Intonational tunes as compositional, but also autonomous, dynamical objects

Khalil Iskarous[1], Jennifer Cole[2], Thomas Sostarics[2]
[1]University of Southern California (USA), [2]Northwestern University (USA)

One of the main kinds of variation in speech is temporal. Not only are gestures spatio-temporally organized, but gestures compose into time-varying organizational structures like syllables, words, utterances, and tunes for intonation. Are the larger phonological structures, like words and tunes, non-autonomous sequential and parallel compositions of gestural dynamical systems (Saltzman and Munhall, 1989), or do they have dynamical autonomy of their own? This type of question has been an important one in the study of phonology for a long time, with different research communities taking the primary units to be temporally-local speech sounds, features, gestures, tones (Jakobson, Fant, and Halle, 1953; Browman and Goldstein, 1989; Pike, 1949; Pierrehumbert, 1980) vs. temporally-global structures like words (Bybee 2001; Vihman, 2019; Davis and Redford 2019, 2023) and tunes (Bolinger 1951; see also Arvaniti 2011). Within the framework of Articulatory Phonology (AP), the question turns on the equilibrium nature of contrastive objects: they are stable equilibria or attractors, so movement tends to some goal and ends there. Then another control algorithm, e.g., the syllable model (Nam and Saltzman 2003), takes over and organizes multiple gestures non-autonomously in simultaneity-sequentiality relations, relegating the larger structures to being control flow algorithms And these algorithms themselves can be implemented as dynamical systems, using limit cycle entrainment, as in Nam and Saltzman (2003), selection-coordination dynamics, as in Tilsen (2016), or dynamical/neural field theory, as in Roon and Gafos (2016).

In this work, we propose a novel account of intonational (nuclear) tunes in American English as being composed of tonal events traditionally called pitch accents, phrase accents, and boundary tones, but where the tune also has its own dynamic. This account builds on recent dynamical models of intonation (Iskarous et al., 2024), but now viewing the contrastive tonal events not as stable attractors, but as *saddles*, or partially stable and partially unstable, which have been used to account for many aspects of neural, cognitive, and motor organization (Kiebel et al. 2009; Meyer-Ortmanns 2023). By making instability a basic aspect of the temporally local units, as saddles, several of these stable-unstable units can exist in a network to compose a whole global organization. Most importantly, the differential equation system for the local events, taken as a whole, has its own autonomy. We will illustrate this with a model of intonational tunes informed by data from an intonation imitation experiment. 37 speakers imitated three phonologically distinct tunes—Fall (H*L-L%), Rise (L*H-H%), RiseFallRise (H*L-H%)—from stimuli resynthesized with variation in the slope and shape of the F0 trajectory in the medial interval of the phrase accent. Tunes were imitated on novel sentences (N=144 trials) with 4 and 6 syllables in the nuclear tune interval. Our goal is to show how the model accounts for both the compositionality and autonomy of the whole tune, while explaining the distinct elastic behavior of F0 in the interval of the phrase accent on one hand, and the more strictly local behavior of F0 in the tune-initial and -final syllables associated with the pitch accent and boundary tone, respectively (Barnes et al., 2021; Grice et al., 2000). This is part of a research program that aims to extend the idea of saddle networks to other types of articulatory and perceptual organization in speech.

References

Arvaniti, A. (2011). The representation of intonation. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *Blackwell Companion to Phonology* (pp. 757–780). Oxford, UK: Oxford University Press.

Barnes, J., Brugos, 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*, 101020. https://doi.org/10.1016/j.wocn.2020.101020

Bolinger, Dwight L. 1951. Intonation: Levels versus configurations. *Word* 7. 199–210.

Bybee, J. (2001). *Phonology and Language Use*. Cambridge University Press.

Davis, M. and Redford, M. (2019). The Emergence of Discrete Perceptual-Motor Units in a Production Model That Assumes Holistic Phonological Representations. *Frontiers in Psychology*. 10.3389./fpsyg.2019.02121.

Davis, M. and Redford, M. (2023). Learning and change in a dual lexicon model of speech production. *Frontiers in Human Neuroscience.* 10.3389/frhum.2023.893785.

Grice, M., Ladd, D. R., & Arvaniti, A. (2000). On the place of phrase accents in intonational phonology. *Phonology*, *17*(2), 143-185.

Iskarous, K., Cole, J., & Steffman, J. (2024). A minimal dynamical model of intonation: tone contrast, alignment, and scaling of American English pitch accents. *Journal of Phonetics*, *104*, 101309. https://doi.org/10.1016/j.wocn.2024.101309

Jakobson, R., Fant, G., and Halle, M. (1953). Preliminaries to Speech Analysis: Distinctive Features and

Kiebel, S., von Kriegstein, K., Dauunizeau, J., and Friston, K. J. (2009). Recognizing sequences of sequences. *PLOS Computational Biology.* https://doi.org/10.1371/journal.pcbi.1000464.

Meyer-Ortmanns, H. (2023). Heteroclinic networks for brain dynamics. *Frontiers in Network Physiology.* 10.3389/fnetp.2023.1276401.

Nam, H., & Saltzman, E. L. (2003). A competitive, coupled oscillator model of syllable structure. In: M.J. Solé, D. Recasens, J. Romero (Eds.), Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, (pp. 2253-2256).

Roon, K., & Gafos, A. I. (2016). Perceiving while producing: modeling the dynamics of phonological planning. *Journal of Memory and Language, 89*, 222-243.

Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. Ecological Psychology, 1, 333–382.

Sorensen, T., & Gafos, A. (2016). Autonomous nonlinear gestural dynamics. Ecological Psychology, 28, 188–215.

Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics, 55*, 53-77.

Vihman, M. (2019). *Phonological Templates in Development*. Oxford University Press.