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

Jason A. Shaw[1] & Kevin Tang[2]

Yale University[1], Heinrich-Heine University[2]

Yale

hhu Heinrich Heine Universität Düsseldorf

Formal tools for sub-phonemic patterns

- Our formalisms tend to privilege discrete symbolic units, as many phonological patterns are insightfully described in these terms.

- Some phenomena are more challenging for this level of description (or just fall outside the scope):
  - Incomplete neutralization
  - Gradual sound change
  - Sub-phonemic change in representations over a lifetime

- In this talk, we explore the potential of Dynamic Neural Fields for capturing sub-phonemic patterns.

Empirical phenomenon: “Leaky Prosody”

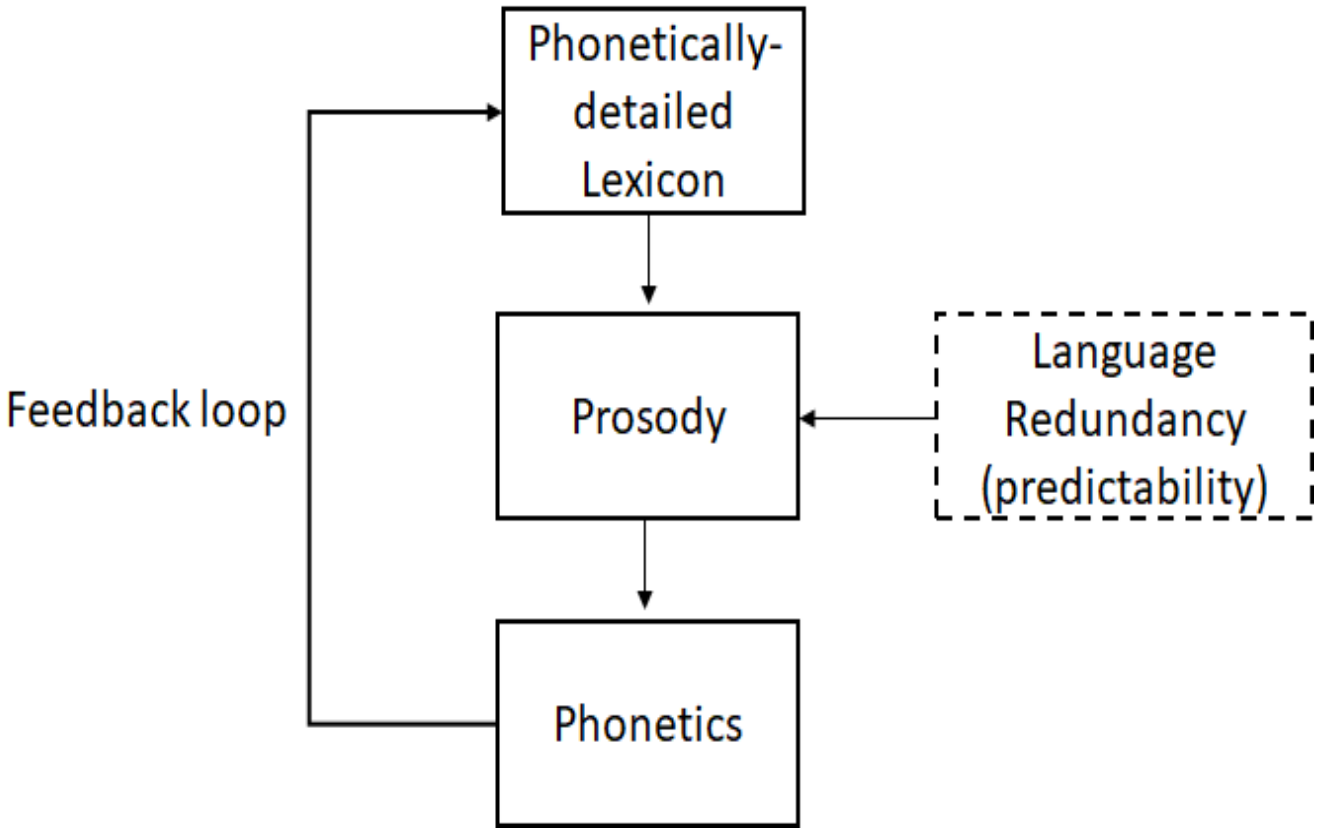
- Lexical items come to take on the phonetic characteristics of the prosodic environments in which they are typically produced (e.g., Seyfarth 2014; Sóskuthy & Hay 2017; Tang & Shaw 2021).

- In Mandarin Chinese, words that tend to attract a high degree of prosodic prominence are produced with relatively high pitch, even in prosodically weak environments; thus, **prosody** from context **leaks into the lexicon** (Tang & Shaw 2021).

- Effects are lexically specific and **sub-phonemic** synchronically but may provide seeds for gradual diachronic change.
  - Frequency/informativity effect on segment count (Zipf 1949; Piantadosi et al. 2012) may derive from frequency/informativity effect on ms duration (Wright 1970; Seyfarth 2014).
  - lexical tone/stress emerging from higher level prosodic prominence/intonation
  - Lexical tone loss in predictable environments.

Architectural sketch (Tang & Shaw 2021)

- Leaky prosody facts suggest that phonetic outputs may feedback into the lexicon.
- *Possibly* imperfect (incomplete) compensation for effect of prosodic environment on phonetic realization.

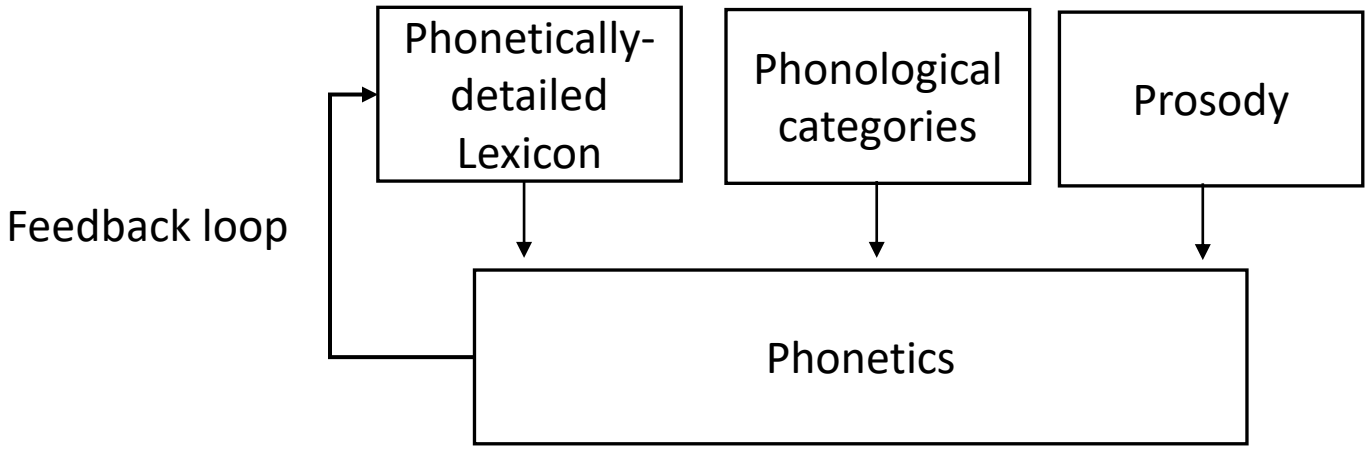


Tang, K., & Shaw, J. A. (2021). Prosody leaks into the memories of words. *Cognition*, *210*, 104601.

