Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona, Spain. August 3–9, 2003. Universitat Autònoma de Barcelona.

1309

The Dynamics of Error

Marianne Pouplier
Yale University, Haskins Laboratories, USA
E-mail: pouplier@haskins.yale.edu

ABSTRACT

We report findings from speech error experiments that challenge particular assumptions about articulatory effort as they have gained widespread currency in recent phonetic and phonological theory. We find that in errors, speakers often add extra gestures that are phonotactically illegal, but crucially create a symmetric frequency pattern. While this increase in number of articulatory events is unexpected under an effort-based approach, from a dynamic perspective we can understand this phenomenon as a transition to an optimally stable mode of coordination. We discuss why articulatory effort cannot be evaluated without reference to dynamic stability relative to the current context of an utterance. Gestures are never *per se* effortful; rather, changes in complex coordination relations can result from the interplay of different stable attractor basins governed by the dynamic characteristics of speech production.

1. INTRODUCTION

In recent phonetic and phonological theory, the question of why articulatory patterns are subject to change is answered from a teleological perspective: Speakers want their language to be 'optimal.' While optimization for listeners means maximal intelligibility, optimal for the speaker is generally believed to mean something quite different: speakers seek to limit 'articulatory cost' by restricting movements of the articulators to a minimum . What saves us from a complete breakdown of all communication is the fact that we are all speakers and listeners at the same time. This internal "tug of war" [19] between the principles of maximal intelligibility and minimum effort leaves its mark on the surface by making speech variable: the speaker-listener conflict is resolved differently by speakers under different situational demands. When the clarity requirement is less stringent, according to a common assumption, the lazy side of speakers wins out - we see the all too well known but little understood reduction and assimilation phenomena of fast speech. Lindblom's highly influential paper on H&H theory [19] has set much of the tone of the debate, and while there is disagreement about the details of measurement, the fundamental assumption that speaker optimization is not only distinct from but even adversary to listener optimization finds broad acceptance. In this paper, we present conceptual and empirical reasons why restricting a measure of effort to minimized articulator movement is overly simplifying. We discuss how variability in language can fall out naturally from taking more global dynamic principles governing speech production into account.

2. LAZY: IS LESS REALLY MORE?

Despite the elusiveness of any kind of effort metric, the concept of 'articulatory effort' is frequently appealed to in order to explain a whole range of phenomena, particularly varying degrees of coarticulation and undershoot, but also markedness relations, inventory structures and diachronic change (e.g., [2, 4, 13, 14, 18, 19, 20, 21, 25, 27]). Lindblom's H&H hypothesis has gained especially high currency within Optimality Theory and Functional Phonology, leading to the formulation of constraints prohibiting 'far' and 'fast' articulator movement or any movement at all, as in LAZY [13, 14], *Fast [2, 4] or *GESTURE [2]. Mohanan [25], for instance, claims that coronal consonants are "more natural" than non-coronals, because [n] compared to e.g. [m] involves the configuration of least effort.

Peter Ladefoged [17] rightly cautions against notions of 'hard' and 'easy' sounds by pointing out that no observation can be made independent of the observer. For a Navaho, he says, an ejective is easier to produce than a dental fricative. But a much more general point can be raised against observations about 'hard' and 'easy' sounds. Any amount of effort that a particular articulator movement is *ex hypothesei* accredited with cannot be stipulated without taking into account that speech production is governed by the dynamics of coordination Any claims that the pure number of gestures produced can serve as an effort metric, that any gesture produced is tantamount to a proportional increase in effort [2, 13, 14, 15] is not tenable as a context-free statement; it lacks both descriptive adequacy as well as explanatory power. We propose here that the composite structure of coordinated, skilled movement is what has to be the object of observation, for the simple reason that the coordination of several gestures is required to form any larger molecular structure in speech (segments, syllables, or larger structures). Our claim that counting the presence or absence of gestures as a measure of effort is misguided receives empirical support from the study of speech errors.

