THE ORGANIZATION OF SPEECH MOVEMENTS: SPECIFICATION OF UNITS AND MODES OF CONTROL

Björn Lindblom

Department of Linguistics, Stockholm University, Stockholm, Sweden
Department of Linguistics, University of Texas at Austin, Texas, USA
lindblom@ling.su.se

ABSTRACT

In the present paper we will address two issues. First, are the units of speech static targets or dynamically specified gestures? The available evidence favors the conclusion that dynamic signal attributes play an important role in speech perception, However, the perceptual significance of speech dynamics does not compel us to conclude that the input to the speech production system is dynamic (gestural). Parsimony would seem to dictate that dynamic motor commands be put on hold until physiological and biomechanical response characteristics of the speech production system are better understood. Second, are speech units articulatory or perceptual? This is a question that derives from much quoted programmatic statements by Stetson and Jakobson. Evidence was reviewed showing that speech movements, like non-speech actions, are adaptively organized and can be planned so as to facilitate the listener's task by enhancing the perceptual correlates of phonetic categories. However, there is also data indicating that these categories have a strong isomorphism with articulatory processes. The implication of this conclusion is that the question of where in the speech chain units are best defined may be a spurious one.

INTRODUCTION

Since sound spectrography and various instrumental techniques came to be widely used research on speech has richly documented the complex relation between the physical attributes of an utterance and its linguistic description. From the very start (Joos 1948, Hockett 1955) it was clear that the overarching challenge was to explain the strong context-dependence and variability of acoustic phonetic patterns. Major handbooks (Hardcastle & Laver 1997) and review chapters (Farnetani 1997, Farnetani & Recasens 1999, Kent, Adams & Turner 1996, Löfqvist 1997) converge in identifying coarticulation as a major contributor to the mismatch between the dynamic and the linguistic perspectives of speech. Hence the large number of studies on this topic (Hardcastle & Hewlett 1999). Experimental work was at first largely focused on 'laboratory speech' but recently interest has been directed also towards spontaneous unscripted speech (Kohler 2000) and a wider range of speaking styles and expressive modes (Fónagy 2001). Such developments have further underscored the magnitude of the task of making sense of the complexities of speech processes.

The present remarks will be focused on the planning of speech movements. In particular we will address the issue of what is controlled and what is a product of execution in the hope of getting somewhat closer to determining the true nature of the basic units of speech. We will examine various schools of thought and try to evaluate the different proposals by drawing selectively on the huge body of experimental data accumulated over the past fifty years.

THE NOTION OF 'STATIC TARGETS'

In 1963 Stevens & House published their seminal study of American English vowels. This is a landmark paper in that it provides the first sketch of a model speech production based on articulatory dynamics. The authors observed that the formant pattern of a given vowel tended to be systematically displaced in the direction of the frequency values of the consonant context. Those shifts were interpreted to indicate that articulatory movements tended to undershoot, that is fall short of, hypothetical vowel targets. This model was further explored by Lindblom (1963). The study demonstrated that undershoot effects in vowels could be predicted from information on vowel duration, consonantal environment and a vowel-specific target F-pattern independent of context.

Subsequent work (e.g., Gay 1978, van Son 1993) failed to replicate the duration-dependence effect reporting minimal undershoot at high speaking rates. However, such discrepancies turn out to be more apparent than real when additional observations and biomechanical constraints are taken into account. The kinematic data of Kuehn & Moll (1976), Nelson et al (1984) Flege (1988) show that subjects with limited undershoot tend to increase movement velocities. According biomechanical analyses (Nelson 1983) an interplay is to be expected between movement duration, extent and velocity (an index of articulatory effort). Consequently, to adequately capture all the facts reported in the literature, undershoot should be described by a model making it a function of all of those three factors.

Important contributions to the debate about targets include the 'window model' proposed by Keating (1990) which replaces positional targets by target ranges in order to better capture underspecification phenomena and coarticulation effects. This idea is also a feature of the DIVA model (Guenther 1995) which describes each phoneme in terms of a set of orosensory dimensions. These representations are established during a learning stage. They take the form of 'convex region targets' that specify the permissible ranges of variation for each individual dimension. With this revision of the traditional notion of target Guenther is able to give a unified and compelling account of a number of key phenomena reported in the literature: motor equivalence, the vowel dependence of place of articulation, speaking rate effects including undershoot as well as anticipatory and perseveratory coarticulation. The mapping of acoustic properties of phonetic segments onto orosensory dimensions is supported by empirical work showing that in the absence of auditory feedback speakers are able to maintain auditory goals (Perkell et al 1997).

