Fitts' law in tongue and lip movements of repetitive speech

Tanner Sorensen,[1*] Mark Tiede,[2] Adam Lammert,[3] Louis Goldstein,[1] Shrikanth Narayanan[1]
[1]University of Southern California, Los Angeles, CA
[2]Haskins Laboratories, New Haven, CT
[3]Worcester Polytechnic Institute, Worcester, MA
[*]tsorense@usc.edu

INTRODUCTION

Prosody refers to the rhythmic grouping of supra-segmental linguistic units into prosodic domains, such as the grouping of pairs of syllables into feet. In speech production, modulating the rate of articulator movement marks the temporal extent of a prosodic domain. Specifically, movement slows down at the edge of the prosodic domain and speeds up inside the prosodic domain. In the theory of prosody that has been developed in Articulatory Phonology (Browman and Goldstein, 1986, 1989), this rate modulation is the result of varying a parameter of gestural stiffness (a result of local clock-slowing at the edges of prosodic domains; see Byrd and Saltzman, 2003). At the edge of the prosodic domain, low gestural stiffness produces slow movements. Inside the prosodic domain, high gestural stiffness produces fast movements (Byrd and Saltzman, 1998; Byrd et al., 2000; Edwards et al., 1991; Beckman and Edwards, 1992).

Lammert et al. (2018) and Sorensen et al. (2020) derived that the gestural stiffness parameter is equal to the inverse square of the slope of a linear equation relating movement time and an index of movement difficulty known as Fitts' law (Fitts, 1954; Lammert et al., 2018; Kuberski and Gafos, 2019). This theoretical derivation led to the prediction that, at the edge of the prosodic domain, low gestural stiffness makes the slope of Fitts' law steep, but inside the prosodic domain, high gestural stiffness reduces the slope of Fitts' law. The primary contribution of the present study was to confirm this prediction empirically.

METHODS

The present study used 54 electromagnetic articulography recordings (9 participants $\times$ 6 trials/participant $\times$ 1 recording/trial) from the dataset of (Tiede et al., 2019). Each recording consisted of a participant repeating a pair of words in time to a metronome (1 word/beat) for 15 s. From time 0 s to time 7.5 s, the metronome rate was 170 bpm. Metronome rate increased from 170 bpm at time 7.5 s to 230 bpm at time 15 s. The pair of words consisted of two identical syllables that comprised a prosodic foot. Each syllable consisted of a bilabial, coronal, or velar onset stop consonant, a vowel nucleus, and a bilabial, coronal, or velar coda stop consonant. Onset and coda were never identical. Each segment provided one data-point for statistical analysis.

The dependent variable was *movement time*. *Movement time* was the duration in seconds of an individual movement. Movement start- and end-points were the times at which the speed of sensor movement rose above and fell below $0.1v_p$, respectively, where v_p is peak velocity. The position of the sensor was the lower lip for [p], the tongue tip for [t], and the tongue dorsum for [k].

The independent variable was *index of difficulty*, calculated as $I_d = -\mathfrak{w}_{-1}(-W/\mathrm{e}A)$, where $\mathfrak{w}_{-1}$ is the lower real branch of the Lambert W function (Corless et al., 1996), W is error tolerance, e is Euler's number, and A is movement amplitude. Sorensen et al. (2020) derived this index from the Task Dynamics model of speech production (Saltzman and Munhall, 1989). Movement amplitude was the three-dimensional path length from the start- to end-point of an individual movement. Error tolerance was a statistic of a sample of movement endpoints corresponding to a particular participant-articulator combination. We calculated separate error tolerance for 27 (9 participants $\times$ 3 segments) different samples. Error tolerance was the sample standard deviation of the movement endpoints relative to the centroid.

The fixed factors under experimental control were *foot position* (levels: foot-initial, foot-final) and *syllable position* (levels: onset, coda). The random factors were *participant* and *articulator* (levels: lips, tongue tip, and tongue dorsum).

We fit a linear mixed effects model to the dependent variable *movement time*. The fixed effects were independent variable *index of difficulty* and factors *foot position* and *syllable position*. The model included all possible interactions. The model included random intercepts and slopes for *index of difficulty* for each *participant-articulator* combination.

RESULTS

Slope of Fitts' law is steeper at prosodic boundaries than inside the prosodic domain The interaction of factors *syllable position* and *foot position* with *index of difficulty* was statistically significant ($F(1, 4589.4) = 14.77$, $p = 1.23 \times 10^{-4}$). This indicates that *syllable position* and *foot position* interacted to determine the steepness of the slope for Fitts' law. At the edge of a prosodic domain, a 1 bit increase in difficulty results on average in a 32 ms increase in movement time. Inside the prosodic domain, a 1 bit increase in difficulty resulted on average in a 27 ms increase in movement time. Specifically, although syllable onsets did not have a significantly steeper slope in foot-initial position than in foot-final position ($z = 2.50$, $p = 1.23 \times 10^{-2}$; not significant according to the Bonferroni-corrected $\alpha = 0.005$ threshold), syllable codas had a steeper slope in foot-final position than in foot-initial position ($z = -2.93$, $p = 3.34 \times 10^{-3}$). We inferred that Fitts' law had a steeper slope at the edge of the prosodic domain (i.e., foot-initial syllable onsets and foot-final syllable codas) than inside the prosodic domain (i.e., foot-initial syllable codas and foot-final syllable onsets).