3. WHY ERRORS ARE OPTIMAL: MORE IS MORE

Speech errors are traditionally thought of as abstract segment substitutions, as for instance when "budget gap" is

ISBN 1-876346-48-5

erroneously pronounced "gudget bap" [29]. In Pouplier & Goldstein [28] and Goldstein et al. [7] we have shown that, due to perceptual biases, traditional, transcription-based studies of speech errors do not necessarily capture an accurate picture of the articulatory events. Instead of e.g., substituting holistically a /k/ for a /t/ in the phrase "top cop", speakers systematically add extra gestures that are communicatively inert and phonotactically illegal. We identified the most frequent type of error to be an addition of a gesture (e.g., a k-like tongue dorsum gesture during /t/) without reduction of the target gesture (e.g., the tongue tip gesture during /t/). In the Goldstein et al. [7] error elicitation study, we used a rapid CVC repetition task with alternating onset consonants but shared rhymes (e.g. "cop top"). We have extended our initial findings to demonstrate that this effect is not confined to repetitive articulatory tasks, but also holds for 'planning' errors. The SLIP technique was employed [1], which uses priming instead of repetition to elicit errors, i.e. subjects pronounce only one word pair at a time. Data for 7 subjects show that with greater than chance probability ($p > 0.01$) the intrusion of a gesture (e.g. tongue dorsum during a production of /t/) is not accompanied with a simultaneous reduction of the target gesture (e.g. tongue tip during /t/). Figure (1) shows an example for a non-errorful and an errorful production of the phrase "tab cab" by the same subject.

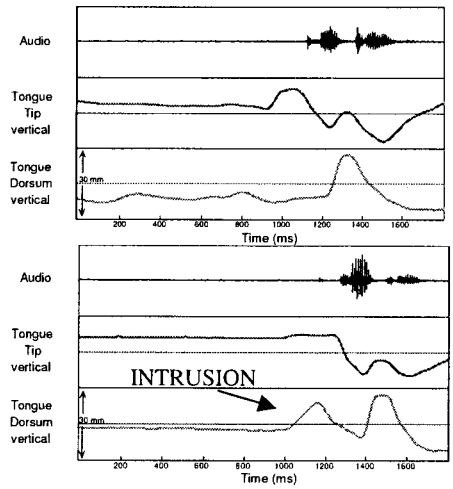

Figure 1: Nonerrorful (top panel) and errorful (bottom panel) production of "tab cab" by one speaker. Bottom panel shows tongue dorsum intrusion during the /t/ of "tab."

From an effort-based approach to articulatory reorganization phenomena, this result seems surprising, since the general premise is that changes to articulatory patterns either result in an articulatory gain (i.e. reduced costs), or at least no net articulatory loss. If we extend this hypothesis to speech errors, we could expect that while errors result in temporal mislocation, effort-wise they would at least be a null-null situation. This would be the case for full segment substitutions, i.e. a co-occurrence of reduction and intrusion, since what is 'added' on one articulator would be 'taken away' from the other.

In Goldstein et al. [7] we interpret the transition to a synchronous production of the initial consonant gestures as instantiation of a hallmark property of nonlinear oscillators, i.e. mode locking. Coordinating multiple events in 1:1 frequency coordination has been shown to be the most stable pattern in movement synchronization tasks: Experiments have demonstrated that under certain conditions (e.g., increased rate), transitions from more complex patterns to 1:1 synchrony can be triggered [8, 9, 26]. While mainly finger-tapping and limb oscillation experiments have been used to investigate this phenomenon, we analyze our articulatory movement data as emergence of a 1:1 frequency pattern in speech production. From a dynamic perspective we can account for this increase in the number of articulatory events as a transition to 1:1 frequency locking. This intrinsically most stable mode of coordination comes to dominate over lexically stable modes of coordination in error-triggering environments. That we find these errors also in non-repetitive task opens the possibility of extending this interpretation to more normal continuous speech.

