A dynamic neural field model of leaky prosody: proof of concept

Jason A. Shaw[1] and Kevin Tang[2]
[1]*Yale University,* [2]*Heinriche-Heine University*

1 Introduction

Sounds patterns in human language are often insightfully described in terms of discrete symbolic units. Accordingly, many of our formal approaches privilege this level of description. However, some phenomena are more challenging for purely discrete models or just fall outside the scope of what such models can explain. These include patterns of incomplete neutralization (Port & Crawford, 1989; Warner, Jongman, Sereno, & Kemps, 2004), gradual sound change (Chen & Wang, 1975), and sub-phonemic changes in representations over the lifespan (Harrington, Palethorpe, & Watson, 2000; MacKenzie, 2017). Accordingly formal approaches to these types of patterns have proposed or adopted some sort of continuous substrate (e.g., Braver, 2019; Bybee, 2002; Pierrehumbert, 2001; Roettger, Winter, Grawunder, Kirby, & Grice, 2014). This paper explores the potential of Dynamic Neural Fields (Schöner & Spencer, 2016) for providing an appropriate substrate to integrate discrete and continuous aspects of sound patterns.

To illustrate the approach, we focus on the empirical issue of "leaky prosody". Recent work has shown that lexical items come to take on the phonetic characteristics of the prosodic environments in which they are typically produced (Seyfarth, 2014; Sóskuthy & Hay, 2017; Tang & Shaw, 2021). Prosodic context often influences the duration, intensity, and pitch with which a word is realized. These phonetic characteristics of prosodic environments can be lexicalized in words that show a distributional skew to a particular type of prosodic environment. For example, in Mandarin Chinese, words that tend to attract a high degree of prosodic prominence are produced with a relatively high pitch (also greater intensity and longer duration), even in prosodically weak environments (Tang & Shaw, 2021). Similarly, words that tend to occur at phrasal boundaries, an environment that lengthens words, end up being longer in duration even in other positions, a result illustrated for New Zealand English (Sóskuthy & Hay, 2017). These effects are lexically specific and sub-phonemic synchronically but may provide the seeds for diachronic change which can be characterized in categorical terms, as in the loss of segments in frequent or informative words (Cohen-Priva, 2017; Cohen Priva, 2015; Piantadosi, Tily, & Gibson, 2011; Zipf, 1949) or the emergence of tone from phrasal prominence (Bang, Sonderegger, Kang, Clayards, & Yoon, 2018; Kang & Han, 2013).

In order to account for the leaky prosody facts in Mandarin, Tang & Shaw (2021) adopt a phonetically-detailed lexicon, as in Exemplar Theory (Pierrehumbert, 2002), prosodic modulation based on language redundancy (Aylett & Turk, 2004; Turk & Shattuck-Hufnagel, 2020), and a feedback mechanism from phonetic output to lexical representation (e.g., Wedel, 2007). A schematic depiction of the proposal is provided in Figure 1. This is a *transformational* architecture in that a phonetically detailed representation of a word, stored in the lexicon, is modulated according to prosodic context, including local predictability, to yield a contextually appropriate phonetic target. The phonetic target then influences the long-term lexical representation through feedback. Lexical representations are updated by experiences with a word, possibly with some,compensation for the effects of prosody, which may be incomplete (Kuzla & Ernestus, 2011; Kuzla, Ernestus, & Mitterer, 2010). The feedback mechanism, whereby context-specific phonetics (even with partial compensation for prosody) update the lexicon, offers a possible account of the leaky prosody facts.

In this paper, we proposal an alternative architecture. Schematized in Figure 2, our alternative, presented here, is a non-transformational *flat* model. Rather than having lexical targets transformed according to

* We would like to thank the Yale Phonologroup community for feedback especially Michael Stern, Manasvi Chaturvedi, Irene Yi, Kevin Roon, Douglass Whalen, Mark Tiede, Natalie Weber, Claire Bowern. All errors are our own.

Proceedings of *AMP 2022*

prosody, we let three forces, a phonetically-detailed lexicon, phonological categories, and prosody, jointly influence the phonetic target. The feedback mechanism, whereby the lexical entry is updated based upon how words are produced in context is retained in the flat model and remains a key part of the explanation for leaky prosody. One of the key advantages of the flat model comes from learnability, as each input to phonetic planning can be acquired through surface-based distributional learning, a point we demonstrate in this paper. This contrasts with the transformational approach which treats speech production as a complex, and hitherto unsolved, optimization problem (Turk & Shattuck-Hufnagel, 2020).

