American English Pitch Accent Dynamics: A Minimal Model

Khalil Iskarous[1], Jeremy Steffman[2], Jennifer Cole[3]

[1] University of Southern California, [2]The University of Edinburgh, [3]Northwestern University
kiskarous@usc.edu, jennifer.cole1@northwestern.edu, jeremy.steffman@ed.ac.uk

ABSTRACT

What is the minimal mathematical model that can generate the F0 trajectories for a system of pitch accents? In this work, we propose a nonlinear coupled dynamical systems theory of American English pitch accents with a single basic parameter. As that parameter increases, F0 profiles for different pitch accents are generated. The terms in the differential equation are based on a novel dynamical analysis of a large database of F0 productions in terms of measurements of F0 peak, peak velocity, and the time to achieve peak velocity. We describe the basic dynamical properties of pitch accents in our database and argue for the proposed model as the simplest one that realizes all the major dynamic F0 properties of the pitch accent system. We argue that the proposed model describes both abstract phonological and concrete phonetic aspects of the system.

Keywords: Intonation; Pitch Accents; Dynamical Systems; Fitzhugh-Nagumo Differential Equation.

1. INTRODUCTION

What are the basic scales or dimensions on which the pitch accents of a language could be contrasted? Current level-based theories of prominence contrast pitch accents in terms of the symbols: L, H, and *, indicating static, relative F0 target values (low, high), and their temporal alignment with a phonological landmark [1,2]. An algorithm translates from the symbolic expressions to F0 trajectories. Yet there is a long research line arguing for the inherent temporal gradience and fundamental variability in the expressions of F0-based prominence [3,4,5,6], emphasizing that it is the F0 *configuration*, not just target values, that are used for pitch-based prominence [3,7]. And polynomial coefficients have been used successfully to describe those configurations [8,9,10].

Another approach is the use of dynamical systems theory to predict F0 trajectories by solving a differential equation that expresses the relation between F0 and its derivative(s) *at every moment in time* [11]. Differential equations are valuable since they bridge the abstract and concrete. The relation between the value of a function and its derivatives is abstract as it holds invariantly (or within stochastic limits) even as the value of the function and its derivative change, but it can also explain concrete details of the shape of the function. Thus, differential equation descriptions can be seen as a bridge between abstract level-based descriptions, and more concrete configuration-based descriptions. Moreover, the differential equation has parameters that can be used as a scale for contrast. The greatest hope for a dynamical account of pitch accents is that they potentially allow for observed properties of the phenomenon to *emerge*, rather than be stipulated symbolically or verbally. This is why we believe that a proposal for an underlying differential equation for a pitch system is worthwhile. For general discussions of differential equation-based approaches to phonology and phonetics, the reader can consult [12,13]. In the original dynamical work on F0 shape [11], a linear differential equation was used. In the present work, we propose a new nonlinear dynamical system of differential equations with a single parameter and show that the variation in shape of pitch contours in Mainstream American English (MAE) requires this additional complexity. Nonlinear differential equations have also recently been used [14,15] to describe the processes of tonal selection, as an application of intentional dynamics to accent and meaning [16].

Our goal is to show, through novel quantitative analysis of F0 dynamics in a large database of MAE intonation, that the minimal model required to account for the dynamical properties is a cubic nonlinear system representing the interaction of F0-raising and F0-lowering forces. The specific system we propose is an instance of the Fitzhugh-Nagumo equation, one of the most fundamental equations in mathematical neuroscience [17]. However, we modify these equations in a novel way, to represent discrete, as opposed to rhythmic movement [18]. We will show that this system for describing pitch accents combines phonetics and phonology in a highly organic way, contributing to the solution of the problem of how phonetics and phonology combine in the description of intonation [19].

2. DATA AND METHODS

The pitch accent model we propose is based on data from 130 MAE speakers, aggregated from several imitative speech production experiments [29,30]. In each experiment, on a given trial, participants heard two model utterances that exemplified a particular tune. F0 in the model utterances was resynthesized based on straight-line approximations from [1]. The participant produced the heard tune on a metrically and syntactically similar sentence (e.g., *heard*: "He answered Jeremy", "Her name is Marilyn"; *produced*: "They honored Melanie"). Here we focus on F0 of the final trisyllabic stress-initial word, which bears the nuclear (phrase-final) pitch accent. Data from 70 speakers producing MAE H*, L+H* and L*+H in all intonational boundary contexts (H-H%, H-L%, L-H%, L-L%) is used, with 48 repetitions per speaker, per accent. Data for L* is taken from two different experiments in which 60 speakers each produced 72 repetitions of L* (and other accents) across all four boundary contexts. F0 was extracted from the nuclear word and time-normalized to 30 samples.