In the present context it is important to understand that the coupled-oscillator view on slips of the tongue sees errors as arising from a move towards optimization; they are instances of optimal stability. In this sense, if we start producing a /t/ and a /k/ at the same time instead of alternating between them, we are reducing articulatory effort. It is not crucial to our point whether preferred or low-cost behavior is the cause or a by-product of stable dynamics (the latter view is for instance taken by Holt et al. [10]). The main point here is that optimization cannot be understood in a context-free form. That is, the global state of a system can only in limited cases be evaluated by examining locally defined states in isolation [16, 32]. A dynamic approach allows us to recognize how the interplay of different stable attractor basins automatically gives rise to multiple, *equally optimal* preferred states.

Beyond speech errors, other cases of apparently unmotivated gestural insertion come to mind which can speculatively be explained on the basis of optimal dynamic stability: The surfacing of an intrusive /r/ in non-rhotic dialects of English can potentially be stated in the context of increased stability through articulatory re-organization. Following Gick's [6] analysis of intrusive or linking /r/s as ambisyllabic, we can interpret them in the context of the Onset Principle [12]. In dynamic terms, they give rise to a more stable prosodic configuration. CV syllable structures are especially stable because of the specific gestural phasing relations pertaining to onsets [3]. Also, because most syllables have onsets, producing one without an onset would disrupt the global rhythmical patterning. Thus it is not surprising that we find consonant epenthesis as means of hiatus resolution (cf. also consonant epenthesis in French, and Zoll [33] for other languages). If we allow articulatory effort measured by number of gestures to shape grammatical principles, then these phenomena can only be

ISBN 1-876346-48-5

explained by assuming that some other constraint of unclear motivation outranks a context-free notion of effort in all these instances. In the conceptualization developed here, the observed pattern is optimal for the dynamics of the production system, more optimal (and less globally effortful) than the alternative forms that carry fewer gestures. Certainly not *all* epenthesis, reduction and deletion phenomena of the world's languages can be explained on the basis of CV stability, but at least some cases of seemingly arbitrary (surface) gestural insertion can be related to the stability properties of preferred prosodic or rhythmic patterns.

4. WALKING AND TALKING

That what is optimal is context-dependent and moreover not restricted to metabolic factors is known from the animal literature. Holt et al. [10] point out that stability to perturbation is a crucially part of an optimal system, as is minimizing the risk of injury by limiting peak impact force. Also the effect of learning deserves mentioning here. Sparrow and Newell [30] report several studies that demonstrate the positive effect of practice on movement costs. They find that with increasing practice, timing and coordination patterns are fine-tuned to minimize metabolic cost. In the light of the fact that the many (rate and style dependent) coordinative patterns of our native language are highly overlearned and practiced it becomes once more clear that we need to allow several coordination modes to be equally optimal, and that effort weightings on the presence or absence of individual gestures cannot play a role in shaping sound inventories of the world's languages.

A famous study of Hoyt and Taylor [11] on gait changes in horses has established a relation between speed, gait and energy expenditure. With increasing speed, horses automatically switch to a different gait. The different gaits correspond to minima of energetic cost (measured by oxygen consumption) at the respective speeds. This means that a gait is optimal only for a given speed, but within their respective speed range, the gaits converge on the same energy minima; they are equally optimal. Note that the gaits are not confined to a single speed; they can be performed at different levels, but at the expense of a higher energy consumption. This lends further support to our claim that there is no context-free statement of effort. Neither a particular speed, nor a particular displacement, nor a particular gesture is *per se* effortful.

Inspired by the animal gait literature and H&H theory, several linguistic studies have ventured to empirically measure articulatory effort in relation to different speaking styles (e.g., [20, 21, 27]). Matthies et al. [21], for instance, try to trace trade-off relations between peak velocity (taken as correlate of effort), duration and displacement at different speaking rates. They predict that in higher-effort clear speech, coarticulation should be reduced to enhance the clarity of the acoustic signal, while in fast speech overlap and undershoot will be increased at the expense of acoustic distinctiveness; yet they find little evidence for a consistent trade-off relation. While the attempt to identify different speech styles as the different 'gaits of speech' is a promising enterprise, these studies still work on the premise that differences in phonetic implementation arise from Lindblom's "tug of war" between adversary speaker and listener demands on communication, the latter perturbing speakers away from their preferred gait-speed combination. This way clear speech, according to Matthies et al. [21], is different from fast speech because we exert more effort in producing it.

