A neural process model of compensation and adaptation as independent mechanisms in speech motor control

Manasvi Chaturvedi (manasvi.chaturvedi@yale.edu)[1] **& Jason A. Shaw**[1]
[1]Department of Linguistics, Yale University

Abstract

Experimental paradigms that perturb auditory feedback induce two distinct but related behavioral responses: (1) compensation, a real-time, within-token, adjustment to speech, and (2) adaptation, an adjustment to speech that persists even after the removal of perturbation, evidencing learning (e.g., Houde & Jordan, 1998; Purcell & Munhall, 2006a, 2006b; Tourville et al., 2008). In this paper, we present a fully dynamic model of one-shot adaptation, i.e., learning after one trial of perturbation (Hantzsch et al., 2022), in the framework of Dynamic Field Theory (DFT; Schöner & Spencer, 2016). The model triggers compensation and adaptation via prediction-perception mismatch driven by lexical-item-based predictions. Compensation derives from resolving multiple inputs into a neural field and via update to the mapping between phoneme representations and production targets, while adaptation relies only on the latter. We discuss how the model, which integrates speech motor control with linguistic representations, accounts for established results while generating novel predictions.

Keywords: speech production model; adaptation; implicit learning; feedback; neural process model

Introduction

Speech production presents a special case of motor control. While common behaviors in motor control, such as compensation and adaptation to perturbed feedback, are also found in speech, they are modulated by the connection that speech has to language (e.g. Bourguignon et al., 2014; Lametti et al., 2014; Mitsuya et al., 2011; Niziolek & Guenther, 2013; Schuerman et al., 2017; Shiller et al., 2009). The need for speech motor control to be integrated more fully into language production has been recognized (Hickok, 2012, 2014; Hickok et al., 2011), but most models of compensation and adaptation in speech are situated in the motor domain and none are implemented in a fully dynamic framework. In this paper, we fill this gap, developing a fully dynamic neural process model of compensation and adaptation that builds on aspects of several existing proposals, including integration of linguistic representations. In doing so, we recognize speech as a motor behavior while also capturing a connection to language absent in other motor behaviors.

When speech is perturbed, speakers correct for this outside perturbation in real-time, and learn a new way of talking in response to the perturbation, a form of implicit learning. Speech perturbation can be applied in two ways – to auditory feedback, such as to vowel formants or pitch (e.g., Houde & Jordan, 1998; Jones & Munhall, 2000), or to somatosensory feedback, such as the location of an articulator like the jaw (e.g., Tremblay et al., 2003). In this paper, we focus on auditory perturbation. The perturbed auditory paradigm leads to two distinct, but related behavioral responses – compensation, a within-token articulatory adjustment (e.g., Purcell & Munhall, 2006b; Tourville et al., 2008), and adaptation, a lasting change in articulation that persists beyond the trial-level perturbation (e.g., Hantzsch et al., 2022; Houde & Jordan, 1998).

Both compensation and adaptation have been shown to be influenced by higher-level linguistic representations (e.g., Bourguignon et al., 2014; Mitsuya et al., 2011; Niziolek & Guenther, 2013). For example, compensation is faster and earlier when perturbation crosses a vowel phoneme boundary (Niziolek & Guenther, 2013), and adaptation is greater when perturbation changes a non-word to a word than when it changes a word to a non-word (Bourguignon et al., 2014).

We present a dynamical systems account of compensation and adaptation behavior with inspiration from Hickok (2012, 2014) and Hickok et al. (2011), who argue for the integration of speech models into a larger linguistic architecture. Specifically, we incorporate two kinds of higher-level linguistic representations: an auditory expectation based on a lexical item representation and a motor command based on a phoneme representation associated with the given lexical item.

Based on this model, we present proof-of-concept simulations that demonstrate that such a model can derive one-shot adaptation, i.e., learning after a single instance of perturbed feedback. Our simulations are informed by the data presented in Hantzsch et al. (2022), who analyzed perturbed auditory feedback across six studies of formant perturbation.

Our model features two *dynamic* mechanisms: 1) inhibition of an on-going production target, and 2) changes in the production target of the phoneme representation. These two mechanisms have unique time courses. Only the second mechanism results in learning, which persists beyond the utterance. Both mechanisms contribute to compensation while only the second drives adaptation.

We implement our model in Dynamic Field Theory (DFT; Schöner & Spencer, 2016). As a fully dynamic model, compensation and adaptation occur at each time step. The model combines insights from different strands of the speech production literature, while implementing these insights in a neural process model that is situated within a broader language ar-

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

chitecture.

We will begin with an overview of DFT, followed by a description of the model architecture. We then report the results and discuss model predictions and avenues for further work.