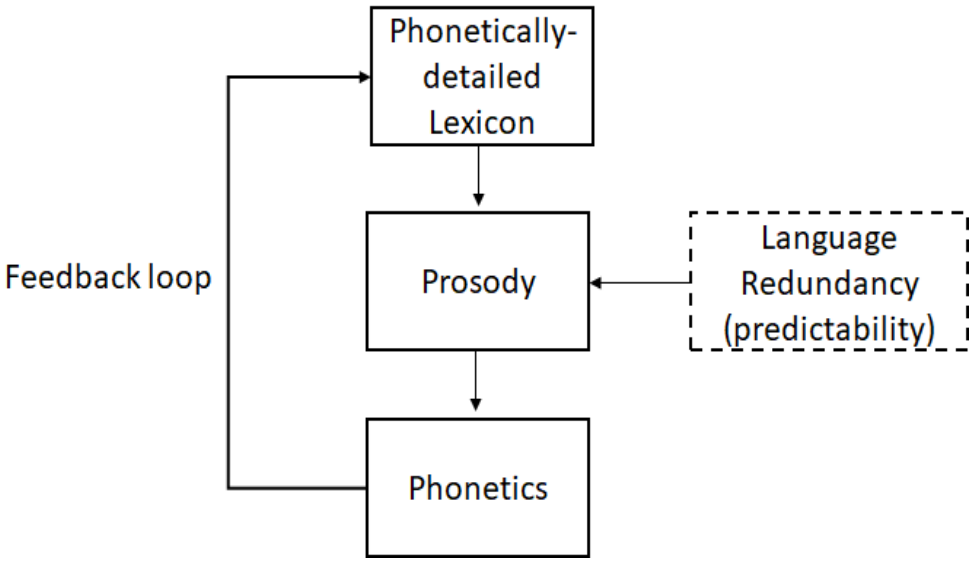


Figure 1. transformational speech production architecture proposed to account for leaky prosody facts in Tang & Shaw (2021), see also (Turk, 2010)

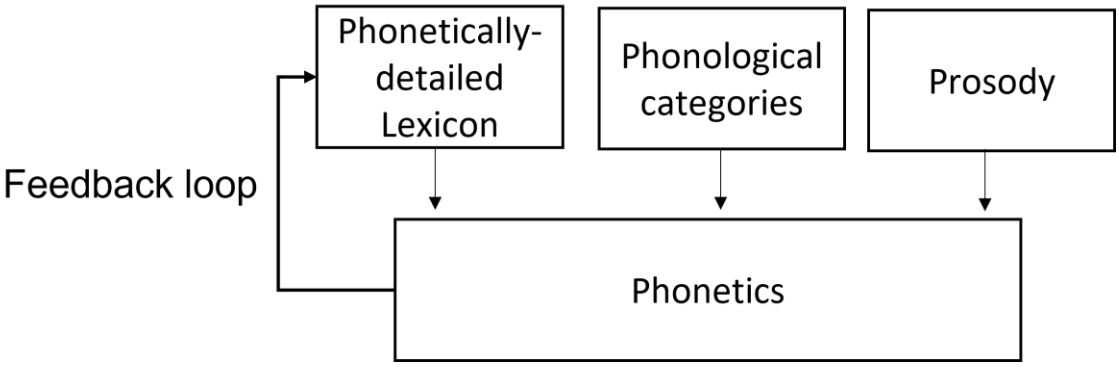


Figure 2. flat (non-transformational) speech production model proposed in the current paper.

The remainder of this paper is organized as follows. In section 2, we introduce Dynamic Field Theory and provide an overview of the model architecture, as situated within this framework. Section 3 provides the formal details of the model. Section 4 presents simulations. We show how pitch planning evolves on a relatively short time-scale in planning a single pitch target, and how feedback drives change in lexical representation over a longer time scale. Section 5 discusses some of the parameters that entered into the model, limitations, and directions for future research.

2 Dynamic Field Theory as a formal framework for the flat model

2.1 *Background* We developed our flat model architecture within the framework of Dynamic Field Theory (Schöner & Spencer, 2016). In this framework, cognitive representations are continuous parameters governed by populations of neurons. In this paper, the continuous parameter of interest is pitch. Populations of neurons sensitive to linguistically-relevant pitch modulation have been localized in left Superior Temporal Gyrus, near other phonetic feature representations (Mesgarani, Cheung, Johnson, & Chang, 2014; Yi, Leonard, & Chang, 2019). In DFT, the distribution of activation across a neural population is represented by a dynamic neural field (DNF). Within our pitch DNF, each field location represents a population of neurons sensitive to a particular pitch value. Activation at each field location evolves over time under the influence of inputs until the DNF stabilizes. A stable activation peak at some location in the field serves as the target for movement.

Stabilization of a pitch DNF over time is illustrated in Figure 3. The z-axis (vertical) represents activation; the x-axis represents the pitch field, where each neuron in the field is selectively tuned to a particular pitch value; the y-axis represents time. In this example, which shows 60 time steps, an activation peak stabilizes at 241 Hz, indicating a pitch target of this frequency.