But if we want to take the implications gait studies have for speech seriously, we have to acknowledge that clear speech is different from fast, casual speech, *because it achieves optimality by structural reorganization.* Both speaking styles, or gaits, are equally optimal at their preferred speed. The above studies are promising as accounts of how the articulatory reorganization plays out in different speakers and in different contexts, but the theoretical premise is misguided. A hypothetical study that would have to be undertaken to test to what extent speaking style is truly analogous to different gaits, would need to measure speakers' oxygen consumption while performing fast speech movements at slow rates. Here, however, we very quickly reach the limits of our empirical possibilities and, at least at this point, we have to leave it open to speculation to what extent talking might indeed be like walking.

As a final point, we should ask ourselves why *a priori* we would expect fast speech to be less intelligible, since communication in most circumstances happens under non-optimal acoustic conditions and yet proceeds at very fast speed. In our view it is questionable to assume that fast, casual speech is non-optimal for the listener because prototypical formant values are not reached. Listeners adapt to changes in articulations under rate [23, 24, 31], i.e. parameter variations such as rate leave the system stable perceptually [5]. How different communicational demands can play out in a dynamic system without appealing to a speaker-internal "tug of war" is for instance discussed by Gafos [5]. Coarticulation, context-dependence and "parallel transmission" [22] are always part of the game. We are, after all, speakers and listeners at the same time.

5. CONCLUSIONS

We have argued that articulatory effort can never be assessed as a context-free statement on gestural production. Also energy considerations alone render an insufficient picture of system optimization. Speech as a coordinative system naturally exhibits different modes of coordination which, from an energy and stability perspective, are equally optimal in a given context. Principles of metabolic cost as they have been associated with individual gestures have no effect on shaping the inventory of languages, nor on phonetic implementation.

ISBN 1-876346-48-5

ACKNOWLEDGMENTS

This paper has greatly benefited from extensive discussions with Louis Goldstein and Khalil Iskarous. Work supported by NIH grant HD-01994.

REFERENCES

[1] B. J. Baars, M. T. Motley, & D. G. MacKay, "Output Editing for Lexical Status in Artificially Elicited Slips of the Tongue," *Journal of Verbal Learning and Verbal Behavior,* 14, 382-391, 1975.

[2] P. Boersma. *Functional Phonology,* The Hague: Holland Academic Graphics, 1998.

[3] C. Browman, & L. Goldstein, "Competing constraints on intergestural coordination and self-organization of phonological structures," *Bulletin de la Communication Parlée,* 5, 25-34, 2000.

[4] E. Flemming, "Phonetic Optimization: Compromise in Speech Production," *University of Maryland Working Papers in Linguistics 5: Selected phonology papers from H-OT-97,* 1997.

[5] A. Gafos. "Dynamics in grammar: comment on Ladd and Ernestus & Bayen," *Laboratory Phonology VIII,* Haskins Laboratories, Yale University, 2002.

[6] B. Gick. "A gesture-based account of intrusive consonants in English," *Phonology,* 16, 29-54, 1999.

[7] L. Goldstein, M. Pouplier, L. Chen, E. Saltzman, & D. Byrd, "Gestural Action Units Slip in Speech Production Errors," submitted.

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

[9] H. Haken, C. E. Peper, P. J. Beek, & A. Dafferthofer, "A model for phase transitions," *Physica D,* 90, 176-196, 1996.

[10] K. G. Holt, S. F. Jeng, R. Ratcliffe, & J. Hamill, "Energetic Cost and Stability During Human Walking at the Preferred Stride Frequency," *Journal of Motor Behavior,* 27, 164-178, 1995.

[11] D. F. Hoyt & R. C. Taylor, "Gait and the energetics of locomotion in horses," *Nature,* 292, 239-240, 1981.

[12] J. Itô, "A Prosodic Theory of Epenthesis," *Natural Language and Linguistic Theory,* 7, 217-259, 1989.

[13] R. Kirchner. "Geminate Inalterability and Lenition," *Language,* 76, 509-545, 2000.

