficient for the increase. From these results, Vatikiotis-Bateson and Munhall argue that we need to develop a better sense of the role of redundancy in production and perception, but offer the hypothesis that redundancies in the visual channel may facilitate the perceiver's spatial and temporal alignment to speech events by multiple means such as highlighting prosodic structure. Lucie Ménard (Chapter 10) also explores audio-visual processing, but from a developmental perspective and within an information-processing framework. She argues, based on work with sensory deprived individuals, that motor goals are multimodal – built up from experience with the acoustic and visible aspects of the signal.

1.2.3 Speech motor control

Motor goals are the principal focus of chapters in Part III. Pascal Perrier and Susanne Fuchs (Chapter 11) provide an extensive introduction to the concept of a goal with reference to motor equivalence, which they define as the "capacity of the motor system to adopt certain (different movement) strategies depending on external constraints." They also link motor equivalence to the concept of "plasticity" and to the workings of the central nervous system (CNS). Takayuki Ito (Chapter 12) and John Houde and Srikantan Nagarajan (Chapter 13) discuss the role of sensory feedback in speech motor control with Ito focused on somatosensory information and Houde and Nagarajan on auditory information. Both contributors review findings from feedback perturbation experiments, and both adopt a neuroscientific approach to explain these findings. Whereas Ito concentrates on contributions from the peripheral

nervous system, Houde and Nagarajan elaborate a CNS model of control based on internal auditory feedback (efferent copy) and an external feedback loop that allows for the correction of errors in prediction based on incoming sensory information. Gary Weismer and Jordan Green (Chapter 14) are also very focused on the CNS, but their objective is to understand whether the "execution" stage of speech production – classically thought to be disrupted in dysarthria – is truly separable from the "planning" stage of speech production. They conclude, based on clinical and experimental data from individuals with dysarthria and apraxia of speech (a "planning" disorder), that the anatomical and behavioral boundary between the two stages is poorly defined. Ben Maassen and Hayo Terband (Chapter 15) appear to confirm Weismer and Green's point regarding fuzzy boundaries by noting that childhood apraxia of speech, a developmental rather than acquired disorder, may be localized "at the level of phonetic planning, and/or motor programming, and/or motor execution, including internal and external self-monitoring systems." They also make the important point that whether the primary deficit is localized in planning or execution, this *motor* disorder has consequences for *language* representation.

1.2.4 Sequencing and planning

Sequencing and planning are at the interface of speech motor processes and the language representations we associate with meaning. It is here that we grapple most directly with Lashley's (1951) serial order problem as applied to speech. Peter MacNeilage (Chapter 16) addresses the problem within an evolutionary framework. He starts with an oral-motor function – chewing – that precedes speech in evolutionary time and hypothesizes that the movements associated with this function were exapted for speech: once coupled with phonation, the up-down jaw movements of chewing yield an amplitude modulated waveform reminiscent of the ones that linguistic systems segment and categorize as consonant–vowel sequences. In Chapter 17, I think about continuities and junctures in development and what these imply for the acquisition of prosodically related temporal patterns. I argue that the acquisition of temporal patterns is due both to the refinement of speech motor skills and the development of a plan, which is suggested to emerge at the transition from vocal play to concept-driven communication. Gary Dell and Gary Oppenheim (Chapter 18), Stefanie Shattuck-Hufnagel (Chapter 19), and Robin Lickley (Chapter 20) all assume a plan based on units that are more closely tied to the abstract representations postulated in most modern linguistic theories. Dell and Oppenheim make an explicit argument against Articulatory Phonology type representations, in favor of atemporal units. Their evidence comes from the finding that the speech errors of inner speech are less subject to the phonemic similarity effects found in the speech errors of overt speech. Shattuck-Hufnagel is less concerned with the specific identity of segment-sized units, and more interested in the macro-structure of the speech plan. She argues, following Lashley (1951), that planning is hierarchically organized and proposes that prosodic structures, from the intonational phrase to

the metrical foot, provide successive frames for planning and execution. In her view, we move from abstract linguistic representation to motor commands as we iterate through the prosodic hierarchy. Finally, Lickley (Chapter 20) considers what happens when there are disruptions at any level in the planning and execution process by describing different kinds of disfluencies and repair strategies. He argues that understanding these in typical speech is critical to being able to define and understand disfluencies that result from developmental or acquired disorders.

