1028

Proceedings of The XIIIth International Congress of Phonetic Sciences

ICPhS 95

Stockholm, Sweden
13 - 19 August, 1995

Editors
Kjell Elenius & Peter Branderud

Volume 2 of 4

The XIIIth International Congress of Phonetic Sciences (ICPhS 95) has been organised by the Department of Speech Communication and Music Acoustics, KTH (Royal Institute of Technology), and the Department of Linguistics, Stockholm University.

INTERGESTURAL TIMING IN SPEECH PRODUCTION: DATA AND MODELING

Elliot Saltzman
Haskins Laboratories, 270 Crown Street, New Haven, CT USA; Center for the Ecological Study of Perception and Action, University of Connecticut; Storrs, CT USA

ABSTRACT

The task-dynamic model of speech production is reviewed, and current developments that incorporate a recurrent network dynamics of intergestural timing are described. Four sets of data are considered that provide a set of temporal patterning benchmarks for the evaluation of any model of intergestural dynamics. These benchmarks are being used to evaluate and to guide the development of the extended task-dynamic model.

INTRODUCTION

This paper examines the nature of intergestural timing in speech production from both a theoretical and an empirical perspective. The theoretical focus is on how a model developed previously in our laboratories, the task-dynamic model, is being extended to incorporate a dynamics of intergestural timing. In the extended model, a recurrent, connectionist network is used to simulate these dynamics. This network plays the role of a utterance-specific central timing network for speech that is bidirectionally coupled to a simplified set of speech articulators. The outputs of the network serve to parameterize the dynamics of the articulators, whose ongoing state is fed back to the network in a manner that modulates its timing. As this work is in its preliminary stages (Saltzman, Mitra, Levy, & Hogden, in progress), discussion will be limited to the directions being taken in these developments.

The empirical focus of this paper is a review of data from four classes of intergestural timing phenomena that provide a significant set of constraints that must be satisfied by this or any model of speech intergestural timing. These data thus serve to establish a set of benchmark criteria in the temporal domain, according to which different models can be evaluated.

THE TASK-DYNAMIC MODEL

Work on the task-dynamic model of speech production has been conducted in collaboration with several of our colleagues at Haskins Laboratories as part of an ongoing project focused on the development of a gesturally-based, computational model of linguistic structures (e.g., [1], [2], [3], [4], [5], [6]). The focus of this model is on the control schemes that underlie gestural patterning, and represents an attempt to reconcile the linguistic hypothesis that speech involves an underlying sequencing of abstract, context-independent units, with the empirical observation of context-dependent interleaving of articulatory movements.

The central thesis of the model is that the spatiotemporal patterns of speech are shaped by a dynamical system with two functionally distinct but interacting levels. The *interarticulator* level is defined according to both *model articulator* (e.g., lips and jaw) and *tract-variable* (e.g., lip aperture and protrusion) coordinates; the *intergestural* level is defined according to a set of *activation* coordinates. Invariant gestural units are posited in the form of context-independent sets of dynamical parameters (e.g., lip protrusion target, stiffness, and damping coefficients), and are associated with corresponding subsets of these coordinates. Each unit's activation coordinate reflects the strength with which the associated gesture "attempts" to shape vocal tract movements at any given point in time. The tract-variable and model articulator coordinates of each unit specify, respectively, the particular vocal-tract constriction (e.g., bilabial) and articulatory synergy whose behaviors are affected directly by the associated unit's activation. The formation and release of such constrictions are governed according to an overall control regime defined as a damped, second-order dynamical system that spans tract-variable and model articulator coordinates, and whose parameters are functions of the current set of gestural activations.

The interarticulator level accounts for the coordination among articulators at a given point in time due to the currently active gesture set. At present, the dynamics of this level are sufficiently well developed to offer promising accounts of spatially goal-directed, multiarticulator movement patterns observed during unperturbed and mechanically perturbed gestures, and during periods of coproduction. In all cases, the evolving configuration of the articulators during a given utterance results from the gesturally- and posturally-specific way that driving influences generated in the tract-variable space are distributed across the associated articulatory synergies.

Intergestural Timing and Sequential Networks

The intergestural level accounts for the relative timing of activation intervals for the gestural units participating in a given utterance, e.g., for vocalic and bilabial gestures in a vowel-bilabial-vowel sequence. At present, this relative timing is accomplished with reference to a *gestural score* that specifies the activation of gestural units over time across parallel tract-variable output channels. Currently, gestural scores are not derived from an underlying implicit dynamics. Rather, they are derived explicitly either "by hand" or according to the rules of Browman and Goldstein's *articulatory phonology* ([1], [2], [3]).