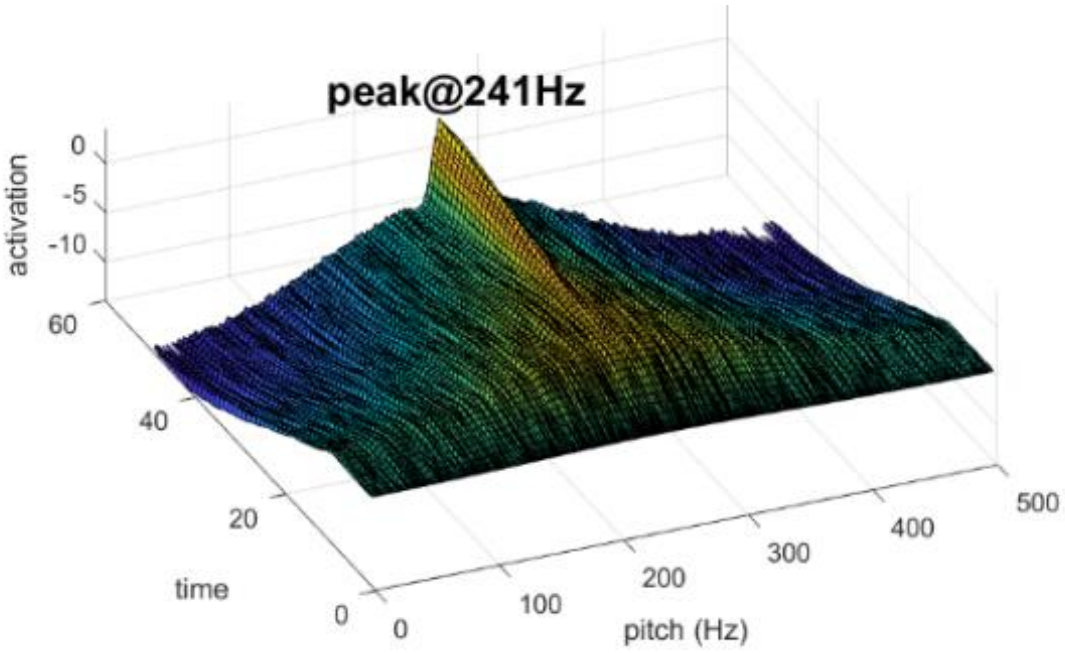


Figure 3. illustration of activation peak stabilization over time in a pitch DNF

DFT has several properties which make it well-suited for developing the flat model we propose in this paper. First, DFT naturally accommodates multiple forces (inputs) on field stabilization by setting parameters to implement selection dynamics (Stern, Chaturvedi, & Shaw, 2022; Stern & Shaw, 2022). Two recent examples come from DFT models of phonetic trace effects in speech errors (Stern et al., 2022) and contrastive hyperarticulation (Stern & Shaw, 2022). The phonetic trace effect is when, in speech errors, sounds that are categorically mis-produced, e.g., [p] in place of [b], still retain some gradient influence of the intended phoneme. For example, the voice onset time (VOT) of [p] produced in error (when [b] was intended) is slightly shorter (closer to [b]) than the VOT of non-errorful [p]. Such sub-phonemic differences in VOT, found in both lab-induced errors from tongue twisters (Goldrick & Blumstein, 2006) and naturally occurring speech errors (Alderete, Baese-Berk, Leung, & Goldrick, 2021) have been modeled as multiple inputs to a DNF representing VOT. Under selection dynamics, strong input from a voiceless stop (long VOT) and weaker input from voiced stop (short VOT) stabilize in a location that is slightly shifted towards the voiced stop, deriving the magnitude of empirically observed trace effects (Stern et al., 2022). A similar account derives contrastive hyperarticulation, the tendency for words with minimal pairs to be hyperarticulated away

from minimal pair competitors. Like the phonetic trace effect, contrastive hyperarticulation occurs in both experimental settings (Baese-Berk & Goldrick, 2009) and spontaneous speech (Wedel, Nelson, & Sharp, 2018). The DFT account involves minimal pair competitors projecting inhibitory input into the field, which drives the location of stabilization away from the target (Stern & Shaw, 2022).

A second useful property of DFT is that it represents cognition—in this case speech production planning—as a time-varying process. This is particularly useful for leaky prosody because our account, at the conceptual level, involves multiple timescales. On a short timescale—the relatively fast process of speech production planning for a single pitch target—prosodic context influences production. On a longer timescale—the relatively slow process of lexical consolidation—the aggregate influences of prosody alter the long term representation of words. Each of these timescales has been modelled within DFT. For example, Roon and Gafos (2016) and Harper (2021) develop DFT models of the millisecond timescale of single consonant production while Gafos and Kirov (2010) capture gradual shift in phonological representations at the longer timescale (see also Tilsen, 2019). Modelling speech production as a cognitive process that unfolds in time distinguishes DFT from stochastic generative models (e.g., Shaw & Gafos, 2015; Shaw & Kawahara, 2018), Exemplar Theories (e.g., Pierrehumbert, 2001), and agent-based models (e.g., Harrington & Schiel, 2017), which treat speech production as a timeless process of statistical sampling.

2.2 *Model overview* Figure 4 provides an overview of our flat model architecture. The pitch planning field (center) is a DNF parameterized for selection dynamics. It receives simultaneous input from a lexical pitch target (lexicon), a phonological pitch target (tone), and a prosodic pitch target (prosody). Over time, given the selection dynamics, the field will stabilize on a pitch target, under the influence of inputs. The stable pitch value serves as the target for a single speech production event (short timescale) and feeds back into the lexicon nudging the long-term representation towards the recent behavior. In the following section, we elaborate on the formal expression of the model.