1.2.5 Language factors

Although many chapters in the Handbook provide evidence for the argument that speech contributes to our understanding of language, this does not contradict the importance of the more widely recognized contribution of language to our understanding of speech and its acquisition. The chapters in this final Part of the Handbook directly address this contribution. Didier Demolin (Chapter 21) makes a strong case for cross-language investigations of speech sound production. He argues that a mainstay of phonetic sciences for over 100 years, the International Phonetic Alphabet (IPA), is based on limited language data and so may improperly circumscribe the capabilities of the human speech production mechanism. Fieldwork studies on speech production provide us with a clearer sense of what is possible, allowing for better documentation and preservation of minority languages. An understanding of diversity and variation also informs theories of sound change. Like Demolin, Taehong Cho (Chapter 22) addresses cross-linguistic diversity. Cho reviews the literature on timing effects at the segmental and suprasegmental level to argue for a phonetic component to the grammar, noting that "fine-grained phonetic details suggest that none of the putative universal timing patterns can be accounted for in their entirety by physiological/biomechanic factors." Jan Edwards, Mary Beckman, and Ben Munson (Chapter 23) are interested in the effects of language-specific sound patterns and social meaning on speech production and phonological acquisition. They review findings from their παιδολογος project and other cross-linguistic research, demonstrating the importance of the social group in speech and language acquisition. They also show that cross-cultural variation in speech sound acquisition is best understood with reference to specific acoustic differences in how the "same" phoneme is produced in different languages and varieties. Finally, Lisa Goffman (Chapter 24) returns to our theme of the fuzzy divide between speech and language to investigate the effects of lexical, morphological, and syntactic structures on the acquisition of speech motor skills. She notes that "though there is little question that [a] more domain specific view is dominant in framing how research on language acquisition has been approached, there have long been powerful suggestions that motor and other factors also play a crucial role in how children approach the language learning task." The studies on speech kinematics in children that she reviews in her chapter indicate that the reverse is also true: language factors affect the acquisition of timing control and articulatory precision in children.

1.3 Conclusion

The *Handbook of Speech Production* is designed to provide the reader with a broad understanding of speech production. Leading international researchers have contributed chapters that review work in their particular area of expertise, outline important issues and theories of speech production, and detail those questions that require further investigation. The contributions bring together behavioral, clinical, computational, developmental, and neuropsychological perspectives on speech production with an emphasis on kinematics, control, and planning in production. The organization of the Handbook is from the most concrete aspects of speech production to its most abstract. Such an organization is designed to encourage careful reflection on the relationship between speech and language, but alternate pathways through the Handbook are always possible. The brief overview of content provided in this Introduction was meant to show how the chapters create a coherent whole, but it will hopefully also help you, the reader, design a personal pathway through the Handbook if that is your wish.

REFERENCES

Bernstein, Lynne E. and Gary Weismer. 2000. Basic science at the intersection of speech science and communication disorders. *Journal of Phonetics* 28: 225–232.

Browman, Catherine P. and Louis M. Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3: 219–252.

Chomsky, Noam. 1959. A review of B.F. Skinner's *Verbal Behavior*. *Language* 35: 26–58.

Keele, Steven W. 1968. Movement control in skilled motor performance. *Psychological Bulletin* 70: 387–403.

Lashley, Karl S. 1951. The problem of serial order in behavior. In L.A. Jeffress (ed.), *Cerebral Mechanisms in Behavior*, 112–131. New York: John Wiley & Sons, Inc.

Pinker, Steven and Paul Bloom. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13: 707–784.

Skinner, Burrhus F. 1957. *Verbal Behavior*. New York: Appleton-Century-Crofts.