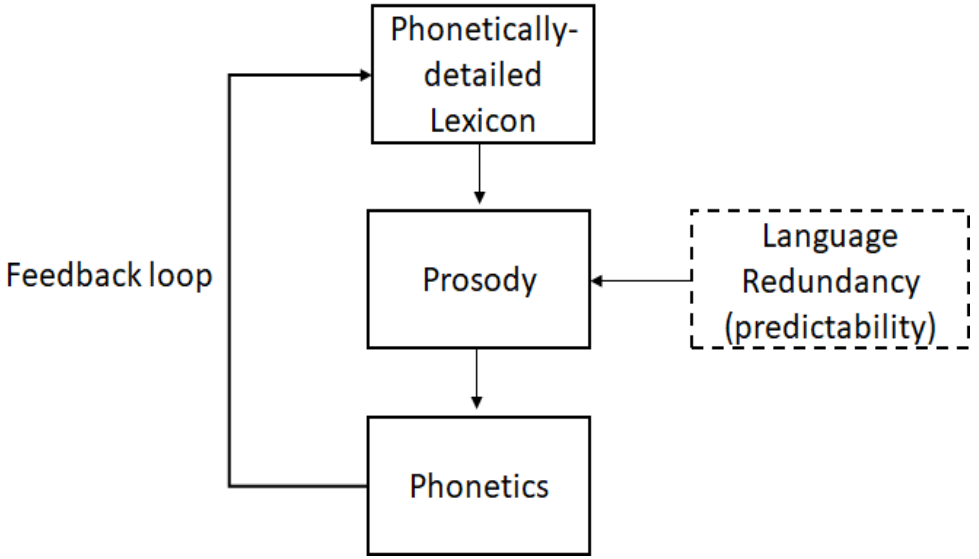
c.f., Turk & Shattuck-Hufnagel (2020)

Today: alternative “flat model”

Today’s flat model



Tang & Shaw (2021) proposal

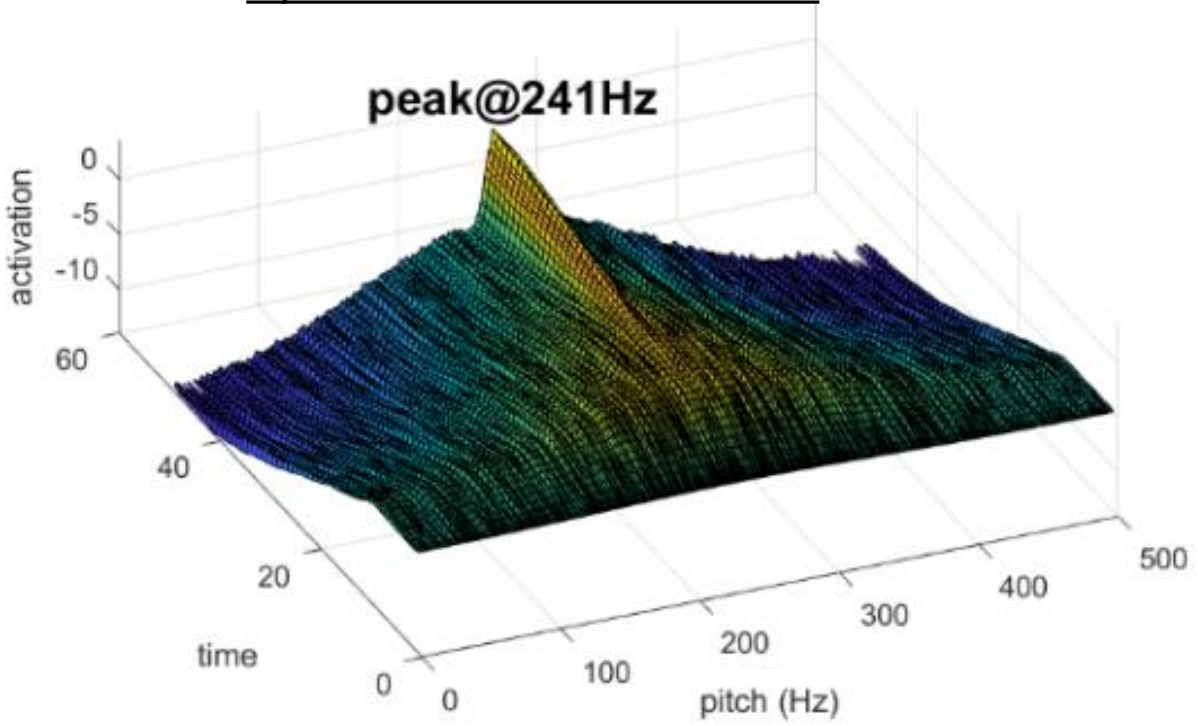


- Potential advantage in learning surface distributions (distributional learning) vs. transformational rule (multi-factor regression/highly parameterized generative model)

Framework: Dynamic Field Theory (Schöner & Spencer 2016)

- Cognitive representations are **continuous** parameters (here, pitch) governed by populations of neurons.
- The distribution of **activation** across a neural population is represented by a dynamic neural field (DNF).
- Activation at each field location **evolves over time** under the influence of inputs until the system **stabilizes**

Dynamic Neural Field of Pitch

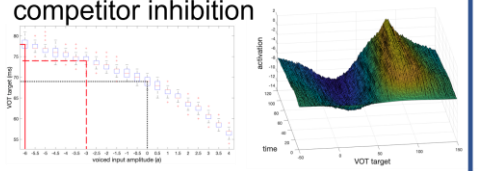


Schöner, G., & Spencer, J. P. (2016). *Dynamic thinking: A primer on dynamic field theory*. Oxford University Press.

DFT: key properties for a flat model of leaky prosody

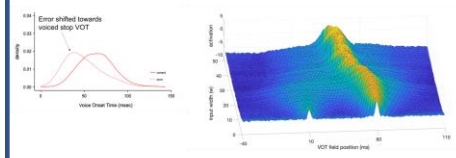
- Multiple inputs to a field can exert influence on stabilization.
- Perception/production modelled as time varying processes, c.f., purely statistical agent-based models (c.f., Harrington & Schiel 2017).
- Nested time scales: Learning occurs token-by-token (slow time scale) in response to production & perception (fast time scale).

Derives contrastive hyperarticulation of VOT from competitor inhibition



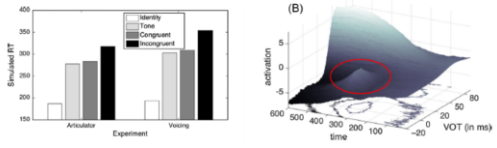
Stern, M. C., & Shaw, J. A. (2022). Neural inhibition during speech planning contributes to contrastive hyperarticulation. *arXiv preprint arXiv:2209.12278*.

Derives trace effect in speech errors from competitor excitation



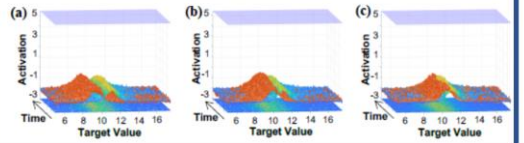
Stern, M. C., Chaturvedi, M., & Shaw, J. A. (2022). A dynamic neural field model of phonetic trace effects in speech errors. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44, No. 44).