A PARADOX

Sensory physiology indicates that perception is more sensitive to changing stimulus arrays than to static ones (Kandel & Schwarz 1981). Physiology textbooks tell us that sensory mechanisms have evolved to detect change. In speech perception such processes clearly play a role as shown by investigations of contrast effects in perceptual categorization (Lotto et al 1998).

In speech production steady-states are rare. Articulatory movement and spectral changes are pervasive. It is curious that the systems for phonetic specification (the IPA, distinctive feature systems (Jakobson et al 1952, Chomsky & Halle 1968) other frameworks (Ladefoged & Maddieson 1996, Maddieson 1984) – should favor static over dynamic attributes. When speech is transcribed phonetically, dimensions are used most of which refer to speech production steady-states, e.g., labial, close, voiced, stop, nasal etc. Similarly, in modeling phonatory and articulatory processes investigators have invoked static targets in combination with system response characteristics (Fujisaki 1983, Fujimura & Erickson 1997).

If perception likes change, why do systems for phonetic specification favor steady-state attributes?

DYNAMIC SPECIFICATION

The "dynamic specification" approach proposed by Winifred Strange responds to this paradox. It is based on a series of experiments (Strange 1989a; Strange 1989b) demonstrating that listeners are able to identify vowels with high accuracy although the center portions of CVC stimuli have been removed leaving only the first three and the last four periods of the vowel segment. In other words, vowel perception is possible also in syllables that lack information on the alleged 'target' but include an initial stop plosion and surrounding formant transitions – so-called 'silent-center' syllables. Strange takes her findings to imply that vowel identification in CVC sequences is based on more than just the information contained within a single spectral slice sampled near the midpoint "target" region of the acoustic vowel segment. Rather the relevant information is distributed across the entire vowel and it includes formant frequency time variations. According to the "dynamic specification" approach,

> *"… vowels are conceived of as characteristic gestures having intrinsic timing parameters (Fowler, 1980). These dynamic articulatory events give rise to an acoustic pattern in which the changing spectrotemporal configuration provides sufficient information for the unambiguous identification of the intended vowels." (Strange 1989a:2135-36).*

The term 'target' has been used with several different meanings. Strange's definition refers to an observable, the spectral cross-section sampled at a vowel's quasi-steady-state midpoint. It thus differs slightly from the meaning used above, *viz.,* target as a virtual (underlying, asymptote) phenomenon - a point in multi-dimensional articulatory control space that may, or may not, be realized articulatorily and acoustically.

THE STATUS OF 'PHONETIC GESTURES'

Proponents of the Motor Theory (Liberman & Mattingly 1985), Direct Realism (Fowler 1986) and Articulatory Phonology (Browman & Goldstein 1992) all endorse Stetson's famous dictum:

> *"Speech is rather a set of movements made audible than a set of sounds produced by movements"* (Stetson 1951).

It is in the spirit of this statement that the 'phonetic gesture' has been proposed as the primitive that phonetic theory needs to resolve its major issues. An important aspect is that this unit is dynamically specified. For an algorithmic definition of gesture the reference is Saltzman & Munhall (1989).

The gesturalist program addresses the paradox just described rejecting static and abstract units. A significant part of the argument is that, if perceptual mechanisms thrive on dynamics, the control units should reflect that fact and should themselves be dynamic.

Another point in favor of gestures is that, unlike abstract phonemes, segments and targets, they can be observed and measured. Munhall & Löfqvist (1992) reported an experiment designed to investigate the ‘blending’ of laryngeal gestures. Data were collected on the phrase “kiss Ted” spoken at different speaking rates so as to yield a broad distribution of [st] durations. The results indicate two separate gestures at slow rates and a merging of the two components into a single peak at the fastest rate. The authors demonstrated that all movement patterns including those for intermediate tempos could be generated from two invariant canonical shapes using the simple strategy of gestural summation.