As the slope of Fitts' law is the reciprocal of information throughput, the present discovery indicates a connection between linguistic prosody and the information throughput of the motor system. Specifically, our finding implies that information throughput is low at the edge of a prosodic domain (32 bits/s on average) and high inside the prosodic domain (39 bits/s on average). This finding provides an information theoretic basis for linguistic prosody: the consequence of rhythmically grouping supra-segmental linguistic units into prosodic domains is to modulate information throughput such that time intervals of low information throughput separate time intervals of high information throughput.

Slope of Fitts' law is steeper for onset than for coda position The interaction of factor *syllable position* with *index of difficulty* was statistically significant ($F(1,4534.1) = 69.88$, $p = 8.28 \times 10^{-17}$). This indicates that the slope of Fitts' law depended on position within the syllable. Specifically, the slope was steeper for movements in syllable onset position than for movements in syllable coda position. Foot-initial syllable onsets had a steeper slope than foot-initial syllable codas ($z = 8.83$, $p = 1.04 \times 10^{-18}$), and foot-final syllable onsets had a steeper slope than foot-final syllable codas ($z = 3.66$, $p = 2.50 \times 10^{-4}$). The present study indicated that the slope for *index of difficulty* was steeper for movements in syllable onset position than for movements in syllable coda position.

Fitts' law holds in all cells of the experiment Fitts' law applied to constriction movements all cells of the experiment. Specifically, the slope of Fitts' law differed significantly from zero for foot-initial syllable onsets ($z = 7.88$; $p = 1.7 \times 10^{-15}$), foot-final syllable onsets ($z = 7.03$, $p = 1.1 \times 10^{-12}$), foot-initial syllable codas ($z = 4.49$, $p = 3.6 \times 10^{-6}$), and foot-final syllable codas ($z = 5.56$, $p = 1.3 \times 10^{-8}$). This confirmed that Fitts' law applied to movements in all cells of the experiment.

ACKNOWLEDGMENTS

The authors acknowledge funding through National Institutes of Health (NIH) Grant Nos. R01DC007124 and T32DC009975, and National Science Foundation (NSF) Grant No. 1514544. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or NSF.

REFERENCES

Beckman, M. E., and J. Edwards, 1992, Intonational categories and the articulatory control of duration, *in* Speech perception, production, and linguistic structure, 1 ed.: Ohmsha, Ltd, 359–375.

Browman, C. P., and L. Goldstein, 1989, Articulatory gestures as phonological units: Phonology, **6**, 201–251.

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

Byrd, D., A. Kaun, S. Narayanan, and E. Saltzman, 2000, Phrasal signatures in articulation, *in* Papers in Laboratory Phonology V: Citeseer, 70–87.

Byrd, D., and E. L. Saltzman, 1998, Intragestural dynamics of multiple prosodic boundaries: Journal of Phonetics, **26**, 173–199.

———, 2003, The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening: Journal of Phonetics, **31**, 149–180.

Corless, R. M., G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth, 1996, On the Lambert W function: Advances in Computational Mathematics, **5**, 329–359.

Edwards, J., M. E. Beckman, and J. Fletcher, 1991, The articulatory kinematics of final lengthening: the Journal of the Acoustical Society of America, **89**, 369–382.

Fitts, P. M., 1954, The information capacity of the human motor system in controlling the amplitude of movement: Journal of Experimental Psychology, **47**, 381–391.

Kuberski, S. R., and A. I. Gafos, 2019, Fitts' law in tongue movements of repetitive speech: Phonetica.

Lammert, A. C., C. H. Shadle, S. S. Narayanan, and T. F. Quatieri, 2018, Speed-accuracy tradeoffs in human speech production: PloS One, **13**, e0202180.

Saltzman, E. L., and K. G. Munhall, 1989, A dynamical approach to gestural patterning in speech production: Ecological Psychology, **1**, 333–382.

Sorensen, T., A. Lammert, L. Goldstein, and S. Narayanan, 2020, Derivation of fitts' law from the task dynamics model of speech production. (arXiv:2001.05044).

Tiede, M., C. Mooshammer, and L. Goldstein, 2019, Noggin nodding: Head movement correlates with increased effort in accelerating speech production tasks: Frontiers in Psychology, **10**, 2459.