Derives effect of perception on production in cue-distractor task



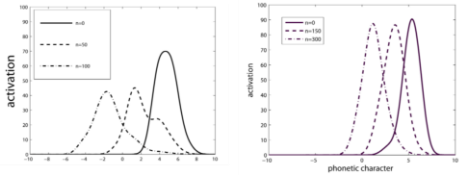
Roon, K. D., & Gafos, A. I. (2016). Perceiving while producing: Modeling the dynamics of phonological planning. *Journal of Memory and Language*, *89*, 222-243.

Derives effect of production variation on perceptual sensitivity



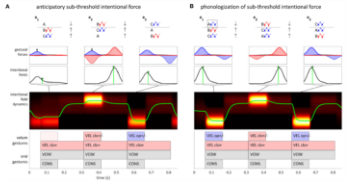
Harper, S. K. (2021). *Individual Differences in Phonetic Variability and Phonological Representation* (Doctoral dissertation, University of Southern California).

Derives effect of word frequency on lenition



Gafos, A., & Kirov, C. (2009). A dynamical model of change in phonological representations: The case of lenition. *Phonological systems and complex adaptive systems: Phonology and complexity*, 219-240.

Derives phonetic precursor to long-distance (dis-)harmony

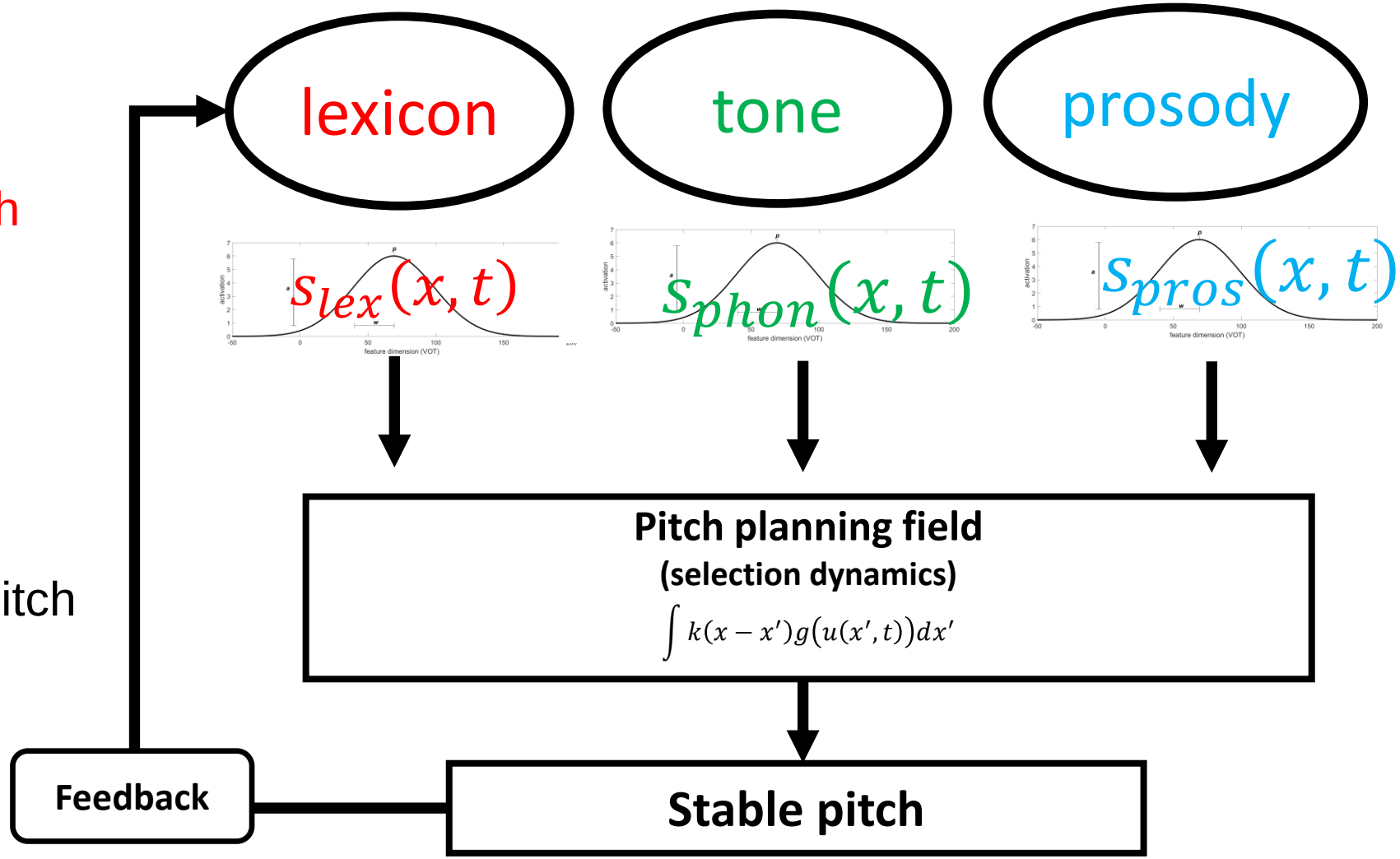


Tilsen, S. (2019). Motoric mechanisms for the emergence of non-local phonological patterns. *Frontiers in Psychology*, *10*, 2143.

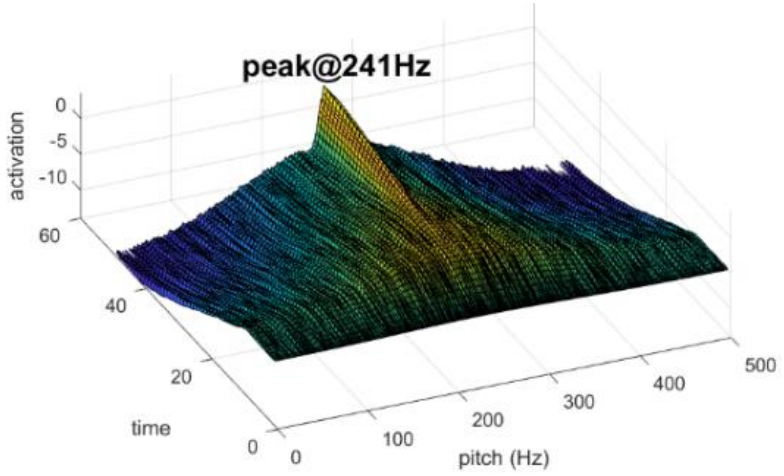
Model overview: pitch target

Pitch input from three sources:

- **lexicon**: lexical pitch target
- **tone**: phonological pitch target
- **prosody**: prosodic pitch target
- Lexicon updated to incorporate stable pitch