A similar analysis is presented by Maddieson (1995) who investigated lip and tongue movements in doubly-articulated and singly-articulated stops, e.g., [aka], [apa] and [akp͡a] in Ewe. The study showed that it is possible to derive the lip and tongue movements for labial-velar stops from the component gestures taken from the singly-articulated stops.

The Swedish psychologist Gunnar Johansson developed a theory of perception which comes close to the ecological framework developed by Gibson (1979) and adopted as a major source of inspiration for Direct Realism work in speech. Johansson came up with the Dot Man, a motion picture illustration of the importance of time-varying stimulus information in perception. The film shows subjects walking, climbing stairs, riding a bike and dancing the Swedish hambo. Subjects have light spots placed on head and joints. The exposure is such that the screen is dark with only the light spots showing. Steady-state images look like stars in the sky but as soon as there is movement the viewer immediately perceives a human body engaged in some action. Although the information from the light spots provides only a sparse sampling of the body features, the motions make sense right away in a highly compelling way. There is definitely an effect of perceptual completion behind these impressions. The main point to be made here is that “the information is in the dynamics”.

In light of these observations: Are we making a mistake hanging on to target theories of speech production? Also we must ask: Are speech units articulatory rather than auditory/perceptual?

THE OUTPUT-ORIENTED NATURE OF SPEECH MOTOR CONTROL

Walking, running, reaching and manipulating things, we constantly face new conditions. We adapt to our changing environment effortlessly because motor mechanisms evolved to deal with unforeseen obstacles. Their default mode of operation is in fact to be compensatory.

This plasticity becomes evident when we consider the way we write. We can easily recognize a colleague’s familiar handwriting whether we find it on a piece of paper or on the blackboard. Different muscle groups are involved. The size of the letters differ. However their shapes are basically similar. The movements are not defined in terms of fixed set of muscles and their contraction patterns. Rather they are recruited in functionally defined groups (“coordinative structures”) and planned and executed in an external coordinate space (Lashley 1951).

In reaching tasks the hand usually moves along a more or less straight line. However, when the movement is described in a diagram with the angle of the elbow plotted against the angle of the shoulder, a curved contour is seen. These differences have been taken to suggest that reaching is planned in extrinsic rather than intrinsic coordinates (Löfqvist 1997).

Examples of output-oriented behavior abound in the speech literature. For instance, the recruitment of respiratory muscles depends on the volume of air in the lungs and is made in order to achieve a fairly constant subglottal pressure during an utterance (Perkell 1997). A second illustration: when speakers increase their vocal effort they typically open their jaws more widely (Schulman 1989). This increases the intensity of the vowels but is in conflict with the production of consonants which involve a higher jaw position and a vocal tract constriction. Schulman's movement records for VbVbV sequences show that the increased jaw opening for loud speech does apply also to consonantal actions but is actively compensated for by more extensive lower and upper lip movements. The purpose of the compensation is to guarantee the bilabial stop closure as well as a reasonably normal closure duration. Third example: In English vowels the velum tends to be higher for high than for low vowels. Perkell (1997) reports that these variations in velum height are positively correlated with pressure measurements: the pressure exerted between the velum and the pharynx wall is higher for high than for low vowels (Moon et 1994) This pattern makes sense when we consider the fact that, acoustically, nasal coupling is known to be spectrally more disruptive for high than for low vowels (House & Stevens 1956). Finally let us also mention the studies of motor equivalence (Perkell et al 1993, Perkell 1997), perturbations of ongoing speech (Abbs 1996, Abbs & Gracco 1984) and actions invoked to compensate for a bite-block (Gay et al 1981). All of these investigations converge on the same conclusion: speech movements are no different from other motor behaviors. They are organized to achieve externally defined goals and to do so in face of disturbances and variations in initial conditions.

THE QUEST FOR CONVERGENCE

The above account has so far painted a rather black-and-white picture of how the field has dealt with some of the major issues. On the one hand, we have the Stetsonian point of view which gives the articulatory level a privileged position. On the other there is Jakobson's position which assumes that the specification of speech sounds is perceptually based and that the stages of the speech chain form an

> *" … operational hierarchy of levels of decreasing pertinence: perceptual, aural, acoustical and articulatory (the latter carrying no direct information to the receiver)." (Jakobson et al 1952).*

Furthermore, we have contrasted the static and the dynamic definitions of phonetic primitives. Are the basic units articulatory or perceptual? Are they static or dynamic? Is the problem really that of making a choice between those alternatives? We will next review some facts indicating that we may need to rethink those options.