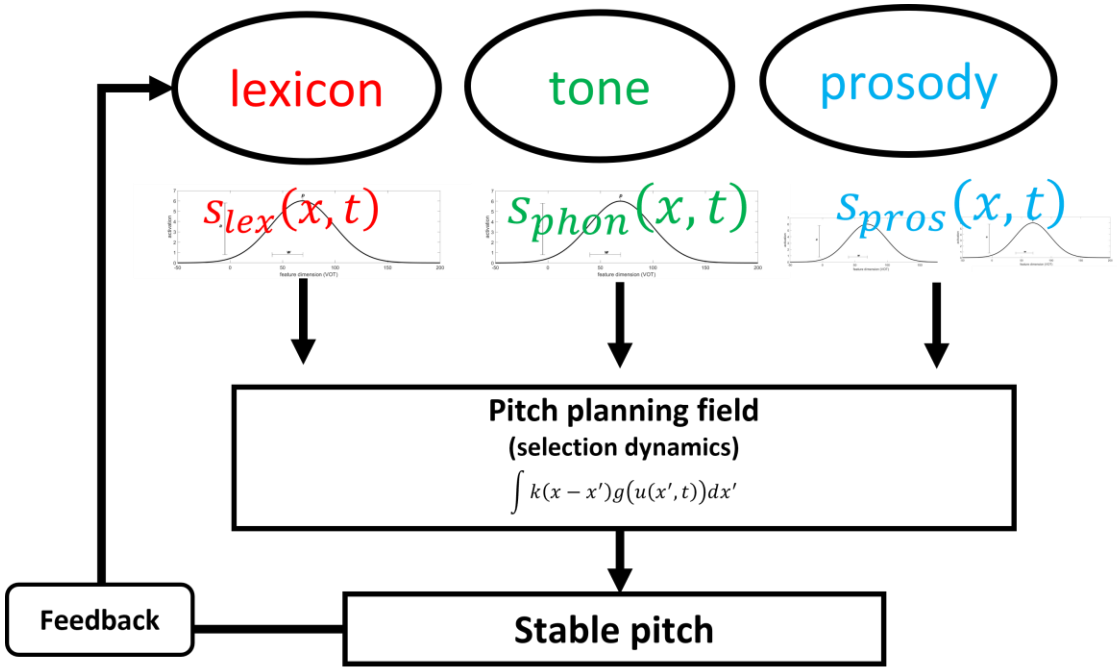


Figure 4. flat model architecture in DFT

3 Formal expression of the model

3.1 *Field evolution* The differential equation governing DNF evolution is provided in (1). The change in activation, $\dot{u}$, over time, *t*, at each field location, *x*, is a function of current activation, u, and four additional parameters: (i) resting activation, *h*; (ii) inputs to the field, $s(x,t)$; (iii) interaction kernel and (iv) noise. The change in activation is weighted by, τ , which dictates the magnitude of change for each time step. For the purposes of the simulations reported on here, τ is held constant, at 20. The noise term is Gaussian-distributed, ξ, of strength *q*. For the simulations here, *q* is held constant, at 1.

When the terms on the right side of the equation sum to zero, activation across the field is stable, i.e., no change. In the absence of inputs to the field, activation will converge on *h*. For all of the simulations reported in this paper, we set *h* to -5. When inputs to the field at some location are greater than 5, then the resting activation will be offset and activation will rise above zero. When this happens, the interaction kernel kicks in, functioning to stabilize a peak in activation. We discuss the formal mechanism of this function in greater detail below, after elaborating on the inputs.

(1) Equation governing change in activation at each field location over time

Change of activation — Resting activation

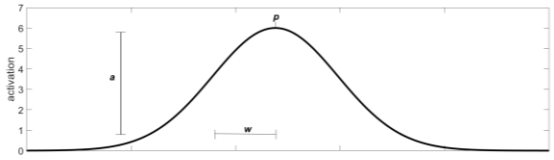
$$\tau\dot{u}(x,t) = -u(x,t) + h + s_{lex}(x,t) + s_{phon}(x,t) + s_{pros}(x,t) + \int k(x-x')g\big(u(x',t)\big)dx' + q\xi(x,t)$$

activation — Inputs to the field — Interaction kernel (property of the field) — noise

3.2 *Inputs to the field* The equation for inputs to the field is given in (2). Inputs take the form of Gaussian distributions with three parameters: (i) the place, *p*, in the field where the distribution is centered, i.e., the mean; (ii) the width, *w*, of the distribution, i.e., the standard deviation, and (iii) the amplitude, *a*, of the distribution. To visualize the shape of the inputs, the equation is plotted below with *a* = 6. An input with this amplitude would drive the field above zero, causing stabilization, given our resting activation level of *h* = -5.

(2) Equation for inputs to the DNF

$$s(x,t) = a\exp\left[-\frac{(x-p)^2}{2w^2}\right]$$

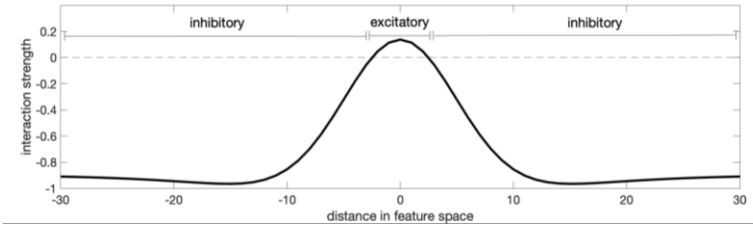


3.3 *Selection dynamics* The equation for the interaction kernel is given in (3). There are three main components to the interaction kernel: (i) a local excitation component, which has a parameter for strength, c_{exc}, and scope, σ_{exc}, and dictates the spread of Gaussian-shaped excitation; (ii) a local inhibition component, which has corresponding parameters, c_{inh}, σ_{inh}; and (iii) a global inhibition component, which covers the entire field with uniform strength c_{glob}. We have set the values of these parameters to ensure selection dynamics. That is, the interaction kernel will promote local activation and inhibit more global activation. The key to deriving selection dynamics from the interaction kernel is to make local excitation stronger and narrower than local inhibition and stronger than global inhibition. This pattern of inequalities is exemplified in the table in (3). Plotting the interaction kernel with the values in the table gives the shape shown below (right), which Schoner and Spencer (2016) refer to as a "Mexican Hat".