Formal expression of the model



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

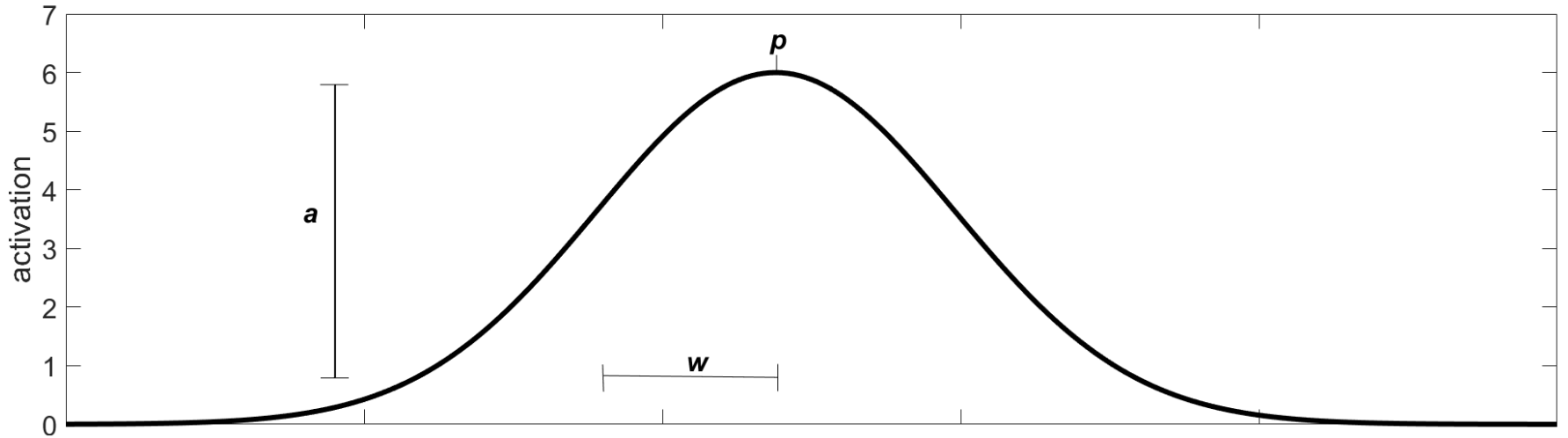
DNF parameter
Resting activation: h = -5
Field evolution speed: τ = 20

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

Formal expression: gaussian inputs

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

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



Simulation inputs are surface distributions

Input parameters based on Tang & Shaw (2021) corpus of 1,655 Mandarin speakers.

- Starting Lexical input = sample of high tone distribution (1/500th)
- Phonological pitch target = high tone distribution (~41,000)
- Prosodic context = distribution of pitch values at two levels of bigram suprisal
  - **Low** predictability (~10,000)
  - **High** predictability (~10,000)

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

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

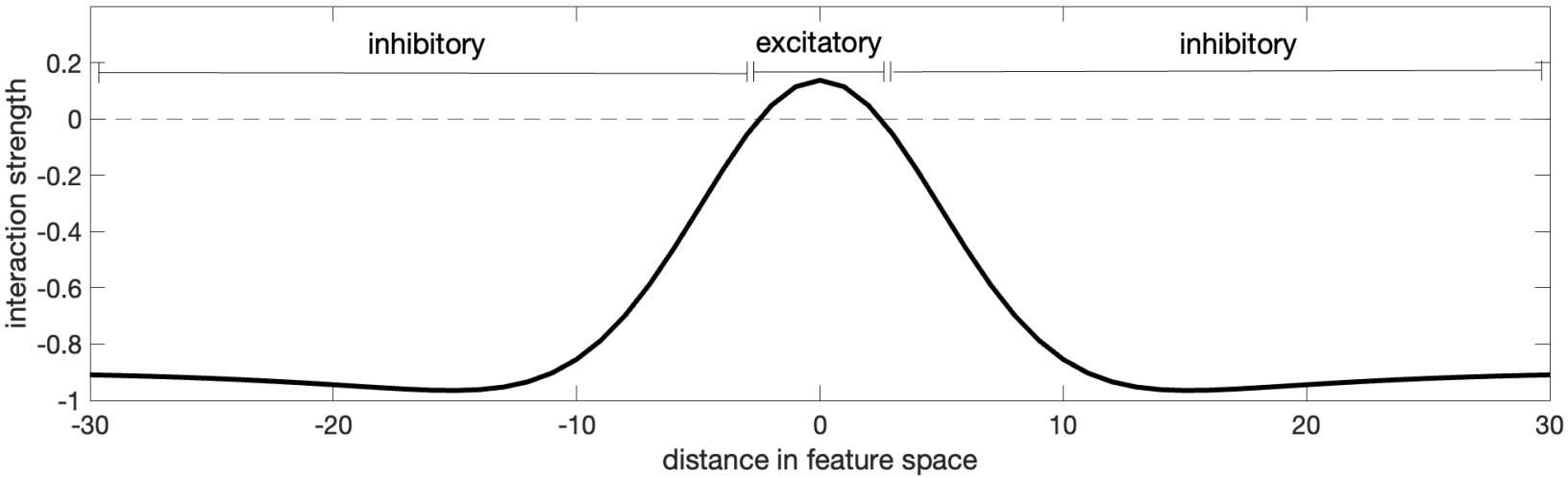
Tang, K., & Shaw, J. A. (2021). Prosody leaks into the memories of words. *Cognition*, *210*, 104601.

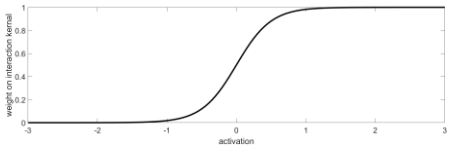
Formal expression: interaction kernel

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

Parameters chosen to ensure **selection dynamics**

Sigmoidal gate





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

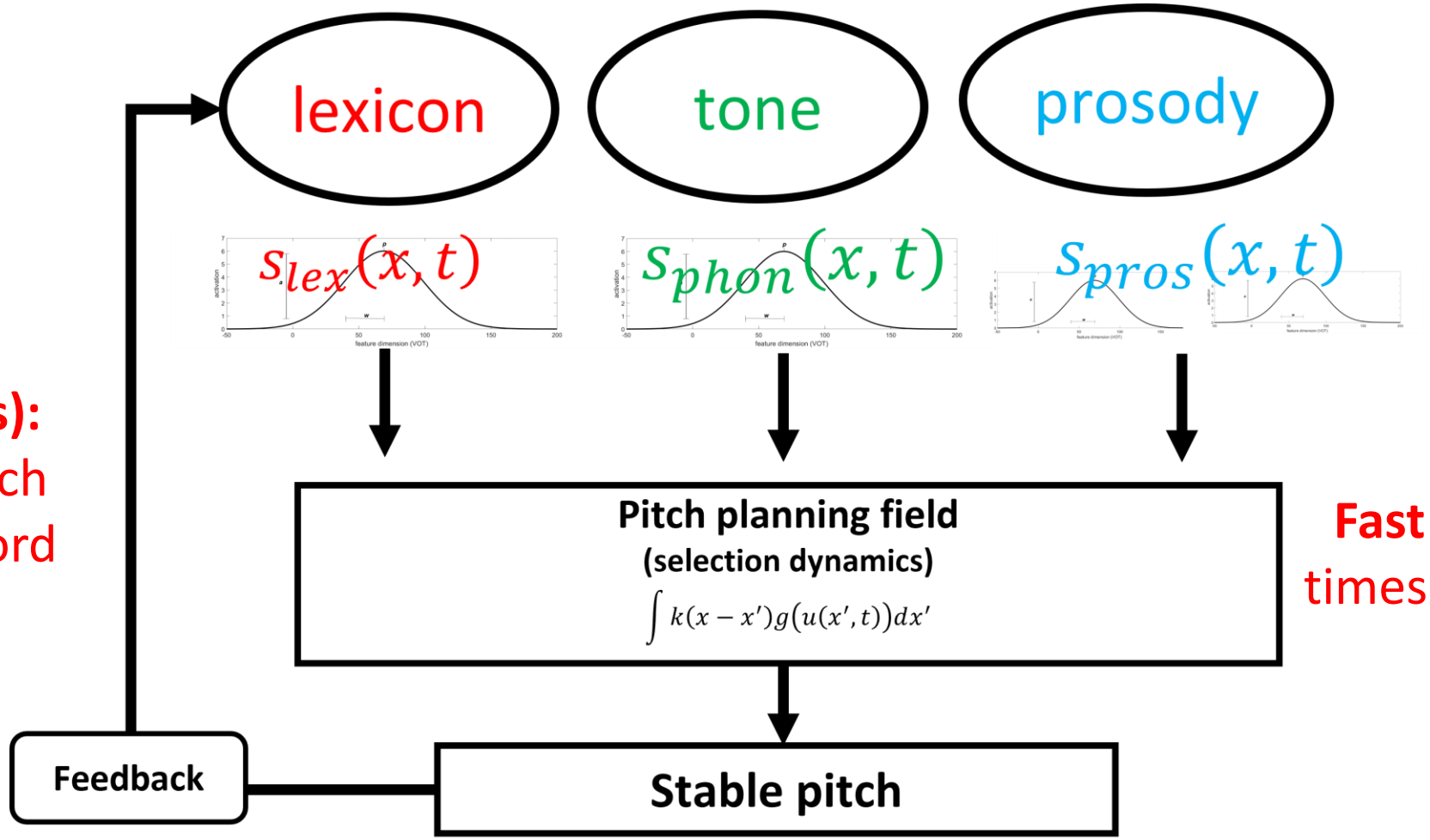
DNF parameters
$c_{exc} = 15$
$c_{inh} = 5$
$c_{glob} = 0.9$
$\sigma_{exc} = 5$
$\sigma_{inh} = 12.5$