The first set of data comes from a perceptual study. Carlsson & Granström (1977) recorded Swedish CV syllables contrasting [l] and [n] in the environment of different vowels. (The Swedish [l] is clear). The acoustic boundary between the consonant and the vowel was identified in the waveforms and a short interval preceding and following it was defined. These micro-segments were then exchanged between [n] and [l] while keeping the vowel constant. The stimuli so processed were then presented to listeners for identification together with the original syllables. The results indicated that these subtle changes were enough to significantly modify the subject's percept from [l] to [n] and conversely. It was inferred from the findings that the dynamics of the closure termination was different in the two cases and contained powerful manner cues. This observation is compatible with what is known about the acoustics of [l] and [n]. The 'release' of [l] would be expected to have a more abrupt, crisper quality than that for [n] in which nasalization extends beyond the CV boundary into the vowel. Hence it appears justified to assume that these attributes were not part of the specification of the motor task but would fall out automatically from the articulation-to-acoustics mapping.

In a similar spirit, let us revisit the work by Öhman (1967). His quantitative account of the vowel-dependent tongue shape variations in [dV] syllables was based on a formula that postulated an invariant target contour for [d]. What does that finding imply about the production of [dV] syllables?

Would there also be an acoustic/auditory target pattern? Undoubtedly. A normal-sounding [dV] would be expected to exhibit a voiced occlusion followed by a release whose spectral shape would reflect both the place of articulation of the consonant and the influence of the following vowel. There would also be the formant transition cues for the dental place. We might reasonably assume that by maneuvering his tongue so as to make contact between the blade and the palate at the dental place of articulation, the talker should be able to produce many of the attributes associated with the expected acoustic pattern. The voiced closure would produce the typical voice bar and would cause the oral pressure build-up necessary for the dental burst. With a certain depth of contact a closure of appropriate duration would be produced.

The point made is that a simple spatial target command would go a long way towards generating the auditory target pattern for [dV]. Admittedly, these two examples may oversimplify the complexities of speech but they should nonetheless open our eyes to the possibility that static point targets in production can be perfectly compatible with temporally distributed, perceptually significant spectro-temporal information.

Note that the choice between Stetsonian and Jakobsonian phonetics does not arise in the above cases. Nor do we need to worry about taking sides whether units are static or dynamic. In the simple examples just given all sides are accommodated.

CLUES FROM NON-SPEECH MOVEMENTS

The above considerations raise the question of how much dynamics motor commands actually need to specify. Work on non-speech movement throws some preliminary light on that question.

Many speech production models include a level at which articulators are represented as spring-mass systems, that is, devices with mass, damping and elasticity. In such mechanisms movement is produced by applying a force of appropriate magnitude to an articulator so as to push it towards an intended goal. However, the coding of articulatory goals in terms of force values turns out not to be a viable approach (Perrier et al. 1996). This is so because the force needed to move an articulator towards a certain goal, by necessity depends on the distance between where the movement starts from and where that goal is. Hence, like kinematically defined gestures (but unlike positional targets), input force is not context-free.

Accordingly it becomes necessary to look for a control level that is located further upstream than the generation of muscle forces. In the literature on the physiology of movement we encounter the Equilibrium-Point (EP) hypothesis also known as the lambda model. It was originally proposed in connection with studies of arm and eye movement and has been applied extensively also to speech notably by David Ostry and his colleagues. The paper by Perrier et al

(1996) provides a helpful overview and can be read in conjunction with the commentaries by other speech physiologists (Abry et al 1998).

The EP hypothesis was developed from the observation that muscles tend to behave like springs. An activated muscle that is stretched produces a restoring force that varies with the degree of stretch. If a limb is moved by some external action, it will return to its original position when the perturbation is removed. Its return path and speed of response are determined by its stiffness. The EP hypothesis assumes that voluntary movements result from shifting the equilibrium-point of the motor system. The input to the model is the threshold length of a given muscle. When the muscle length reaches the desired value, muscle activation, and thus force generation, cease. The point in space at which force is zero defines an equilibrium-point.