(3) Equation for interaction kernel

$$k(x-x') = \frac{c_{exc}}{\sqrt{2\pi}\sigma_{exc}} \exp\left[-\frac{(x-x')^2}{2\sigma_{exc}^2}\right] - \frac{c_{inh}}{\sqrt{2\pi}\sigma_{inh}} \exp\left[-\frac{(x-x')^2}{2\sigma_{inh}^2}\right] - c_{glob}$$

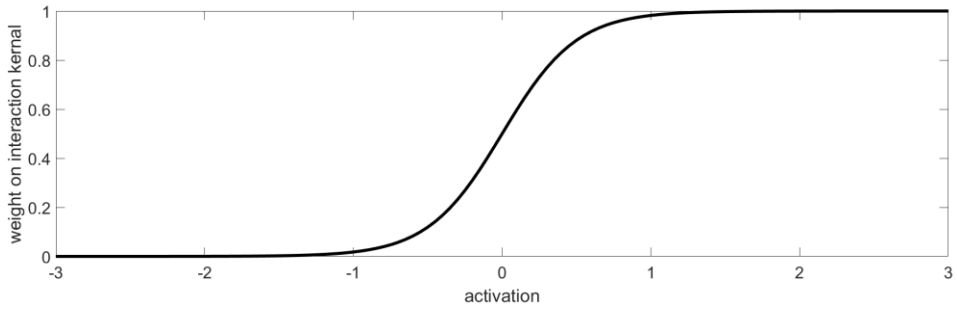
DNF parameters for interaction kernal	
c_{exc} = 15	**"Mexican Hat"**
c_{inh} = 5	
c_{glob} = 0.9	
σ_{exc} = 5	
σ_{inh} = 12.5	



3.4 *Sigmoidal gate* The interaction kernel is gated by $g(u)$, a sigmoidal function, which is shown in (4). The gate prevents the interaction kernel from exerting much influence on field dynamics until activation, at some location in the field, crosses zero. When that happens, $g(u)$, switches abruptly from zero to 1, essentially turning on the activation kernel, which, given the dynamics in (3), functions to create a stable peak at that location in the field. The gate has one parameter, β, which was set to 4 for the simulations below.

(4) Equation for sigmoidal gate

$$g(u) = \frac{1}{1+\exp(-\beta u)}$$



4 Simulations

Using the model specified in the preceding section, we ran two types of simulations. The first simulated the effect of different prosodic positions on pitch. This serves to establish the validity of the flat model in deriving the effects of prosodic context on pitch targets. The second simulated lexical learning as a time varying (longer timescale) process. Here, we sought to derive the leaky prosody facts from lexical updating. We focus on the high tone (T1) of Mandarin. Both simulations made use of the COSIVINA toolbox (Schneegans, 2021).

4.1 *Input parameters* In keeping with the potential learnability advantage of a flat model, we set all input parameters based to values extracted from a large corpus of spontaneous speech. For this purpose, we used the Tang & Shaw (2021) corpus of 1,655 Mandarin speakers. The input parameters for S_{phon}, the

phonological input to the field, was based on the distribution of high tone pitch values found in the corpus. Across the ~41,000 instances of high tones, the average maximum pitch value was 238 Hz and the standard deviation was 94 Hz. To initialize the starting distribution of a lexical item for the S_{lex} input, we sampled 1/500th of the total number of high tones (N = ~86) in the corpus and calculated the mean (241 Hz) and standard deviation (99) of the sample. For the S_{pros} inputs, we divided all of the words in the corpus (~400,000) into 24 equal-spaced bins based upon their local bigram predictability. The assumption, also taken up in Tang & Shaw (2021) is that bigram predictability is directly related to prosodic prominence. This is admittedly a very coarse-grained index of prosodic structure. However, even when more sophisticated linguistic factors are factored into the analysis of prominence, it still seems that local predictability (as well as informativity) play a reliable role in prominence (e.g., Anttila, Dozat, Galbraith, & Shapiro, 2020). Of the 24 equally spaced predictability bins, we choose two bins (4th and 12th), each with ~10,000 tokens, to represent high prominence (low predictability) and low prominence (high predictability) field inputs. Our high prominence bin had a mean of 233 Hz (SD = 100) and our low prominence bin had a mean of 226 Hz (SD = 92). We set the amplitude of all of the inputs to be 6, high enough to individually overcome the resting activation, h = -5.

Input parameters		
$\boldsymbol{S_{lex}}$ **(1st run)**	$\boldsymbol{S_{phon}}$	$\boldsymbol{S_{pros}}$ **(high, low)**
$\boldsymbol{a = 6}$	$\boldsymbol{a = 6}$	$\boldsymbol{a = 6}$
$\boldsymbol{p = 241}$	$\boldsymbol{p = 238}$	$\boldsymbol{p = 233, 226}$
$\boldsymbol{w = 99}$	$\boldsymbol{w = 94}$	$\boldsymbol{w = 100, 92}$

Table 1. Input parameter values estimated from the Tang & Shaw (2021) corpus

4.2 *Short time scale* For the first simulation, we demonstrate how the flat model, based upon surface-based input parameters, faired in capturing the effect of prosodic context on pitch. Figure 5 shows the evolution of the field with the same lexical (S_{lex}) and phonological (S_{phon}) inputs but differing prosodic inputs. The left panel shows the high prominence condition, which stabilizes at 241 Hz; the right panel shows the low prominence condition, which stabilizes at 234 Hz.