Simulations

1. **Speech production planning as a time varying process (fast time scale)**: establish effect of prosodic context on pitch target
   - Initialize two words with identical pitch targets
   - Simulate one in a high prominence environment; one in a low prominence environment.

2. **Lexical learning as a time varying process (slow time scale)**: derive leaky prosody from updating the lexicon
   - update lexical representations based on where the field stabilizes on each fast time scale simulation

Nested timescales

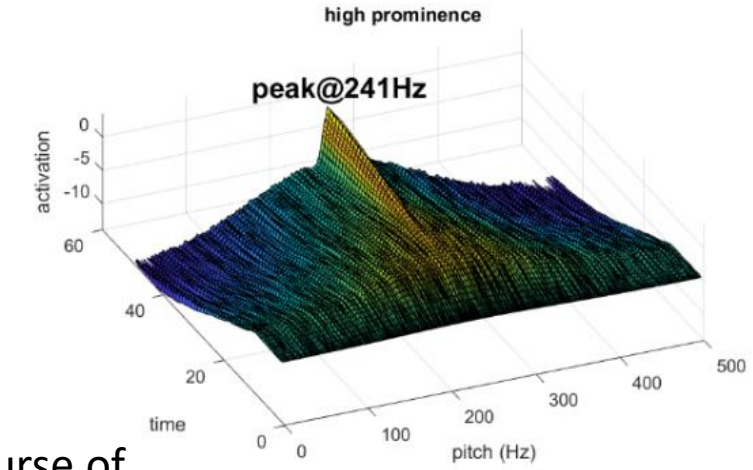


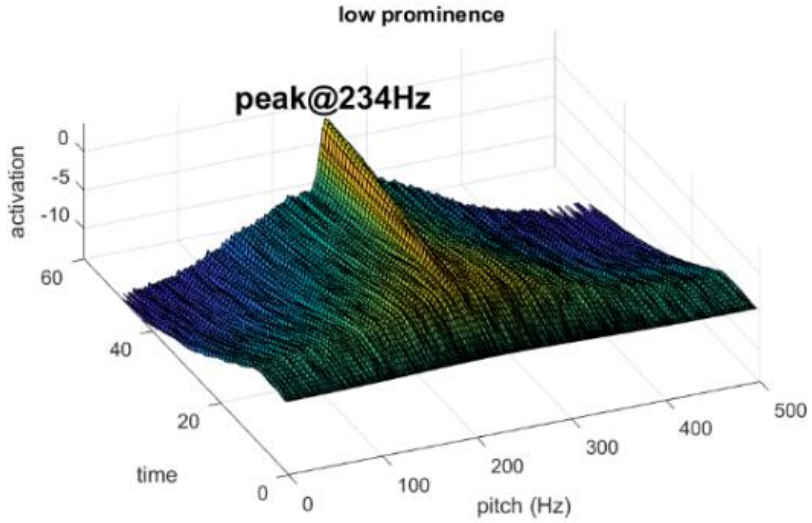
Slow (e.g. hours): learning with each exposure to a word

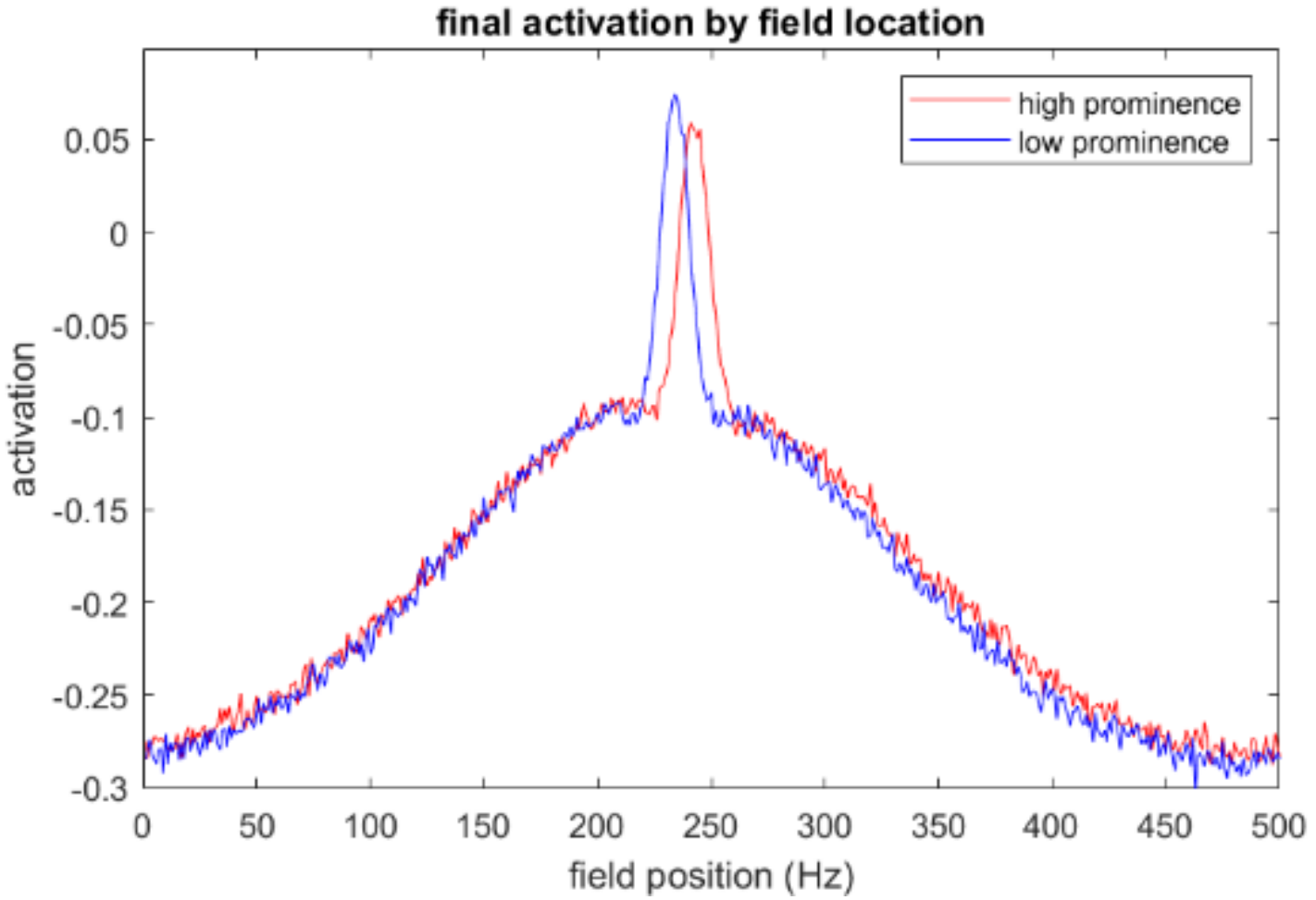
Fast (e.g., msec): timescale of speech planning