The EP hypothesis provides a parsimonious account of how the central nervous system plans and executes movements. EP control bypasses the problem of “inverse dynamics”. It suggests that the brain finds a unique and smooth trajectory by letting that path be automatically determined by the dynamics of the physiological system. The parsimony lies in the fact that neural control levels do not need an explicit representation of limb dynamics. Nor is any preplanning required in the case of compensatory tasks since viscoelastic forces provide immediate corrective action in response to an external perturbation.

There are competing answers to why movements take one specific path rather than others that are in principle possible. A central idea of those accounts is that uniqueness and kinematic characteristics are products of optimization. An often-mentioned fact about of natural motions is that the brain seems to have a preference for the ‘smoothest’ option. It has been suggested that motions are smooth because control processes are guided by criteria related to minimal energy consumption (Nelson 1983). An empirically successful and widely embraced variant of this thinking is the ‘minimum jerk’ model (Hogan & Flash 1987) which finds the least ‘jerky’ pathway by optimizing the acceleration pattern. A third criterion, also in this category, is ‘minimum muscle force change’ (Soechting & Flanders 1998). A recent proposal takes a different route (Harris & Wolpert 1998). It argues that the dimension optimized is ‘precision’, not ‘efficiency’ as in the three studies just mentioned. This work starts from the observation that neural signals are inherently noisy. Because of the noise the same command does not guarantee reaching the same destination. The noise is activity-dependent: the stronger the command the larger the variability. The ‘minimum variance’ model of Harris and Wolpert optimizes the precision of endpoint control and in so doing generates motions exhibiting both smoothness and the kinematic properties closely matching those found in natural movements.

How much dynamics do motor commands actually need to specify? The evidence reviewed appears to converge on the following points. 1.Target control. First, all accounts appear to assume that the nervous system has information on the current location of the limb and a specification of where it has to go. Both EP and non-EP models postulate positional targets. 2. Determinants of trajectory shape: stiffness tuning and optimization. Another shared theme is that, once information on ‘here’ and ‘there’ is available, physiological constraints determine the shape of the path between those points. According to the EP hypothesis, movement paths are determined by the stiffness of the activated muscles. Since stiffness can be controlled separately from muscle lengths there exists a possibility of varying the time course of the trajectory. In non-EP approaches trajectories are derived by means of various forms of optimization. While the choice of optimality criterion may be a matter of lively debate (Kawato 1999), that debate seems to be, not about optimization per se, but about the parameter(s) to be optimized.

Needless to say the above remarks do not resolve the issue of static targets vs. dynamic gestures in phonetics. However, it does provide some preliminary clues: It shows that the notion of ‘target’ as an object of control is pervasively paralleled in recent neuro-computational models of movement. It also suggests that, if the production of ‘phonetic gestures’ is assumed to be analogous to trajectory formation in point-to-point arm movements, they are better seen as interpolation phenomena arising as by-products of efficiency- and/or precision-optimized execution rather than as explicitly specified at the input level of speech production control.

SUMMARY

In the present review we addressed two issues. First, are the units of speech static targets or dynamically specified gestures? The available evidence favors the conclusion that dynamic signal attributes play an important role in speech perception, However, the perceptual significance of speech dynamics does not compel us to conclude that the input to the speech production system is dynamic (gestural). Parsimony would seem to dictate that dynamic motor commands be put on hold until physiological and biomechanical response characteristics of the speech production system are better understood.

Second, are speech units articulatory or perceptual? This is a question that derives from much quoted programmatic statements by Stetson and Jakobson. Evidence was reviewed showing that speech movements, like non-speech actions, are adaptively organized and can be planned so as to facilitate the listener’s task by enhancing the perceptual correlates of phonetic categories. However, there is also data indicating that these categories have a strong isomorphism with articulatory processes. The implication of this conclusion is that the question of where in the speech chain units are best defined may be a spurious one.

REFERENCES

Abbs J H (1996): “Mechanisms of speech motor execution and control”, pp 93-111 in Lass N J (1996): *Principles of experimental phonetics*, Mosby:St Louis, Missouri.

Abbs J H & Gracco V L (1984): “Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech”, *J of Neurophysiology 51*:705-723.

Abry C & Badin P (1998): *Bulletin de la Communication Parlée* 4:5-110, Special issue on the Equilibrium Point Hypothesis,

Browman C P & Goldstein L (1992): “Articulatory phonology: An overview”, *Phonetica 49*:155-180.