[14] R. Kirchner. *An Effort Based Approach to Consonant Lenition,* New York: Routledge, 2001.

[15] K. Kohler, "Segmental Reduction in Connected Speech in German: Phonological Facts and Phonetic Explanations," in *Speech Production and Speech Modelling*, W.J. Hardcastle & A. Marchal, Ed., 69-92. Dordrecht: Kluwer, 1990.

[16] P. N. Kugler, & M. T. Turvey. *Information, Natural Law, and the Self-Assembly of Rhythmic Movement,*. Hillsdale, NJ: Lawrence Erlbaum, 1987.

[17] P. Ladefoged, "Some reflections on the IPA," *Journal of Phonetics,* 18, 335-346, 1990.

[18] Lindblom, "Economy of Speech Gestures," in *The Production of Speech,* P.F. MacNeilage, Ed., 217-246. New York: Springer, 1983.

[19] B. Lindblom, "Explaining Phonetic Variation: A Sketch of H and H Theory," in *Speech Production and Speech Modelling*. W.J. Hardcastle & A. Marchal, Ed.s, 403-439. Dordrecht: Kluwer, 1990.

[20] B. Lindblom, J. H. Davis, S. A. Brownlee, S.-J. Moon, & Z. Simpson. "Energetics in phonetics: A preliminary look," in *Language Production*, O. Fujimura, B.D. Joseph, & B. Palek Ed., 401-415. Prague: Karolinum Press, 1998.

[21] M. Matthies, P. Perrier, J. S. Perkell, & M. Zandipour, "Variation in Anticipatory Coarticulation With Changes in Clarity and Rate," *Journal of Speech, Language, and Hearing Research,* 44, 340-353, 2001.

[22] I. G. Mattingly, "Phonetic Representation and Speech Synthesis by Rule," in *The Cognitive Representation of Speech*, T. Myers, J. Laver, & J. Anderson, Ed.s, 415-420. New York: North Holland, 1981.

[23] J. Miller, & A. M. Liberman, "Some effects of later-occurring information on the perception of stop consonant and semivowel," *Perception and Psychophysics,* 25, 457-465, 1979.

[24] J. L. Miller, "Some Effects of Speaking Rate on Phonetic Perception," *Phonetica,* 38, 159-180, 1981.

[25] K. P. Mohanan, "Fields of Attraction in Phonology," in *The Last Phonological Rule. Reflections on Constraints and Derivations,* J. Goldsmith, Ed., 61-116. Chicago: University of Chicago Press, 1993.

[26] L. Peper, *Tapping Dynamics*, PhD Dissertation, University of Amsterdam, 1995.

[27] J. S. Perkell, M. Zandipour, M. L. Matthies, & H. Lane, "Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues.," *JASA*, 112, 1627-1641, 2002.

[28] M. Pouplier, & L. Goldstein, "Asymmetries in Speech Errors: Production, Perception and the Question of Underspecification," submitted.

[29] S. Shattuck-Hufnagel, "Sublexical Units and Suprasegmental Structure in Speech Production Planning," in *The Production of Speech,* .P.F. MacNeilage, Ed., 109-136. NY: Springer, 1983.

[30] W. A. Sparrow, & K. M. Newell, "Metabolic energy expenditure and the regulation of movement economy," *Psychonomic Bulletin & Review* (5(2)), 173-196, 1998.

[31] Q. Summerfield, "Articulatory Rate and Perceptual Constancy in Phonetic Perception," *Journal of Experimental Psychology: Human Perception and Performance,* 7, 1074-1095, 1981.

[32] M. T. Turvey, E. Saltzman, & R. C. Schmidt, "Dynamics and Task-specific Coordinations," in *Making them move: Mechanics, control, and animation of articulated figures*, N.I. Badler, B.A. Barsky, & D. Zeltzer, Ed.s, 157-170. San Mateo, CA: Morgan Kaufmann, 1991.

[33] C. Zoll. "Ghost Segments and Optimality," *Proceedings of WCCFL 12*, 183-199, 1993.