Fast time scale: single trial, high vs. low prominence

Time course of speech planning

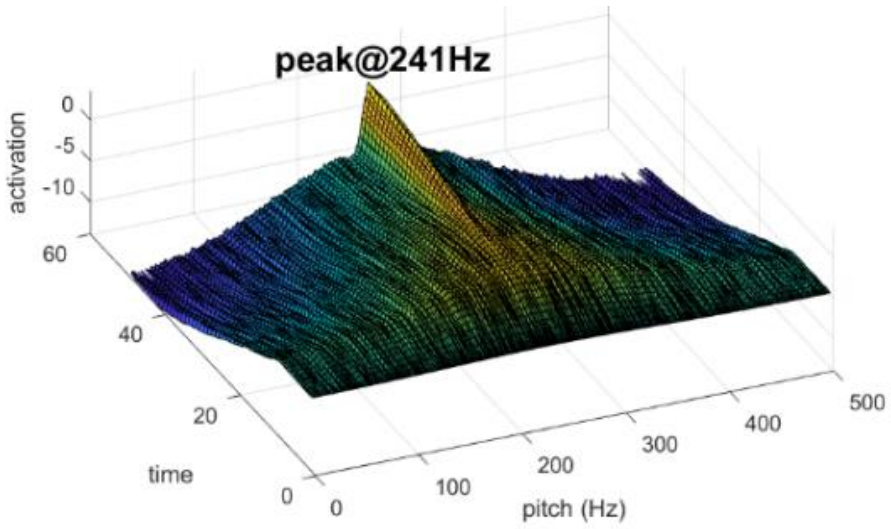


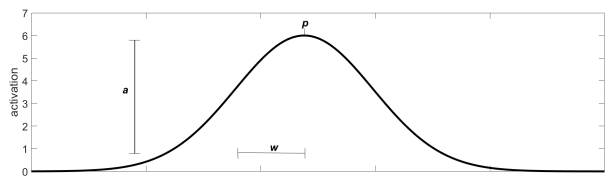




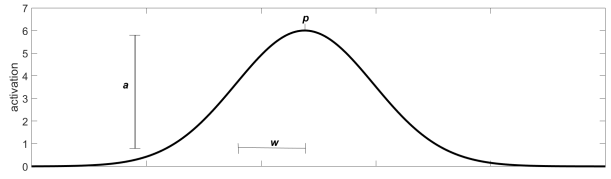
Updating the lexical input from single trial

Stable pitch
(single trial output)

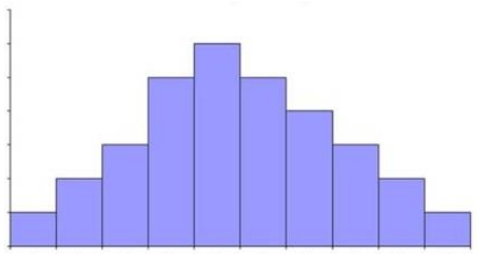




Lexical input



Updated lexical input

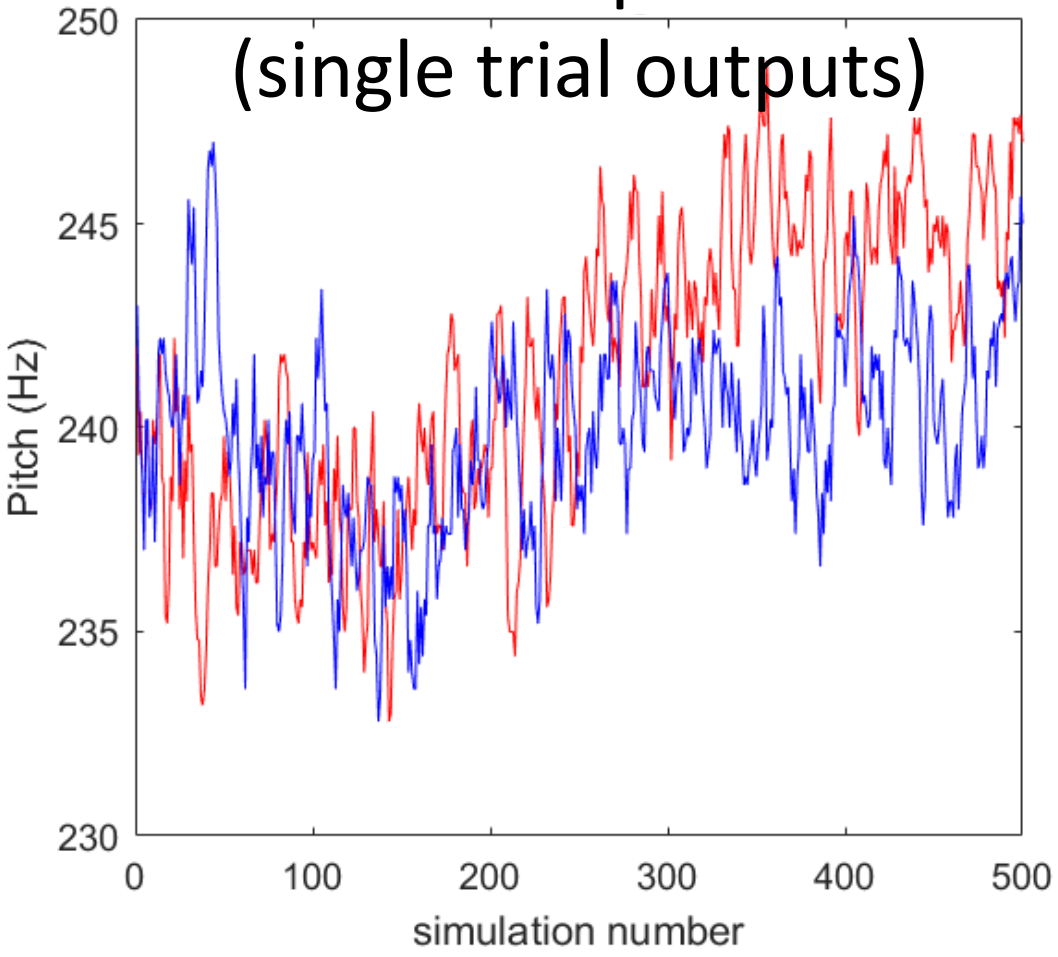


Samples from lexical input distribution; one sample is randomly selected and replaced with the new pitch value.