Carlsson R & Granström B (1977): *Perception and synthesis of speech,* doctoral dissertation, KTH, Stockholm, Sweden.

Davis H (1954): “Twenty-Five Acoustical Years of Speech and Hearing”, *J Acoust Soc Am 26(5)*:607-611.

Diehl R L (1986): “Coproduction and direct perception of phonetic segments: A critique”, *J of Phonetics 14(1)*:61-66.

Fant G (1960): *Acoustic Theory of Speech Production*, Mouton:The Hague.

Farnetani E (1997): “Coarticulation and connected speech processes”, pp 371-404 in Hardcastle W J & Laver J (eds): *The Handbook of Phonetic Sciences*, Blackwell:Oxford.

Farnetani E & Recasens D (1999): “Coarticulation models in recent speech production theories”, pp 31-65 in Hardcastle W J & Hewlett N (eds): *Coarticulation, theory, data and techniques*, CUP:Cambridge.

Folkins J W & Abbs J H (1975): “Lip and jaw motor control during speech: Responses to resistive loading of the jaw”, *J Speech Hearing Res 18*:207-220.

Fónagy I (2001): *Languages within Language*, John Benjamins:Amsterdam.

Fowler C A (1980): “Coarticulation and theories of extrinsic timing”, *J of Phonetics 8*:113-133.

Fowler C A (1986): “An event approach to the study of speech perception from a direct-realist perspective”, *J of Phonetics 14(1)*:3-28.

Fujimura O & Erickson D (1997): “Acoustic phonetics”. pp 65-115 in Hardcastle W J & Laver J (1997): *The Handbook of Phonetic Sciences*, Blackwell:Oxford.

Fujisaki H (1983): “Dynamic characteristics of voice fundamental frequency in speech and singing”, pp 39-55 in MacNeilage P F (1983): *The production of speech,* New York: Springer-Verlag.

Fukson O I, Berkinblit A G & Feldman A G (1980): “The spinal frog takes into account the scheme of its body during the wiping reflex”, *Science 209*:1261-1263.

Gay T (1978): “Effect of speaking rate on vowel formant movements.” *J Acoust Soc Am 63(1)*: 223-230.

Gay T, Lindblom B & Lubker J (1981): “Production of bite-block vowels: Acoustic equivalence by selective compensation”, *J Acoust Soc Am 69(3)*:802-810.

Gibson J J (1979): *The ecological approach to visual perception*, Houghton Mifflin:Boston, MA.

Granit R (1979): *The purposive brain*, MIT Press:Cambridge, MA.

Guenther F H (1995): “Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production”, *Psychological Review 102*:594-621.

Hardcastle W J & Laver J (1997): *The Handbook of Phonetic Sciences*, Blackwell:Oxford.

Hardcastle W J & Hewlett N (1999): *Coarticulation, theory, data and techniques*, CUP:Cambridge.

Hardcastle W J & Marchal A (1990): *Speech Production and Speech Modeling*, Dordrecht:Kluwer.

Harris C M & Wolpert D M (1998): “Signal-dependent noise determines motor planning.” *Nature 394*: 780-784.

Henke W L (1966): *Dynamic articulatory model of speech production using computer simulation*, doctoral dissertation, MIT, Cambridge, MA.

Hockett M (1955): *A manual of phonology*, Indiana University Publications in Anthropology and Linguistics XI, Bloomington.

Hogan N & T Flash (1987): “Moving gracefully: quantitative theories of motor coordination.” *Trends Neurosci 10*: 170-174.

Houde J F & Jordan M I (1998): “Sensorimotor adaptation in speech production”, *Science 279*:1213-1216.

Houde J F & Jordan M I (1998): “Sensorimotor adaptation in speech production”, *Science 279*:1213-1216.

House A S & Stevens K N (1956): “Analog studies of the nasalization of consonants”, *JSHD 21*:218-232.

Jakobson R, Fant G & Halle M (1952): *Preliminaries to Speech Analysis*, Acoustics Laboratory, M.I.T., Techn Report No.13.

Joos M (1948): “Acoustic phonetics”, *Language* 24:2, supplement.