Dynamic Field Theory

Dynamic field theory (DFT) is a framework to model perception and action behavior via the dynamics of neural fields (Schöner & Spencer, 2016). DFT has previously been applied to different aspects of speech, including perception-production links (Jenkins & Tupper, 2016; Roon & Gafos, 2016), lexical competitor effects (Stern & Shaw, 2023; Stern et al., 2022), bilingualism (Pintado-Urbanc, 2025), and phonological patterns (Shaw, 2025, 2026) (see Stern, 2025, for a review).

In DFT, activation over neural fields evolves according to the equation in (1). By relating the rate of change of activation, $\dot{u}(x,t)$ to the current activation, $u(x,t)$, the equation defines a *point attractor* system, where activation converges to a particular value under the influence of the resting level, h, external input(s), *s(x,t)*, and noise, $q\xi(x,t)$.

$$\tau\dot{u}(x,t) = -u(x,t)+h+s(x,t)+\int k(x-x')g(u(x',t))dx'+q\xi(x,t) \quad (1)$$

Inputs are modeled as Gaussian distributions according to the equation in 2, where a is the amplitude of the distribution, p is the center and w is the width.

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

Activation peaks arise as a result of instabilities under the influence of the interaction kernel in (3), which defines how neurons in a field interact. Generally, this kernel is defined such that neurons close together excite each other, and neurons farther away inhibit each other, reflecting a *stabilization dynamics.* Such dynamics arise because of the c/σ_{exh}, c/σ_{inh}, and c_{glob} parameters. These parameters dictate how each neuron x affects every other neuron, x', via excitatory or inhibitory Gaussian activation distributions. c sets the magnitude for excitatory and inhibitory within-field interactions, while σ sets the width. c_{glob} refers to a global inhibitory contribution by every above threshold neuron. This is because the kernel (in (3)) is sigmoidally gated by ($g(u(x',t))$; (1)) – neurons are only able to influence one another if they cross an activation threshold, set to 0. The resting level h is generally set to -5, meaning neurons at rest are inhibited and cannot influence each other.

$$k(x-x') = \frac{c_{exc}}{\sqrt{2\pi}\sigma_{exh}}\exp[-\frac{(x-x')^2}{2\sigma_{exh}^2}]-\frac{c_{inh}}{\sqrt{2\pi}\sigma_{inh}}\exp[-\frac{(x-x')^2}{2\sigma_{inh}^2}]-c_{glob} \quad (3)$$

By modeling the dynamics of *fields*, DFT privileges activation evolution over neuronal populations rather than single neurons. This choice is grounded in neurophysiological findings showing that behavior reflects the activity of neuronal populations (e.g. Georgopoulos et al., 1986; Lee et al., 1988). Since fields are defined over continuous dimensions but result in discrete states (activation peaks), DFT allows us to capture variation in continuous space, while still relating it to stable categories (see Stern, 2025, for a discussion about this point). For example, previous models have used inhibitory inputs into fields defined over production dimensions to model *hyperarticulation* effects resulting from lexical competition – a shift in continuous phonetic measures without a resulting change in category (Stern & Shaw, 2023). Additionally, the space of fields allows the modeling of inputs as Gaussian distributions, capturing the representation of action targets as ranges or spaces rather than single values, similar to the convex target regions in DIVA (Guenther, 2016). With noise, this captures the natural variation seen in speaker's productions. The structured variability of continuous dimensions of speech is a natural consequence of modeling via neural fields.

DFT architectures are well suited to learning, allowing us to model changes in parameters *over time* in response to neural activation dynamics. Further, similar to how neurons in one field can influence each other, fields can also be coupled together, driving synchronous changes in multiple perception and action dimensions. Such coupling can be inhibitory or excitatory. Finally, fields can be coupled to *nodes*, representing populations of neurons that are tuned to discrete categories rather than continuous values. This choice also has neural grounding, reflecting monosemanticity, such as the activation of neurons in the superior temporal gyrus that respond categorically to phonetic-acoustic features for consonants and vowels (Chang et al., 2010; Mesgarani et al., 2014; Oganian et al., 2023). Note, however, that the neural populations in these studies are still sensitive to continuous values (such as voice onset time (VOT), or the first and second formant of vowels). This pattern can be achieved via coupling between nodes and fields (modeled as Gaussian inputs of the form in 2) – a property the current model leverages to model learning, as we will discuss below.

In summary, DFT models perception and activation behavior as a result of activation dynamics in continuous *fields*. The properties of these fields have important consequences for representing behavior. DFT has already been applied to various aspects of language production and perception, including meaning and conceptual structure (Kati et al., 2024; Sabinasz et al., 2024; Stern & Piñango, 2026). This provides the opportunity to build larger architectures, encompassing the different facets of language, and going beyond traditional subfield boundaries (in the vein of Hickok (2012, 2014) and Hickok et al. (2011)).