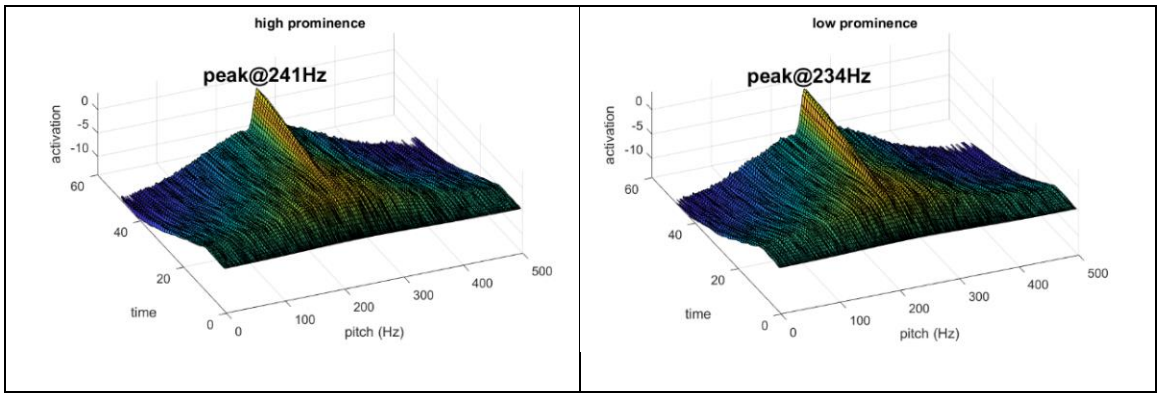


Figure 5. DNF evolution for a high tone word produced with high (left) and low (right) prominence

To better visualize the effect, Figure 6 shows activation across the entire field at the last timestep in the simulations. The difference in activation peak, ~7 Hz, is on the order of magnitude reported in the literature (Tang & Shaw, 2021).

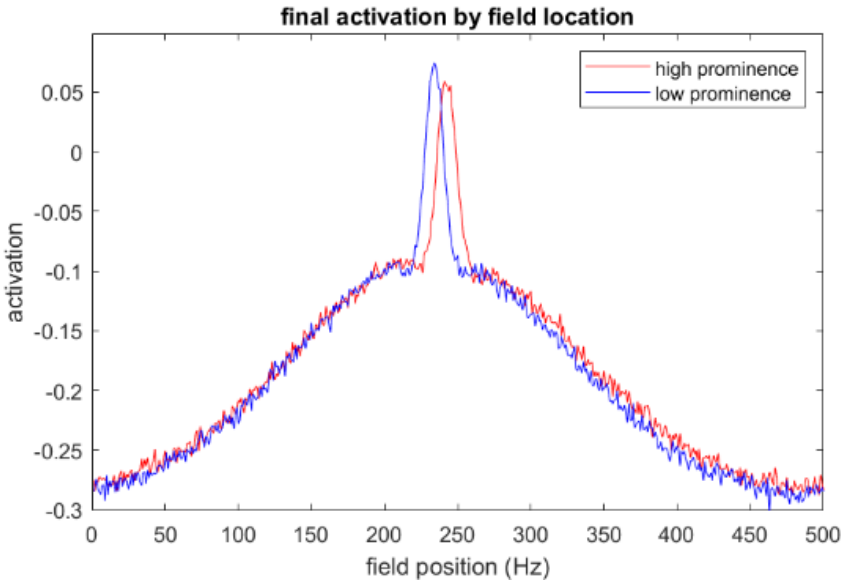


Figure 6. Activation across the pitch field for a high tone word produced with high and low prosodic prominence

4.2 *Long time scale* Having established an effect of prosodic structure, operationalized as local predictability, on single word production (short time scale), we now consider a longer time scale. We simulated two words, 500 times each. The words start out with the same lexical representation, *Slex*. One word is produced systematically in a low prominence position and the other in a high prominence position. At the end of each production of a word, i.e, a short time scale simulation, we updated the lexical representation, i.e., the S_{lex} input, for each word with the new pitch value (based on the location of the stable activation peak in the pitch field). To update, we sampled 86 tokens (the same number used to initialize the distribution) from each S_{lex} and replaced one sampled value (selected at random) with the stable pitch value from the simulation. We then recalculated the S_{lex} parameters, *p* and *w*, based on the new distribution. The new parameters of S_{lex} then served as input to the next production cycle. This feedback loop allows each token to nudge the underlying S_{lex} distributions in the direction of the stable pitch target. Since prosody leaves an impact on the stabilization process, it can come to influence the lexical representation through feedback over many productions.

The simulations results are shown in Figure 7. The left side of the figure shows where the field stabilizes on each short time scale simulation run. There is variation—within a 20 Hz range— from trial to trial in where the pitch DNF stabilizes. The tendency is for the word in a high prominence position to stabilize at a higher pitch value but this is not absolute. On some trials, through the influence of noise, the low prominence word ends up with a higher pitch target. This happens more often at earlier simulations than at later simulations. The reason for this is that the lexical representations start to diverge over time. This is shown in the right panel of Figure 7. The high and low prominence lines gradually diverge, showing consistent separation from about 130th run of the simulation. This is the leaky prosody effect. A small local effect of prosody, if consistently applied, can drive lexical separation between words that started completely homophonous.