We have adopted the recurrent, sequential network architecture of Jordan ([7], [8], [9], [10]; see also [11], [12], [13], [14]) at the intergestural level, as a means of patterning the gestural activation trajectories for the task-dynamic model. In the network, a separate output unit exists for each distinct gesture in a sequence, with the values of the output units corresponding to the activation values of the associated gestures. The network additionally includes a set of *state* units that determine among themselves a dynamical flow with an intrinsic time scale specific to the intended sequence. The dynamics of activity among the state units are defined by weighted recurrent connections: a) among the state units themselves, and b) from the output units to the state units. A set of intermediate *hidden* units are connected to both the state and output units by two respective layers of weighted paths. Temporal ordering among the output units is an implicit consequence of the network architecture and the sequence-specific set of weight values associated with each connection path.

Using *back-propagation* training methods (e.g., [15]), sequential networks can learn *single* sets of weights that will allow them to produce *several* different sequences. The information that is used to define the *teaching vector* sequences applied to the network during training will be obtained from the same articulatory phonology rules that are presently used to generate gestural scores. Such information will include gestural targets stiffness and damping coefficients, and the required times or phases of target attainment. Each simulated sequence is produced (and learned) in the presence of a corresponding constant activation pattern in a set of *plan* units. These units provide a second, external set of inputs to the network's hidden layer, in addition to the internal inputs provided by the state units.

A simplified pilot network

Pilot investigation now under way is focused on the behavior of a simplified model system. This system consists of a sequential network whose outputs represent the activations of gestures that are defined in a small set of "tract variables," whose dynamics are first order. The activation nodes act only to insert their associated target values into the tract-variable dynamics; the tract-variable stiffness coefficients are fixed. There are no model articulators at this point, equivalent to the simplifying assumption that tract variables and model articulators are defined in a one-to-one manner. These simplifications were adopted in order to gain familiarity and to sharpen intuitions regarding the types of intergestural timing phenomena that are intrinsic to such coupled dynamical systems, while keeping the "controlled" articulatory system as simple as possible.

FOUR BENCHMARK CLASSES OF TEMPORAL PHENOMENA IN SPEECH

This section focuses on four empirical classes of phenomena related to intergestural temporal patterns: 1) the temporal extent of anticipatory coarticulation; 2) the durational changes that result when mechanical perturbations are introduced to the articulators during speaking; 3) the effects of linguistic boundaries on the relative timing of gestures associated with the same phoneme; 4) the continuous sliding and aggregation effects, and discontinuous phase transition effects, that accompany increases of speaking rate.

Emphasis is placed on the first two classes of phenomena, since they are being addressed currently in our pilot modeling efforts. The remaining two classes will be reviewed more briefly, since they are not being addressed presently in our model, although they will provide significant constraints in guiding its future development.

Temporal Extents of Anticipatory Coarticulatory Fields

The definition adopted in this paper of a gesture's anticipatory field of coarticulation is the time from gesture onset to the time of target attainment. In a review of the literature on anticipatory lip rounding (e.g., [17], [18], [19]) and velum lowering (e.g., [20]), and on transconsonantal vowel-to-vowel coarticulation (e.g., [21]), Fowler and Saltzman [22] concluded that anticipatory coarticulation fields are temporally constrained, and that they do not typically extend very far backward in time from the time of target attainment. This interpretation is consistent with that provided by Bell-Berti & Harris' [23] frame model of coarticulation, and contrasts with the extensive degrees of anticipation that are possible in *look-ahead* models (e.g., [24]; in fairness, however, it should be noted that the anticipatory feature-spreading used in Henke's [24] model looked ahead only to the immediately following segment, although unlimited anticipation was allowed in principle).

In the task-dynamic model, this definition corresponds (roughly) to the interval from the time of gestural activation onset to the time of target attainment. Currently, the activation trajectories are specified in the gestural score as simple step functions, whose values change discretely between zero (the gesture is inactive) and one (the gesture is maximally active). Consequently, a gesture's anticipatory field is simply proportional to the gesture's intrinsic time constant, which is itself a function of the gesture's fixed stiffness and damping parameters. The onset of a given activation wave is specified according to the rules of articulatory phonology so that the gesture attains its target at the appropriate point in the simulated utterance (see also [16]).