A differential equation describes the functional relation at every moment of time between the function's value, e.g., and the value of its derivatives (e.g., velocity $\frac{dx}{dt}$ and maybe acceleration $\frac{d^2x}{dt^2}$). To determine a differential equation, dynamical properties of the data are examined, and the simplest differential equation accounting for those properties is proposed. The equation can then be solved and predicted solutions can be assessed for their fit to the data. In the study of speech dynamics, the peak velocity (PV) and the time at which peak velocity is reached (TTPV) are of paramount importance in the attempt to induce a differential equation [20,12,21], and they will be two of the measurements we present as clues, along with the error-corrected extremum (max/min) F0 during the pitch accents. Though TTPV is not frequently used as a measure in intonation modeling, it serves to quantify the amount of delay in pitch with respect to some supralaryngeal event like a stressed vowel. Indeed, the notion of delay is one of the earliest innovations of the Autosegmental-Metrical approach to tone, in Goldsmith's introduction of the star diacritic * to signify alignment of tonal to non-tonal tiers [23].

3. EMPIRICAL RESULTS

The top panel of Figure 1 shows mean curves of each pitch accent over the entire data, a qualitative view of the F0 trajectories (a) and F0 velocity (b) for the initial 2/3 of the entire nuclear tune interval (to exclude most of the final region that implements the boundary-marking tones). Theses panels show the extrema of the pitch accents and where they are reached. F0 trajectories are normalized to start at the same value. The trajectories show that these data are highly representative of observations made throughout the last few decades about the pitch accents of MAE: the rises rise to different extents and show different alignments with the stressed syllable. Note that H* reaches a lower F0 peak than the bitonal accents. Since this paper is about the dynamics of F0, it is also important to view the F0 velocity curves (b), which shows the expected earlier rise of L+H* vs. L*+H [2], for instance, by a later achievement of peak velocity for the latter.

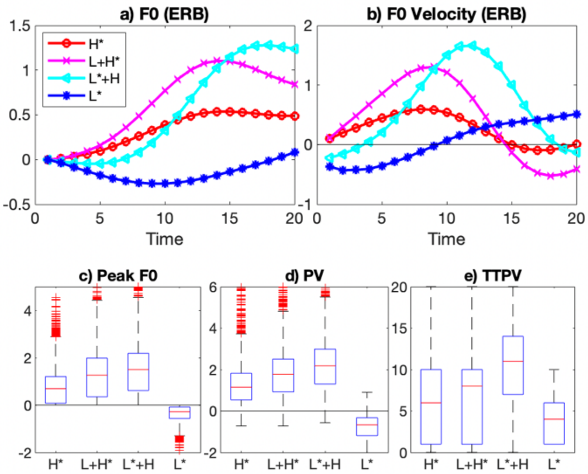

Figure 1: a) Mean F0 trajectories. b) Mean F0 velocity. c) Peak F0, d) Peak velocity, e) Time to peak velocity by pitch accent type.

We do not aim in this paper to provide a statistical determination of the phonetic properties of the AM pitch accent labels. Our goal, rather, is to glean from these dynamical measurements clues as to the relation between F0 and its derivatives. We use the effect size metric *Cohen's d*, which is a measure of the difference in means between two distributions, normalized by pooled standard-deviation [22], to assess the difference between accents in their dynamical F0 measures. To judge magnitude of effect, Cohen [22] argued that a .5 difference is a *medium* effect, and a .8 difference is *large*. Within the rises, for Peak F0, there is a medium effect for H* vs. L+H* (d = -0.58), a large effect for H* vs. L*+H (d = -.77), but there is no effect for L+H* vs. L*+H. For PV, there is a medium effect for H* vs. L+H* (d = -0.50), a large effect for H* vs. L*+H (d = -.81), but there is no effect for L+H* vs. L*+H. For TTPV, there is no effect for H* vs. L+H*, a medium effect for H* vs. L*+H (d = -.55), and a medium effect for L+H* vs. L*+H (d = -.53). L* is distinct from the rises with a large effect for Peak F0 and PV. Considering all three rising accents, we observe significant overlap in the distributions for each F0 measure, resulting in mostly medium effect sizes, signalling a gradient scale of dynamic variation. But despite this variation, the Cohen's d above shows a pattern among the MAE rising pitch accents: *the* *higher F0 goes, the faster it rises, and the later it rises*. This will be our first clue to the underlying dynamic. Of course, it's possible to have a model in which Peak F0, PV, and TTPV are independently controlled, but the variation we see in Figure 1 points to joint control of these dynamical measures. Since a dynamical model captures the relation between F0 and its derivatives, it should be able to capture at least some aspects of this covariation, with as few parameters as possible.