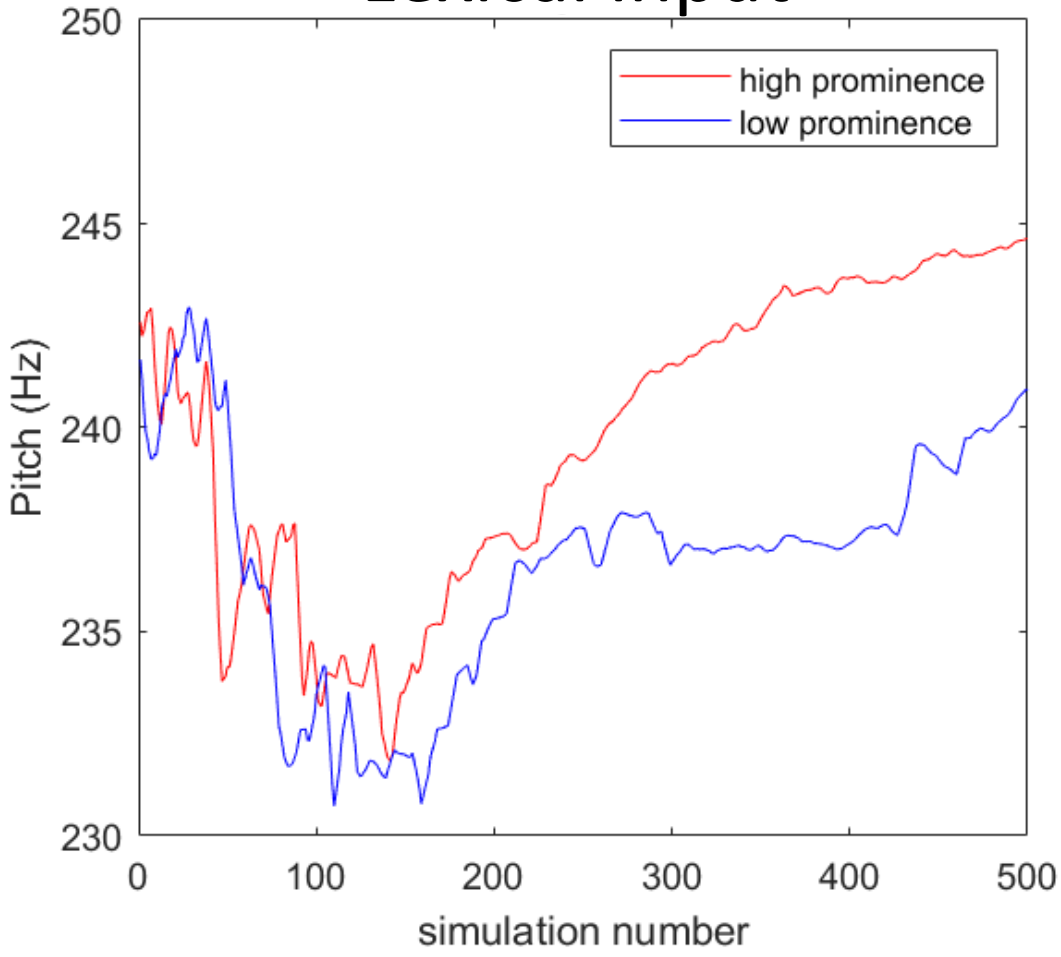
Slow time scale: lexical drift over 500 trials

Stable pitch

(single trial outputs)



Lexical input



Discussion: achievements

- Leaky prosody effect derived from simple assumptions
  - A1: production inputs come from **surface distributions**
    - **Lexical target**: sample of distribution of f0 for high tone category
    - **Phonological tone**: complete distribution of f0 for high tone category
    - **Prosodic context**: distribution of f0 at a given level of surprisal
  - A2: inputs jointly influence pitch target
  - A3: flat model → stabilization instead of transformations

- Trial-by-trial variability
- Small lexical differentiation emerges over time from learning

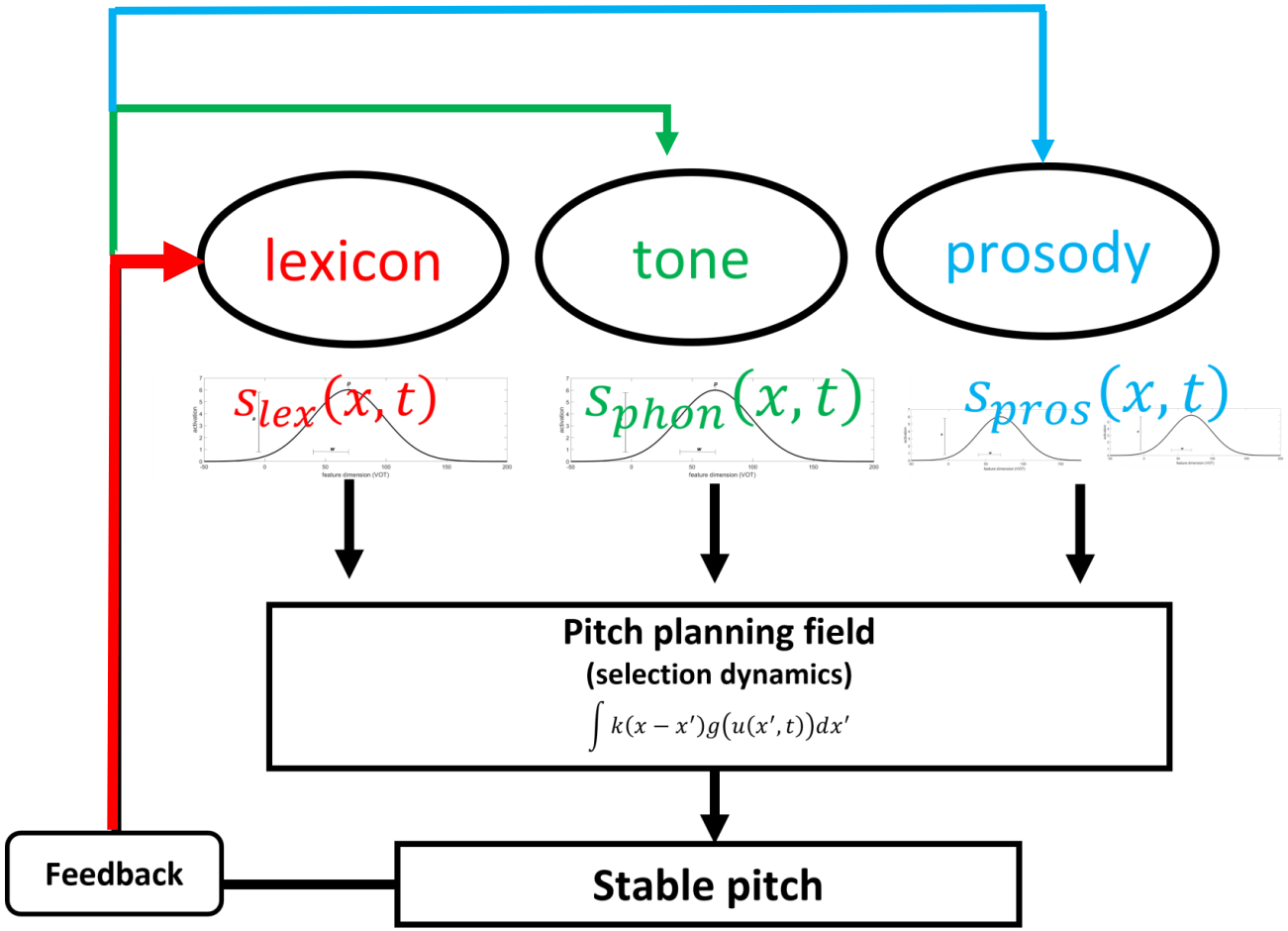
Discussion: limitations

- Just one tone (high)
- Just two lexical items
- Just one feature dimension (pitch)
- No talker normalization (flat model)
- No signal transformations (ERB, MEL)