Anticipatory behavior in the pilot network

In the pilot network, each first-order tract-variable is represented by a linear unit whose inputs are current target value and tract-variable position. The unit's output is current tract-variable velocity, which is fed into a linear, self-recurrent unit that provides a discrete time, Euler integration to generate the next tract-variable position. Taken together, therefore, the tract-variable and integrator units represent a model of the *forward dynamics* from current tract-variable state and target inputs to the next tract-variable state. Additionally, a delay line from the integrator unit is used to feed back current tract-variable state to the hidden units of the sequential network (see [8], [10], [25], [26] for related treatments).

Teaching vectors are applied intermittently at the tract-variable integrator units, and output errors are measured. At all other times, "don't care" conditions exist and no errors are defined. When errors are defined, however, they are propagated backward through the fixed tract-variable forward dynamics, and applied to the sequential net's output units. From this point on, these backpropagated errors are used in the usual manner to train the weights inside the sequential net. Because of the downstream recurrence implicit in the tract-variable dynamics, and the delayed tract-variable feedback into the sequential network, the *back-propagation through time* (e.g., [15]) training procedure is used.

It is known from previous work [7] that sequential nets whose outputs directly represent the distinctive features of speech will display relatively unconstrained temporal fields of anticipatory behavior. Hence, we hypothesize that "frame-model-like" behavior will require the addition of appropriate sets of *side constraints*, perhaps to the sequential net's output units, during the network's training phase (see [8] for detailed discussion of such constrained optimization methods). Under such training conditions, the anticipatory fields of activation waves should be limited to durations proportional to the tract-variable time constants of the activated gestures.

Effects of Mechanical Perturbations: Phase Resetting

Transient mechanical perturbations delivered to the speech articulators during repetitive and discrete speech sequences can alter the underlying timing structure of the ongoing sequence, and induce systematic shifts in the timing of subsequent movement elements ([27], [28], [29], [30], [31], [32], [33]). We have recently used such methods to examine the sequential dynamics governing bilabial and laryngeal gestures, both within segments and between successive syllables, in repetitive [...pæpæ...] and discrete [pə'sæpæpl] utterances ([34], [35], [36]).

Analyses of the repetitive utterances indicated that steady-state shifts occurred for both lips and larynx, but that no steady-state shifts occurred in the relative phasing of these articulators. These effects occurred only when the perturbation was delivered within a "sensitive phase" of the cycle. During this period, the downwardly directed lower lip perturbation opposed the just-initiated, actively controlled bilabial closing gesture for /p/. Thus, the sensitive period corresponded (roughly) to the acceleration portion of the closing gesture (Kawato, personal communication). When perturbations were delivered at the system's sensitive phase, the bilabial and laryngeal gestures were phase-advanced as a relatively coherent unit, maintaining their relative phasing as they were advanced in absolute time. Durational changes observed in other phases were systematic yet transient peripheral responses to the perturbation, and did not indicate central phase-resetting. Finally, the patterns observed in the discrete utterances resembled the transient effects seen in the repetitive utterances, indicating that similar dynamics underlie the production of both types of utterance.

Phase resetting: Implications for the pilot network

The fact that steady-state phase shifts exist supports the hypothesis that central intergestural dynamics can be "permanently" reset by peripheral articulatory events. Thus, any hypothesized central timing network for speech cannot not unidirectionally drive the articulatory periphery; rather, central and peripheral dynamics must be coupled bidirectionally, so that feedback information from the biomechanical periphery also can influence the state of the central "clock." In this regard, it is interesting to recall that feedback connections from the interarticulator level (tract-variable state) to the intergestural level (sequential network hidden units) were required in designing the pilot network used to produce unperturbed gestural sequences (see above section *Anticipatory behavior in the pilot network*).

It is unlikely, however, that this architecture will be sufficient to simulate the phase-resetting results. One reason is that such resetting occurred only during the syllable's sensitive phase, when the lip-opening perturbations opposed the actions of the just-initiated bilabial closing gesture. Additionally, although changes in syllable duration were found for other perturbed phases, these changes were simply transient effects, and did not indicate resetting of the central "clock". If one hypothesizes that efferent commands to the periphery are strongest near gestural initiation, then this pattern of results implies that the timecourse of afferent sensitivity mirrors that of efferent strength. In turn, such an hypothesis suggests a modeling possibility that we will investigate, namely, that "sensory" information regarding a perturbed tract-variable trajectory, perhaps in the form of a tract-variable error signal (i.e., current target – current tract-variable position), might be multiplicatively gated into the sequential network as a function of the activation state of gestures associated with the tract-variable.