Another observation from Figure 1 is that, as has been remarked in the literature, L*+H is often "scooped", falling before it rises. This is crucial to a dynamical model, because for the same value of F0 the system can output a negative velocity (for the falling part of the scooped rise) or a positive velocity (for the rising part). This means that for the dynamic model of pitch accents, velocity is not a single-valued function of F_0, and instead requires two variables, each with its own differential equation, not just one [17]. Another alternative is to use acceleration as part of the model, however, we have found that the resulting model does not capture MAE pitch accent shapes without having several controllable parameters. As will be seen for the proposed pitch accent model, these two variables can be interpreted as the level of excitation of the forces for raising and lowering pitch, respectively.

To summarize, the dynamical clues are: 1) rise later-rise more; 2) the dynamical system needs to have at least two interacting differential equations.

4. DIFFERENTIAL EQUATION MODEL

We first motivate the proposed differential equation to account for the dynamical properties apparent from Figure 1, and then offer a brief interpretation for how it operates in the phonetic and phonological systems. Instead of presenting the entire differential equation directly, we will build it bit by bit from the simplest interesting differential equation, so that the properties gained from each term are clear. The hierarchy of differential equations have polynomial functions on the right-hand side: starting from the simplest linear polynomial and ending with a cubic polynomial. While it's possible to consider more complex, non-polynomial functions, our focus on polynomials is justified by Taylor's Theorem showing that any function can be approximated with a polynomial of arbitrary degree. Also, we consider only 1-parameter differential equations and models with velocity, but no acceleration, as we seek a minimally complex model with the fewest number of controllable parameters.

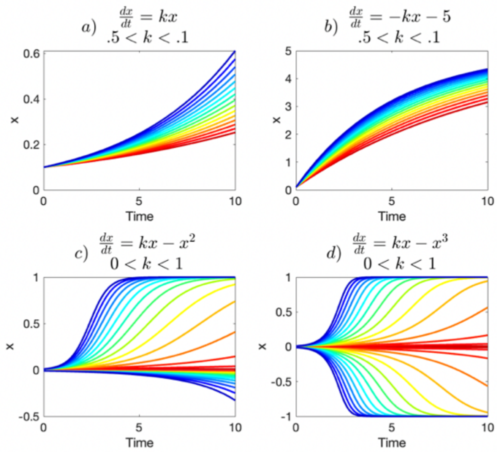

Figure 2: A hierarchy of 1-paramater differential equations. k is swept from a low number (red) to a high number (blue).

Figure 2 shows the solutions of several differential equations with a single parameter *k,* with trajectories generated with different values of *k* within the specified range. In (2a) the slope function $\frac{dF_0}{dt}$ (where x represents F_0) is linear and positively related to the value F0, which yields exponential growth with the rate of growth set by the parameter *k*. The F0 trajectories generated by different *k* values with this model are like those of rising pitch accents, but unlike empirical F0 trajectories, these trajectories rise without bound, so the model fails. In (2b) *k* is negative and a constant is added, which yields bounded growth and generates earlier and later rise patterns, as seen for the different rising accents in the empirical data. But in this model velocity peaks at time 0, where curves are steepest, contrary to the empirical data (Figure 1), so the model fails. For a quadratic polynomial as in (2c), as x approaches 1, x and x^2 balance (become equal in value), which means their difference becomes 0, at which point the trajectory flattens at an equilibrium value. Here it's possible to reach equilibrium with peak velocities at time points later than 0, as in the empirical F0 rises. However, note that the trajectories generated from this model exhibit the pattern "rise-later, rise-less" (trajectories where the rise starts later, rise to a lower F0 peak value), the opposite to the empirical pattern for rises, so this model also fails. The trajectories above 0 are for an initial small positive x. If x starts as a small negative, however, which we show below 0, there is an exponential fall. For a cubic polynomial (d), there is balance of the linear and cubic terms at two equilibria 1 and -1, so there can be stable (bounded) rises and falls as observed in the empirical trajectories. In this model we see the germ of the tonal constructs Low (L) and High (H), emergent here as stable F0 goals at -1 and 1. As already noted, to get to the Low equilibrium requires a negative initial condition. Also, this model, like that in (2c), generates the pattern of "rise-later rise-less", opposite to the empirical pattern, so again, the model fails.