Kaburagi T Honda M (1996): “A model of articulator trajectory formation based on the motor tasks o vocal-tract shapes”,*J Acoust Soc Am 99(5)*:3154-3170.

Kandel E R & Schwarz J H (1981): *Principles of neural science*, Edward Arnold:London.

Kawato M (1999): “Internal models for motor control and trajectory planning”, *Current Opinion in Neurobiology 9*:718-727.

Keating P A (1990): “The window model of coarticulation: Articulatory evidence”, pp 451-470 in Kingston J & Beckman M E (eds), *Papers in laboratory phonology I: Between the grammar and physics of speech*, CUP:Cambridge.

Kelso J A S, Saltzman E L & Tuller B (1986): “The dynamical perspective on speech production: Data and theory”, *J of Phonetics 14(1)*:29-59.

Kent R D, Adams S G & Turner G S (1996): “Models of speech production”, pp 3-45 in Lass N J (1996): *Principles of experimental phonetics*, Mosby:St Louis, Missouri.

Kohler K J (2000): “Investigating unscripted speech: Implications for phonetics and phonology”, *Phonetica 57*:85-94.

Kozhevnikov V A & Chistovich L A (1965): "Speech Articulation and Perception", US Dept. of Commerce Publ. JPRS:30 543;TT: 65-31233 (translated from Russian).

Kröger B J (1993): “A gestural production model and its application to reduction in German”, *Phonetica 50*:213-233.

Kühnert B & Nolan F (1999): “The origin of coarticulation”, pp 7-30 in Hardcastle W J & Hewlett N (eds): *Coarticulation, theory, data and techniques*, CUP:Cambridge.

Lashley K (1951): “The problem of serial order in behavior”, pp 112-136 in Jeffress L A (ed): *Cerebral mechanisms in behavior*, Wiley:New York.

Ladefoged P & Maddieson I (1996): *The sounds of the world’s languages*, Oxford:Blackwell.

Lass N J (1996): *Principles of experimental phonetics*, Mosby:St Louis, Missouri.

Levelt W J M (1989). *Speaking: From intention to articulation*. MIT Press: Cambridge, MA:

Liberman A & Mattingly I (1985): "The motor theory of speech perception revised," *Cognition* 21:1-36.

Lindblom B (1963): "Spectrographic study of vowel reduction", *J Acoust Soc Am 35*:1773-1781.

Lindblom B, Lubker J & Gay T (1979): ”Formant frequencies of some fixed-mandible vowels and a model of motor programming by predictive simulation,” *J Phonetics 7*:147-161.

Lindblom B (1990): "Explaining phonetic variation: A sketch of the H&H theory", 403-439 in Hardcastle W J & Marchal A (eds): *Speech Production and Speech Modeling*, Dordrecht:Kluwer.

Löfqvist A (1990): “Speech as audible gestures”, 289-322 in Hardcastle W J & Marchal A (eds): *Speech Production and Speech Modeling*, Dordrecht:Kluwer.

Löfqvist A (1997): "Theories and models of speech production", pp 404-426 in Hardcastle W J & Laver J (eds): *The Handbook of Phonetic Sciences*, Blackwell:Oxford.

Lotto A J, Kluender K R & Holt L L (1997): "Perceptual compensation for coarticulation by Japanese quail", *J Acoust Soc Am 102*:1134-1140.

MacNeilage P F (1970): "Motor control of serial ordering of speech", *Psychological Review 77*:182-196.

MacNeilage P F (1983): *The production of speech,* New York: Springer-Verlag.

Maddieson I (1984): *Patterns of sound*, CUP:Cambridge.

Maddieson I (1995): "Gestural economy", pp. 574-577 in Elenius K & Branderud P (eds): *Proceedings ICPhS 95 Stockholm*, vol 4.

Mauk C (2003): *Undershoot in two modalities: Evidence from fast speech and fast signing*, Ph D dissertation, Dept of Linguistics, University of Texas at Austin.

Moon J B, Kuehn D P & Huisman J J (1994): "Measurement of velopharyngeal closure force during vowel production", *Natural Center for Voice and Speech SPR 6*:53-60.

Morrish E C E (1988): "Compensatory articulation in a subject with total glossectomy", *British J of Disorders of Communication 23*:13-22.

Munhall K & Löfqvist A (1992): "Gestural aggregation in speech: laryngeal gestures," *J Phonetics 20*:111-126.