Effects of Linguistic Boundaries

Many so-called phonemes are produced as coordinated *constellations* [1] of gestures, with characteristically different patterns of intergestural phasing depending on the constellation's location relative to a linguistic (e.g., syllable) boundary ([3], [37], [38], [39], [40]). For example, /l/ is produced with two tongue gestures — tongue tip raising and tongue body retraction. In certain dialects of English, word initial /l/'s ("light" /l/'s) are produced using roughly synchronous gestures; in word final /l/'s ("dark" /l/'s), the gestures are produced asynchronously, with gestural onset for the tongue tip aligned roughly with time of target attainment for the tongue body. Similary, the nasal consonant /m/ is produced word-initially with roughly synchronous velic lowering and bilabial closing gestures; in word final position, these gestures are asynchronous, with gestural onset for velic lowering aligned roughly with the bilabial closing gesture's time of target attainment.

Linguistic boundaries: Constraints on modeling

It is possible that these timing differences are incorporated explicitly into the phonological rules that, in our modeling scheme, are used to genererate the set of tract-variable-target teaching vectors for a given sequence. It is also possible, however, that the boundaries themselves are represented as non-tract-variable elements in the teaching vectors. For example, one possibility is that such boundary elements might serve to dampen the activations (i.e., drive the output units of the sequential network toward zero) of all the gestures in the simulated sequence in proportion to the strength (e.g., [41], [40]) of the associated boundary. Such a boundary element would serve to reduce the magnitudes of tract-variable gestures within the boundary "gesture's" temporal field. If the anticipatory field is much larger than the carryover field, this could produce pre-boundary gestural reduction similar to that reported in the literature. Whether such boundary elements could also alter the relative timing of nearby gestures is an intriguing possibility that will be explored in future developments of our model.

Effects of Increased Speaking Rate

A striking phenomenon that accompanies increases in speaking rate is that the gestures associated with temporally adjacent phonemes tend to "slide" relatively continuously into one another with a resultant increase in temporal overlap (e.g., [42], [3], [43], [44], [45], [46], [47], [48]). For example, Hardcastle [43] has shown with electropalatographic data that the tongue gestures associated with producing the (British English) consonant sequence /kl/ tend to slide into one another with experimentally manipulated increases in speaking rate.

Continuous increases in speaking rate can also produce discontinuous transitions of intergestural phasing ([49], [50], [51]). In these studies, when subjects spoke the syllable /pi/ repetitively at increasing rates, the relative phasing of the bilabial and laryngeal gestures associated with the /p/ did not change from the pattern observed at a self-elected, comfortable rate. However, when the repeated syllable /ip/ was similarly increased in rate, its relative phasing pattern switched relatively abruptly at a critical speed, from that observed for a self-elected, comfortable rate to the pattern observed for the /pi/ sequences.

Speaking rate: Constraints on modeling

At this point, we can only speculate as to the dynamical underpinnings of these continuous and discontinuous changes as a function of speaking rate. For example, it is possible that the continuous patterns of intergestural sliding might result from relatively simple changes in the values of a control parameter or parameter set at the intergestural level; e.g., the increased parallelism might be simulated by having rate increases serve to decrease the effective inhibition among the sequential net's output nodes.

The discontinuous, intergestural phase transitions may be viewed as nonequilibrium phase transitions of the type seen in other physical and biological systems (e.g., [52]). In such a framework, the rhythmic units in question are characterized as nonlinearly coupled, limit cycle oscillators; the transition is characterized as a bifurcation from a modal pattern that becomes unstable with increasing rate to another modal pattern that retains its stability. The obvious implications for model development are that when the model system is trained to perform extended repetitive /pi/ or /ip/ sequences, the rhythmically active gestures should be similarly characterizable, and should display corresponding intergestural phase transitions.

ACKNOWLEDGEMENTS

This work was supported by NIH Grant DC-00121 and NSF Grant DBS-9112198 to Haskins Laboratories. I am grateful to Dani Byrd and Philip Rubin for discussion and helpful comments on an earlier draft of this paper.