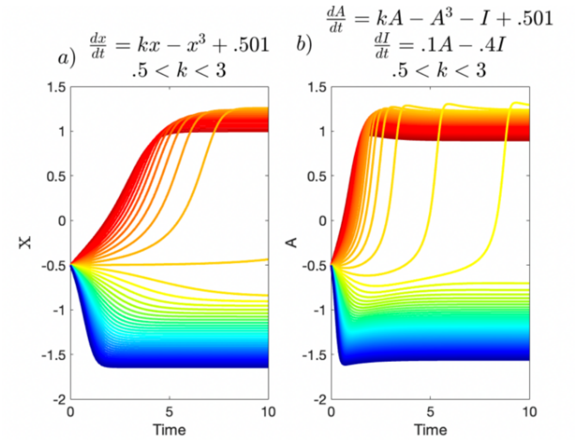

Figure 3: Cubic Models with broken L/H symmetry (a) and, in addition, (b) to an Activator/Inhibitor dynamic.

Figure (3a) shows the result of breaking the symmetry between the high equilibrium at 1 and the low equilibrium at -1 of Figure (2d) by adding a small constant .501. All the curves in (3a) rise from the same baseline -0.5 and the constant 0.501 is chosen to be just large enough to release the system from the initial inhibited value. As k varies, both the L and H equilibria are stable. Lower k (red) is H*-like, rising early to a low extent, and as k rises (orange), we get to intermediate L+H*-like, then L*+H trajectories, before L* for the highest k. Therefore, a key dynamic property of the data in Figure 1 *emerges* from adding that constant .501. However, the differential equation in (3a). has a major problem: rises can only rise, there is no ability to fall then rise, as we see in many scooped L*+H.

Our proposed model is in Figure (3b). As discussed below, we interpret A as F0. This model, with the help of the interacting variable I, is also able to generate L*+H scooped rises. There are two variables, an Activator A, and an Inhibitor I. The A equation is the same as in (3a), except that now A's change also depends on the value of the second variable, I. The negative sign on I signifies that I inhibits A. I depends on A, but positively due to the positive coefficient of the A term in the function for $\frac{dI}{dt}$. As the value of A increases (an F0 rise), it increases I, which in turn inhibits A. This Activator-Inhibitor dynamic is fundamental to mathematical biology in general [24] and theoretical neuroscience in particular [25, 26]. The specific instantiation we developed in (3b) is a version of the Fitzhugh-Nagumo equation [17] that we modified via the equation for I so it will not generate oscillatory/rhythmic trajectories, but instead generate trajectories that achieve one of two discrete equilibria, corresponding to the H/L tone targets of the MAE pitch accent system. Note how variation in a single parameter k, as it is gradiently swept, generates rises with properties as we observe, as well as falls. Even small effects like the earlier fall for L* vs. the rises is captured. We can therefore say that from the variation of one parameter *emerges* a set of correlated dynamical properties. Note however, that the intonation system of MAE does not need to use every value of k. Categories of pitch accents can emerge from the selection of certain k values (or regions in k space) as the conventional signalling values of the language. Other dialects of English, or different languages, could divide the k-scale differently.

We can also interpret A to represent the level of motor factors leading to a rise in F0 (e.g., Cricothyroid muscle) and I to be the level of activation of motor factors leading to a lowering of F0 (e.g., Thyroarytenoid muscle). The coupled system of differential equations represents the internal structure of the motor system interaction through activation and inhibition, as we know muscular systems to be organized since Sherrington's work [27]. We propose that the motor system is phonologically parameterized by k, and that values of k are specified at least in part by the prominence setting system. Examination of other languages could also reveal the need for different constants in the equation, or even more or fewer terms in the equation, while remaining in the Activator-Inhibitor framework. Therefore, the model is a non-dualistic model where cognition and the motor system interact directly in a language-dependent way.

5. CONCLUSION

We have proposed a coupled nonlinear system of differential equations representing the motor-system governing pitch control, with one phonological parameter whose variation leads to phonetic F0 trajectories representing pitch accents. It's possible that exposure to a dialect of a language leads to a few regions of the scale being more prominent than others, resulting in dialectal variation of the pitch accent system within a language [22]. Finally, though the system we have proposed represents the same motor system that all humans possess, there can be phonologically induced differences in the terms and parameters of the interaction, while maintaining the A/I interaction, leading to different systems of pitch variation in different languages.

Acknowledgments: Project supported by NSF BCS-1944773 to JC and PAC-1246750 to KI.

6. REFERENCES

[1] Pierrehumbert, J. 1980. *The phonology and phonetics of English intonation*. Ph.D. thesis, MIT. Distributed 1988, Indiana University Linguistics Club.