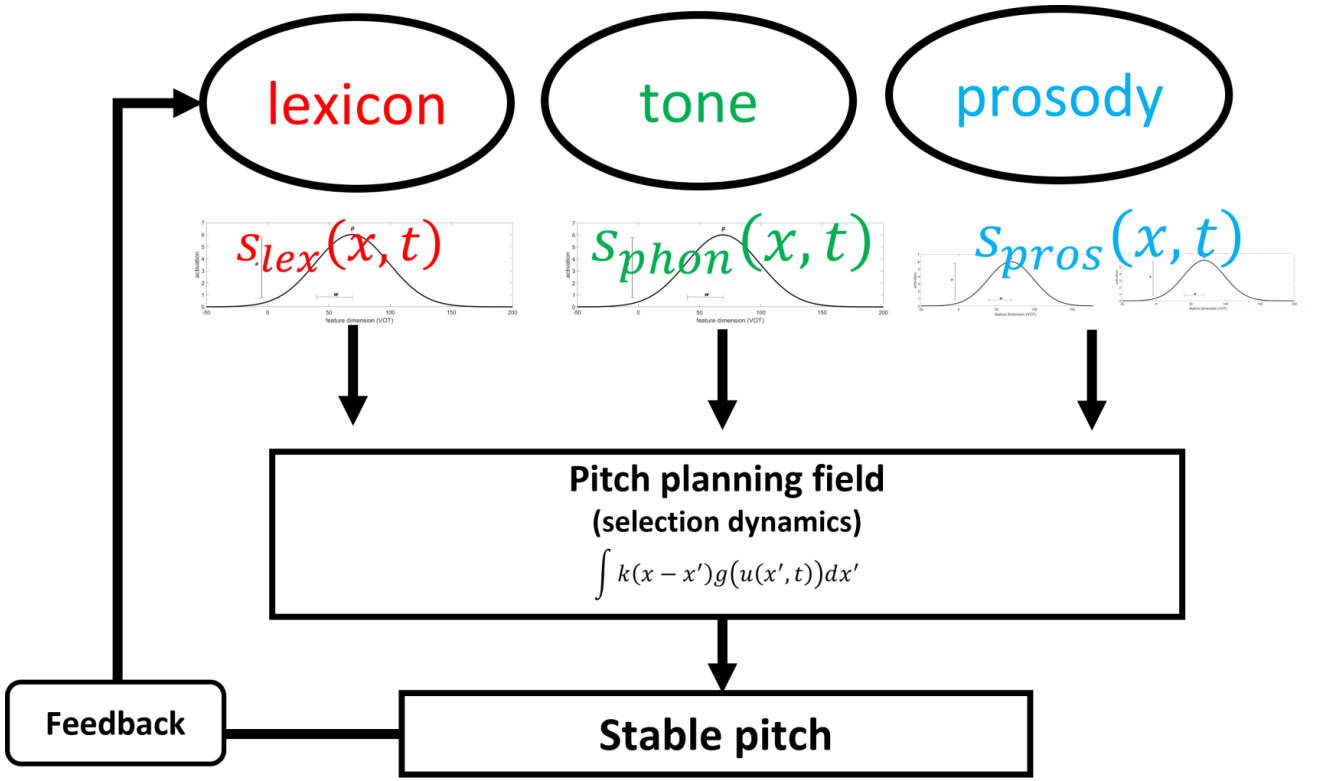
Discussion: parameter space

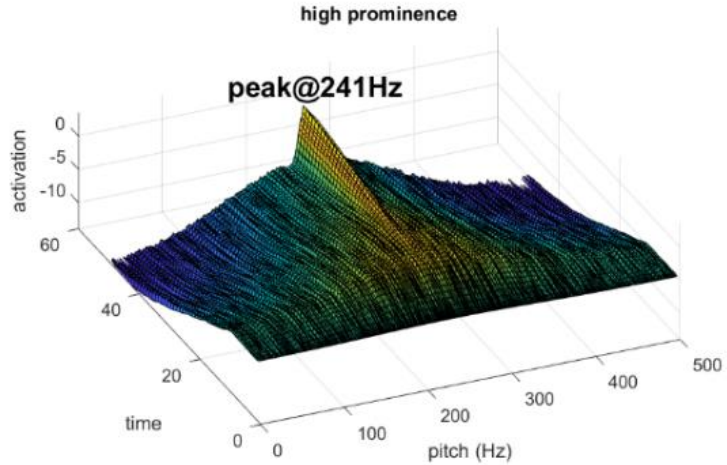
- Only lexical inputs (not phonological/prosodic) updated
  - **Stable phonological input** works against lexical drift.
  - Should persist even if phonological representations are updated…
  - Unless enough words shift in the same direction

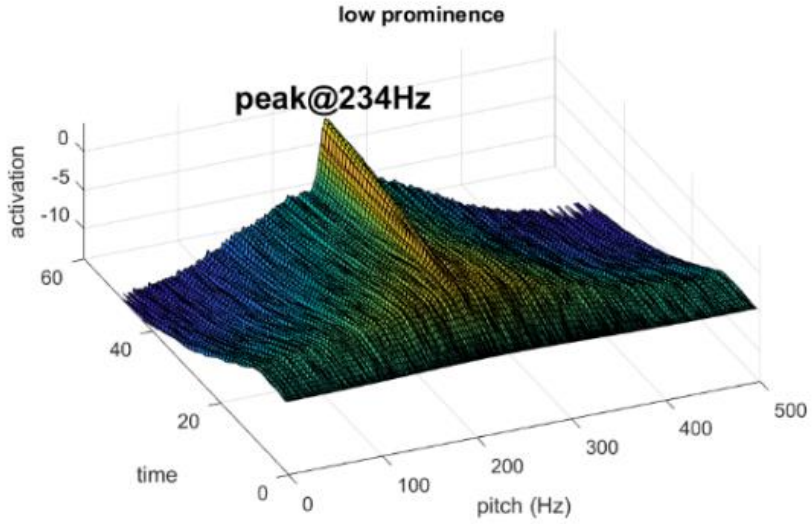
- Amplitude of inputs the same (> *h* 'rest level') for lexical, phonological, prosodic targets
  - Having lexical, phonological, and prosodic inputs leads to ***faster stabilization***.
  - Predicts we should be able to have a pitch target with just one input.



Thank you!







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