REFERENCES

[1] Browman, C., & Goldstein, L. (1986). Towards an articulatory phonology. In C. Ewan, & J. Anderson (Eds.), *Phonology yearbook 3* (pp. 219-252). Cambridge: Cambridge University Press.

[2] Browman, C. P., & Goldstein, L. (1991). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology: I. Between the Grammar and the Physics of Speech.* (Pp. 341-338) Cambridge, England: Cambridge University Press.

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

[4] McGowan, R.S. & Saltzman, E. (1995). Incorporating aerodynamic and laryngeal components into task dynamics. *Journal of Phonetics, 23,* 255-269.

[5] Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research, Series 15,* 129-144.

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

[7] Jordan, M. I. (1986). *Serial order in behavior: A parallel distributed processing approach* (Tech. Rep. No. 8604). San Diego: University of California, Institute for Cognitive Science.

[8] Jordan, M. I. (1990). Motor learning and the degrees of freedom problem. In M. Jeannerod, (Ed.). *Attention and Performance XIII.* Hillsdale, NJ: Erlbaum.

[9] Jordan, M. I. (1992). Constrained supervised learning. *Jounal of Mathematical Psychology, 36,* 396-425.

[10] Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science, 16,* 307-354.

[11] Bailly, G., Laboissière, R., & Schwartz, J. L. (1991). Formant trajectories as audible gestures: An alternative for speech synthesis. *Journal of Phonetics, 19,* 9-23.

[12] Grossberg, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines (Vol. 1).* New York: Academic Press.

[13] Guenther, F. H. (1994). *Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production.* (Technical Report CAS/CNS-94-012). Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems, Boston, MA.

[14] Kawato, M. (1989). Motor theory of speech perception revisited from minimum torque-change neural network model. *Proceedings of the 8th Symposium on Future Electron Devices, October 30-31, Tokyo, Japan* (Pp. 141-150).

[15] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McLelland, (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Vol 1. Foundations.* Cambridge, MA: MIT Press.

[16] Coker, C. H. (1976). A model of articulatory dynamics and control. *Proceedings of the IEEE, 64,* 452-460.

[17] Boyce, S. (1990). Coarticulatory organization for lip rounding in Turkish and English. *Journal of the Acoustical Society of America, 88*, 2584-2595.

[18] Gelfer, C., Bell-Berti, F., & Harris, K. (1989). Determining the extent of coarticulation: Effects of experimental design. *Journal of the Acoustical Society of America*, *86*, 2443-2445.

[19] Perkell, J. S., & Matthies, M. L. (1992). Temporal measures of labial coarticulation for the vowel /u/: Within- and cross-subject variability. *Journal of the Acoustical Society of America*, *91*, 2911-2925.

[20] Bell-Berti, F., & Krakow, R. (1991). Anticipatory velar lowering: A coproduction account. *Journal of the Acoustical Society of America, 90*, 112-123.

[21] Magen, H. S. (1989). *An acoustic study of vowel-to-vowel coarticulation in English*. Unpublished Ph.D. dissertation, Yale University, New Haven CT.

[22] Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech, 36*, 171-195.

[23] Bell-Berti, F., & Harris, K. (1981). A temporal model of speech production. *Phonetica*, *38*, 9-20.

[24] Henke, W. (1966). *Dynamic articulatory models of speech production using computer simulation*. Unpublished Ph.D. dissertation, MIT, Cambridge, MA.

[25] Kawato, M. (1990). Computational schemes and neural network models for formation and control of multijoint arm trajectory. In W. T. Miller, III, R. S. Sutton, & P. J. Werbos, (Eds.). *Neural networks for control*. Cambridge, MA: MIT Press.

[26] Massone, L. L. E., & Myers, J. (1995). The role of plant properties in point-to-point arm movements. Poster presented at *5th Annual Conference on Neural Control of Movement*, Key West, FL.

[27] Gracco, V. L., & Abbs, J. H. (1989). Sensorimotor characteristics of speech motor sequences. *Experimental Brain Research, 75*, 586-598.

[28] Kollia, H. (1994). *Functional organization of velar movements following jaw perturbation*. Unpublished Ph.D. dissertation, Department of Speech and Hearing Sciences, City Universtiy of New York, New York, NY.

[29] Löfqvist, A., & Gracco, V. L. (1991). Discrete and continuous modes in speech motor control, *Perilus XIV*, Institute of Linguistics, University of Stockholm, Stockholm, Sweden, 27-34.