[2] Beckman, M., Hirschberg, J., Shattuck-Hufnagel, S. 2005. The original ToBI system and the evolution of the ToBI framework. In: Jun, S-A. (ed), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press, 9–54.

[3] Bolinger, D.L. 1952. Intonation: Levels Versus Configurations. *Word*. 7, 199–210.

[4] Bolinger, D.L. 1961. *Generality, Gradience, and the All-or-None.* Mouton.

[5] Kohler, K. 2005. Timing and communicative functions of pitch contours. *Phonetica*. 62, 88–105.

[6] Cole, J. 2015. Prosody in Context: A review. *Language, Cognition, and Neuroscience*, 30, 1–31.

[7] Barnes, J., Burgos, A., Veilleux, N, Shattuck-Hufnagel, S. 2021.On (and off) ramps in intonational phonology: Rises, falls, and the Tonal Center of Gravity. *Journal of Phonetics,* 85, 1–20.

[8] Kochanski, G., Grabe, E., Coleman, J, Rosner, B. 2005. Loudness predicts prominence; fundamental frequency lends little. *JASA,* 118, 1038-1054.

[9] Grabe, E., Kochanski, G., Coleman, J. 2007. Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and Speech,* 50, 281-310.

[10] Reichel, U. D. 2017. CoPaSul Manual: Contour-based, parametric, and superpositional intonation stylization. *arXiv:1612.04765*.

[11] Prom-on, S., Xu, Y., Thipakorn, B. 2009. Modeling tone and intonation in Mandarin and English as a process of target approximation. *JASA,* 125, 405-424.

[12] Sorenson, T., Gafos, A. 2016. The Gesture as an Autonomous Nonlinear Dynamical System. *Ecological Psychology,* 28, 188-215.

[13] Iskarous, K. 2017. The relation between the continuous and the discrete: A note on the first principles of speech dynamics. *Journal of Phonetics,* 64, 8-20.

[14] Roessig, S., Mücke, D., Grice. M. 2019. The dynamics of intonation: Categorical and continuous variation in an attractor-based model. *PLOS ONE,* 15, e0231221.

[15] Roessig, S., Mücke, D. 2019. Modeling Dimensions of Prosodic Prominence. *Frontiers in Communication,* 4, 10.3389.

[16] Kelso, J.A.S. 1994. The informational character of self-organized coordination dynamics. *Human Movement Science,* 13, 393-413.

[17] Izhikevich, E. 2010. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting.* MIT.

[18] Hogan, N., Sternad, D. 2007. On rhythmic and discrete movements: reflections, definitions and implications for motor control. *Experimental Brain Research,* 4, 181, 13-30.

[19] Ladd, R. 2022. The Trouble with ToBI. In: Barnes, J. Shattuck-Hufnagel, S. (eds), *Prosodic Theory and Practice*. MIT.

[20] Perrier, P., Abry, C., Keller, E. 1988. Vers une modelisation de la langue. *Bulletin de la Communication,* 2, 181, 45-63.

[21] Iskarous, K., Pouplier, M. 2022. Advancements of phonetics in the 21st century: a critical appraisal of time and space in Articulatory Phonology. *Journal of Phonetics,* 2, 95, 101195.

[22] Cohen, J. 1988. *Statistical Power nalysis for the Behavioral Sciences.* Routlage.

[23] Goldsmith, J. 1974. *English as a tone language*. Distributed in 1981 in D. Goyvaerts (ed). *Phonology in the 1980's. Story-Scientia.*

[24] Turing, A. 1952. The Chemical Basis of Morphogenesis. *Philosophical Transactions of the Royal Society B,* 237.

[25] Fitzhugh, R. 1961. Impulse and physiological states in models of nerve membrane. *Biophysical Journal,* 1, 445-466.

[26] Amari, S-i. 1977. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics,* 27, 77-87.

[27] Sherrington, C. 1932. Inhibition as a Coordinative Factor. *Nobel Prize Lecture*.

[28] Burdin, R.S., Holliday, N., Reed, P.E. 2022. American English pitch accents in variation: Pushing the boundaries of mainstream American English. *Journal of Phonetics,* 2, 94, 101163.

[29] Steffman, J., Shattuck-Hufnagel, S., Cole, J. 2022. The rise and fall of American English pitch accents: Evidence from an imitation study of rising nuclear tunes. Proceedings of Speech Prosody 2022, Lisbon.

[30] Cole, J., Steffman, J., Shattuck-Hufnagel, S., Tilsen, S. (in press, 2023). Hierarchical distinctions in the production and perception of nuclear tunes in American English. Laboratory Phonology.