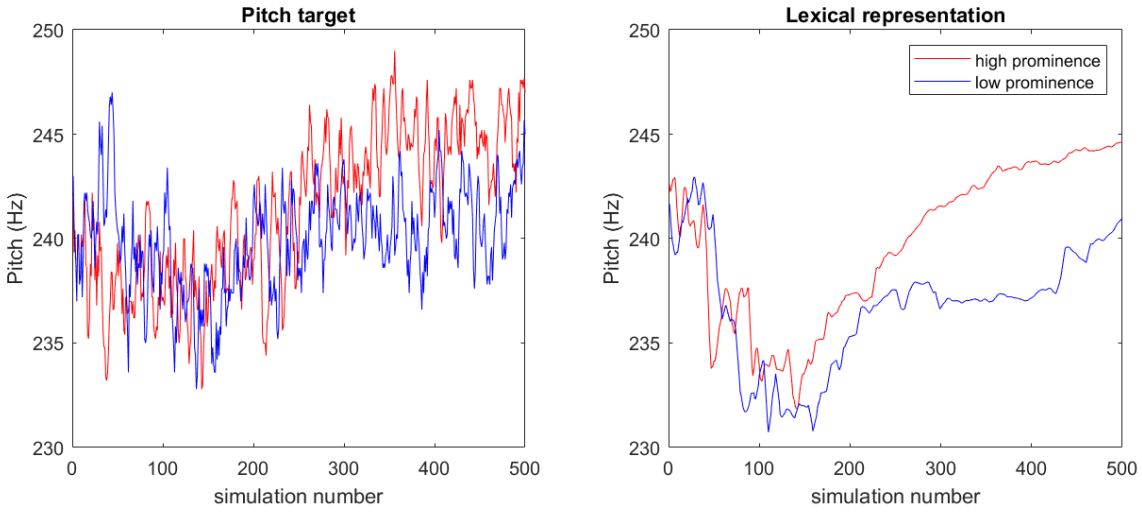


Figure 7. The stable pitch target on each of 500 simulations (left); the value of S_{lex} after updating (right).

5 Discussion

To summarize, our flat DFT model derived the leaky prosody facts of Mandarin pitch. We demonstrated that a small influence of prosody, estimated solely from bigram predictability, could over time cause divergence between two lexical items of the same phonological category. The production inputs to the model were surface distributions calculated from a spontaneous speech sample (~400,000 tokens; 1,655 speakers). The phonological input was based on the complete distribution of maximum pitch values for high tone syllables in the corpus. The lexical input was initialized as a sample of the high tone category. The prosodic input was the distribution of pitch values at fixed levels of surprisal (bigram predictability). We allowed these three inputs to jointly condition the evolution of the pitch DNF. The dynamics of the field ensured stabilization at a fixed location in the field, which varies from trial to trial due to noise. By updating the lexical input based on the location of field stabilization, we showed that a small degree of lexical differentiation emerges over time.

While the results serve as a promising proof of concept, there are many limitations of the current study. We just modelled one tone (Mandarin high tone, T1), just two lexical items, and just one feature dimension, pitch. Moreover, we didn't consider talker normalization or neurophysiologically plausible signal transformations (e.g., ERB, Mel). Additionally, we implemented assumptions about learning, i.e. that mental representations are faithful summaries of experience, which are likely overly simplistic (Olejarczuk, Kapatsinski, & Baayen, 2018). There are many directions in which this work can be expanded to represent more realistic scenarios.

The model has the potential to make interesting predictions for sound change. In the simulations reported here, we only updated lexical inputs based on the location of field stabilization. Of course, phonological and prosodic representations also have to be learned, so a more realistic model would update these as well. In the current simulations, since only the lexical input was updated, the phonological input (tone) functioned to work against lexical drift. That is, since the phonological input does not vary from run-to-run, it represents a constant force for stabilization at the same location in the field; this works against lexical drift. However, even if we updated the phonological representation on each run of the simulation, the anti-drift force of phonology would still persist to some degree, in most realistic situations. The reason is that, typically, there will always be more occurrences of a phonological category than of a lexical item that contains that category. For example, there will always be at least as many instances of the high tone category as there are instances of any particular word that contains that a high tone. Thus, the phonological category itself will be more stable than any given lexical item. If, however, several words of the same phonological category all shift in

the same direction, this could pull the entire phonological category, which would in turn pull the remaining lexical items along. These patterns of sound change are relatively straightforward predictions of the theory, although they require feedback to both the lexical and the phonological representations. The key components that lead to these predictions are (i) a flat model with lexical, phonological, and prosodic inputs to (ii) a DNF with selection dynamics and (iii) feedback to long-term representations at both the phonological and lexical level.

Another consideration in future work is the amplitude of the inputs. We set the amplitude of all three inputs to our pitch DNF to be $a = 6$ so that each one individually could drive the field to stabilize, given a resting activation of $h = -5$. The presumption is that a speaker could plan a pitch target on the basis of any one of these inputs without the other. This would mean, for example, being able to hum the pitch of a tone category or prosodic position without activating a lexical item. Having sufficiently strong inputs from each of these sources at once allows the field to stabilize faster than if there were only one input. This makes the prediction that speech planning is faster when all three of these sources, lexical, phonological, prosodic, are engaged in tandem.

6 Conclusion

We showed that leaky prosody, as evidenced in Mandarin Chinese, can be derived from a flat model of speech production. Lexical, phonological, and prosodic inputs each exert forces on a Dynamic Neural Field representing pitch. Notably, the forces exerted by these inputs reflect surface distributions in a corpus of spontaneous speech. The model parameters are present in the ambient speech and can be acquired through naïve distributional learning. Our simulations showed that the flat model derives the short timescale effect of prosodic prominence on pitch production as well as the longer timescale effect of lexical drift. Pitch targets in words consistently produced in different prosodic environments gradually come to take on (lexicalize) the influence of those environments.