[30] Munhall, K., Löfqvist, A., & Kelso, J. A. S. (1994). Lip-larynx coordination in speech: Effects of mechanical perturbations to the lower lip. *Journal of the Acoustical Society of America*, *95*, 3605-3616.

[31] Saltzman, E. (1992). Biomechanical and haptic factors in the temporal patterning of limb and speech activity. *Human Movement Science, 11*, 239-251.

[32] Saltzman, E., Kay, B., Rubin, P., & Kinsella-Shaw, J. (1991). Dynamics of intergestural timing. *Perilus XIV*, 47-56, Institute of Linguistics, University of Stockholm, Stockholm, Sweden.

[33] Saltzman, E., Löfqvist, A., Kinsella-Shaw, J., Rubin, P., & Kay, B. (1992). A perturbation study of lip-larynx coordination. In J. J. Ohala, T. M. Nearey, B. L. Derwing, M. M. Hodge, & G. E. Wiebe, (Eds.). *Proceedings of the International Conference on Spoken Language Processing (ICSLP '92): Addendum*. Edmonton, Canada: Priority Printing. Pp. 19-22.

[34] Löfqvist, A., Saltzman, E., Kinsella-Shaw, J., Rubin, P., & Kay, B. (1994). Phase resetting in speech. II. Discrete utterances. *Journal of the Acoustical Society of America*, *95*, (Abstract).

[35] Saltzman, E., Löfqvist, A., Kinsella-Shaw, J.,Rubin, P., & Kay, B. (1994). Phase resetting in speech. I. Repetitive utterances. *Journal of the Acoustical Society of America*, *95*, 2823, (Abstract).

[36] Saltzman, E., Löfqvist, A., Kinsella-Shaw, J., Kay, B., & Rubin, P. (1995). On the dynamics of temporal patterning in speech. In F. Bell-Berti, & L. Raphael, (Eds.). *Studies in speech production: A Festschrift for Katherine Safford Harris*. Woodbury, New York: American Institute of Physics.

[37] Browman, C. P., & Goldstein, L. (1995). Gestural syllable position effects in American English. In F. Bell-Berti & L. Raphael, (Eds.). *Studies in speech production: A Festschrift for Katherine Safford Harris.* Woodbury, New York: American Institute of Physics.

[38] Delattre, P. (1971). Consonant gemination in four languages: An acoustic, perceptual, and radiographic study, Part I. *IRAL*, 31-52.

[39] Krakow, R. A. (1989). *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures*. Unpublished Ph.D. dissertation, Yale University.

[40] Sproat, R., & Fujumura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics, 21*, 291-311.

[41] Lehiste, I. (1980). Phonetic manifestation of syntactic structure in English. *Annual Bulletin of the Research Institute of Logopaedics and Phoniatrics*, *14*, 1-27.

[42] Barry, M. C. (1991). Temporal modelling of gestures in articulatory assimilation. In *Proceedings of the XIIth International Congress of Phonetic Sciences*. (Pp. 7-8). Aix-en-Provence: University of Provence.

[43] Hardcastle, W. J. (1985). Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication*, *4*, 247-263.

[44] Munhall, K., & Löfqvist, A. (1992). Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics*, *20*, 111-126.

[45] Nittrouer, S. (1991). Phase relations of jaw and tongue tip movements in the production of VCV utterances. *Journal of the Acoustical Society of America*, *90*, 1806-1815.

[46] Nittrouer, S., Munhall, K., Kelso, J. A. S., Tuller, B., & Harris, K. S. (1988). Patterns of interarticulator phasing and their relation to linguistic structure. *Journal of the Acoustical Society of America*, *84*, 1653-1661.

[47] Shaiman, S., & Porter, R. J. Jr. (1991). Different phase-stable relationships of the upper lip and jaw for production of vowels and diphthongs. *Journal of the Acoustical Society of America*, *90*, 3000-3007.

[48] Zsiga, E. C. (1994). Acoustic evidence for gestural overlap in consonant sequences. *Journal of Phonetics*, *22*, 121-140.

[49] Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986a). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, *14*, 29-60.

[50] Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986b). Intentional contents, communicative context, and task dynamics: A reply to the commentators. *Journal of Phonetics*, *14*, 171-196.

[51] Tuller, B. & Kelso, J. A. S. (1991). The production and perception of syllable structure. *Journal of Speech and Hearing Research*, *34*, 501-508.

[52] Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*, 347-356.