Munhall K G Vatikiotis-Bateson E & Kawato M (2000): "Coarticulation and physical models of speech production", in Broe M B & Pierrehumbert J (eds): *Papers in Laboratory Phonology V*. CUP:Cambridge.

Nelson, W. L. (1983). "Physical principles for economies of skilled movements." *Biol Cybern 46*: 135-147.

Nelson, W. L., J. S. Perkell, et al. (1984). "Mandible movements during increasingly rapid articulations of single syllables: Preliminary observations." *J Acoust Soc Am 75(3)*: 945-951.

Nooteboom S G & Eefting W (1992): "To what extent is speech production controlled by speech perception?", 439-449 in Tohkura Y, Vatikiotis-Bateson E & Sagisaka Y (eds): *Speech perception, production and linguistic structure*, IOS Press:Amsterdam.

Ohala J J (1986): "Against the direct realist view of speech perception", *J of Phonetics 14(1)*:75-82.

Öhman S E G (1967): "Numerical model of coarticulation", *J Acoust Soc Am 41*:310-320.

Perkell J, Matthies M, Svirsky M A, & Jordan M P (1997): "Trading relations between tongue-body and lip rounding in production of the vowel /u/: A pilot "motor equivalence" study", *J Acoust Soc Am 93(5)*:2948-2961.

Perkell J S (1997): "Articulatory processes", pp 333-370 in Hardcastle W J & Laver J (eds): *The Handbook of Phonetic Sciences*, Blackwell:Oxford.

Perkell J, Matthies M, Lane H, Guenther F, Wilhelms-Tricarico R, Wozniak J & Guiod P (1997): "Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models", *Speech Communication 22(2-3)*:89-302.

Perrier P, Ostry D J & Laboissière R (1996): "The equilibrium point hypothesis and its application to speech motor control." *J Speech & Hearing Res 39*: 365-378

Potter R K Kopp A G & Green H C (1947): *Visible Speech*, New York.

Recasens D, Palarès D M & Fontdevila J (1997): "A model of lingual coarticulation based on articulatory constraints" *J Acoust Soc Am 85(1)*: 295-312.

Saltzman E L & Munhall K G (1989): "A dynamical approach to gestural patterning in speech production", *Ecological Psychology 1*:333-382.

Schulman R (1989): "Articulatory dynamics of loud and normal speech" *J Acoust Soc Am 85(1)*: 295-312.

Sherrington C S (1986): *Man on his nature*, MacMillan:London.

Soechting, J F & Flanders M (1998): "Movement planning: kinematics, dynamics, both or neither?", pp 332- 349 in. Harris L R & Jenkin M (eds): *Vision and Action*, CUP:Cambridge.

Stetson R H (1951): *Motor Phonetics: A Study of Movements in Action,* North Holland:Amsterdam.

Stevens K N & House A S (1963): "Perturbation of vowel articulations by consonantal context. An acoustical study", *J Speech Hearing Res 6*:111-128

Stevens K N & House A S (1998): *Acoustic Phonetics*, MIT Press:Cambridge, MA.

Stevens K N (1972): "The quantal nature of speech: Evidence from articulatory-acoustic data," pp. 51-66 in Denes P B & David Jr E E, (eds): *Human Communication: A Unified View*, , New York: McGraw Hill.

Stevens K N (1989): "On the quantal nature of speech," *J Phonetics 17*:3-46.

Strange W (1989a): "Dynamic specification of coarticulated vowels spoken in sentence context." *J Acoust Soc Am 85(5)*: 2135-2153.

Strange W (1989b). "Evolving theories of vowel perception." *J Acoust Soc Am 85(5)*: 2081-2087.

Studdert-Kennedy M (2003): "How did language go discrete?" in Tallerman M (ed): *Evolutionary Prerequisites of Language*, Oxford, Oxford University Press.

Tremblay S, Shiller D M & Ostry D J (2003): "Somatosensory basis of speech production", *Science 423*:866-869.

van Son R J J H (1993): *Spectro-temporal features of vowel segments*, Univ. of Amsterdam: 195 "Theories and models of speech production", pp 404-426 in Hardcastle W J & Laver J (eds): The Handbook of Phonetic Sciences, Blackwell:Oxford.