References

Alderete, J., Baese-Berk, M., Leung, K., & Goldrick, M. (2021). Cascading activation in phonological planning and articulation: Evidence from spontaneous speech errors. *Cognition, 210*, 104577.

Anttila, A., Dozat, T., Galbraith, D., & Shapiro, N. (2020). Sentence stress in presidential speeches. *Prosody in Syntactic Encoding, Berlin/Boston: Walter De Gruyter*, 17-50.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*(1), 31-56.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes, 24*(4), 527-554.

Bang, H.-Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T.-J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics, 66*, 120-144.

Braver, A. (2019). Modelling incomplete neutralisation with weighted phonetic constraints. *Phonology, 36*(1), 1-36.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change, 14*(3), 261-290.

Chen, M., & Wang, W. S.-Y. (1975). Sound change: Activation and implementation. *Language, 51*, 228-281.

Cohen-Priva, U. (2017). Informativity and the actuation of lenition. *Language, 93*(3), 569-597.

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology, 6*(2), 243-278.

Gafos, A., & Kirov, C. (2010). A dynamical model of change in phonological representations: The case of lenition. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupé (Eds.), *Approaches to phonological complexity* (pp. 225-246). Berlin: Mouton de Gruyter.

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes, 21*(6), 649-683.

Harper, S. K. (2021). *Individual Differences in Phonetic Variability and Phonological Representation.* University of Southern California.

Harrington, J., Palethorpe, S., & Watson, C. I. (2000). Does the Queen speak the Queen's English? *Nature, 408*(6815), 927-928.

Harrington, J., & Schiel, F. (2017). /u/-fronting and agent-based modeling: The relationship between the origin and spread of sound change. *Language, 93*(2), 414-445.

Kang, Y., & Han, S. (2013). Tonogenesis in early Contemporary Seoul Korean: A longitudinal case study. *Lingua, 134*, 62-74.

Kuzla, C., & Ernestus, M. (2011). Prosodic conditioning of phonetic detail in German plosives. *Journal of Phonetics, 39*(2), 143-155.

Kuzla, C., Ernestus, M., & Mitterer, H. (2010). Compensation for assimilatory devoicing and prosodic structure in German fricative perception. *Laboratory Phonology, 10*, 731-757.

MacKenzie, L. (2017). Frequency effects over the lifespan: A case study of Attenborough's r's. *Linguistics Vanguard, 3*(1).

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science, 343*(6174), 1006-1010.

Olejarczuk, P., Kapatsinski, V., & Baayen, H. (2018). Distributional learning is error-driven: The role of surprise in the acquisition of phonetic categories. *Linguistics Vanguard, 4*(S2).

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences, 108*(9), 3526-3529.

Pierrehumbert, J. (2001). Exemplar dynamics: word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137-157). Amsterdam and Philadelphia: Benjamins.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 101-139). New York: Mouton de Gruyter.

Port, R., & Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics, 17*, 257-282.

Roettger, T. B., Winter, B., Grawunder, S., Kirby, J., & Grice, M. (2014). Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics, 43*, 11-25.

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

Schneegans, S. (2021). COSIVINA: A Matlab Toolbox to Compose, Simulate, and Visualize Neurodynamic Architectures (Version 1.4).

Schöner, G., & Spencer, J. P. (2016). *Dynamic thinking: A primer on dynamic field theory*: Oxford University Press.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition, 133*(1), 140-155.

Shaw, J. A., & Gafos, A. I. (2015). Stochastic Time Models of Syllable Structure. *PLoS One, 10*(5), e0124714 0124711-0124736.

Shaw, J. A., & Kawahara, S. (2018). Assessing surface phonological specification through simulation and classification of phonetic trajectories. *Phonology, 35*(3), 481-522. doi: 10.1017/S0952675718000131

Sóskuthy, M., & Hay, J. (2017). Changing word usage predicts changing word durations in New Zealand English. *Cognition, 166*, 298-313.

Stern, M. C., Chaturvedi, M., & Shaw, J. A. (2022). *A dynamic neural field model of phonetic trace effects in speech errors.* Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society.

Stern, M. C., & Shaw, J. A. (2022). Neural inhibition during speech planning contributes to contrastive hyperarticulation. *arXiv preprint arXiv:2209.12278*.

Tang, K., & Shaw, J. A. (2021). Prosody leaks into the memories of words. *Cognition, 210*, 104601.

Tilsen, S. (2019). Motoric mechanisms for the emergence of non-local phonological patterns. *Frontiers in psychology*. doi: https://doi.org/10.3389/fpsyg.2019.02143

Turk, A. (2010). Does prosodic constituency signal relative predictability? A smooth signal redundancy hypothesis. *Laboratory Phonology, 1*(2), 227-262.

Turk, A., & Shattuck-Hufnagel, S. (2020). *Speech Timing: Implications for Theories of Phonology, Speech Production, and Speech Motor Control* (Vol. 5): Oxford University Press, USA.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics, 32*(2), 251-276.

Wedel, A., Nelson, N., & Sharp, R. (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language, 100*, 61-88.

Wedel, A. B. (2007). Feedback and regularity in the lexicon. *Phonology, 24*(01), 147-185. doi: doi:10.1017/S0952675707001145

Yi, H. G., Leonard, M. K., & Chang, E. F. (2019). The encoding of speech sounds in the superior temporal gyrus. *Neuron, 102*(6), 1096-1110.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley Press.