Model Specifications

In this section, we provide an overview of the architecture and describe the choice of certain parameter values. The model was built using COSIVINA (Schneegans, 2012), with a few modifications. The complete details, including new COSIVINA elements and code for simulations, are available

on OSF: https://osf.io/y6qnt

The model consists of 5 main components: 1) Lexical item (LI) and phoneme nodes, 2) Production fields defined over articulator goals, such as constriction degree (CD) and constriction location (CL) (Browman & Goldstein, 1989)[1], 3) Perception fields defined over acoustic dimensions, such as F1 (Percept), 4) A prediction field that encodes acoustic expectations for words, i.e, the 'Lexical item efference copy' (LIEC), and 5) A Contrast field that encodes *error* as a result of mismatch between prediction (LIEC) and Percept (i.e., a difference in activations in the two fields).

The architecture involves multiple levels of coupling. Figure 1 provides a schematic, with green arrows representing excitatory coupling and red arrows representing inhibitory coupling. Activation in the LI nodes flows to the phoneme level before being sent to the production field. At the same time, the LI nodes send input into the LIEC (expectation) field, representing one's expectation of the word based on prior experience. The Percept field, on the other hand, is coupled to the phoneme nodes, facilitating speech perception. Finally, the Contrast field is coupled inhibitorily to the production field, representing suppression of movements perceived as 'wrong,' i.e., those that induce activation in the Contrast DNF via Percept-LIEC comparison.

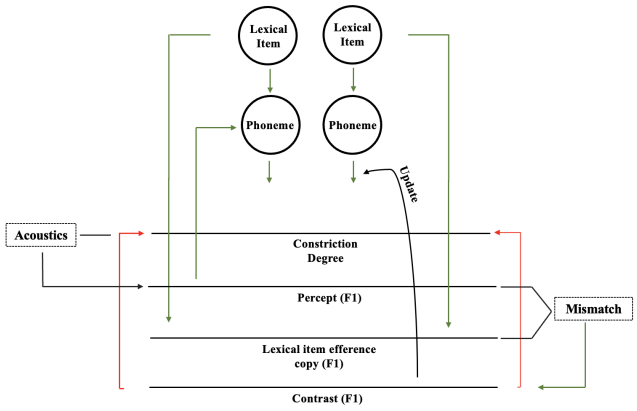

Figure 1: Schematic of DNF model.

Finally, a mismatch between prediction and percept leads to an *update* in the phoneme node to production field coupling, representing a change in the target CD value for that phoneme. In the current model, this is done via a simple error-based learning equation which updates the p parameter of this coupling relation, representing the location of the input into the field ($p_{phoneme}$):

$$p_{phoneme} = p_{phoneme} - \eta * (p_{percept} - p_{expectation}) \quad (4)$$

[1]For one-shot adaptation, we will focus on the former, since the studies in (Hantzsch et al., 2022) perturbed the first formant (F1), which is related to this articulatory dimension.

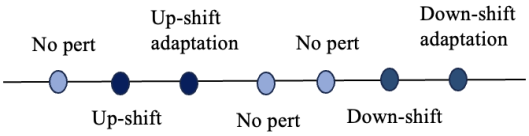

Figure 2: Simulation trial structure.

Equation 4 computes an update to the $p_{phoneme}$ parameter based on the difference between the current percept (the location of maximum activation in the Percept field), and the expectation (the position of the lexical item to LIEC coupling relation). This difference is multiplied by the learning rate, η (set to 0.0009), before being subtracted from the current $p_{phoneme}$. As we will see, this forms the basis for adaptation in the presence of altered feedback.

A key characteristic of the model is that adaptation derives from adjusting the coupling between the phoneme node, a categorical linguistic representation, and an articulatory goal (CD). Notably, this update is driven by the comparison of percept to expectation, and is independent from the inhibitory coupling of the Contrast field to the production field. Thus, adaptation is mediated by the linguistic representation–a phoneme node coupled to the production field–whereas compensation also involves direct influence of the articulatory goal, via field-to-field coupling. Both, however, arise as a result of a mismatch between expectation (dictated by the lexical item), and percept.

Trial Structure

In order to model one shot learning, we stay close to the trial structure of the studies in Hantzsch et al. (2022). In order to capture *one-shot* adaptation, perturbation has to be inconsistent in direction, and sandwiched between trials where there is no perturbation (Hantzsch et al., 2022, p. 2). The results reported here are based on the trial structure in Figure 2, with the perturbed trials forming around 30% of the trials, similar to the numbers across studies reported in Hantzsch et al. (2022) (average of 40.2% across 6 studies).

To assess compensation and adaptation, we calculated normalized compensation and adaptation trajectories in response to a perturbation of 120Hz, following Hantzsch et al. (2022). We used the location of maximum activation in the production field at each timestep to calculate production trajectories. This was done for every trial in Figure 2, across simulations.

First, baseline production trajectories were calculated by averaging the trajectories of no perturbation trials which were preceded by no perturbation trials (no learning expected). This mean value was then subtracted from each up and down-shift trajectory (from trials in which feedback was perturbed), giving us a difference from baseline. The same subtraction was done for adaptation trials. We report the results below.

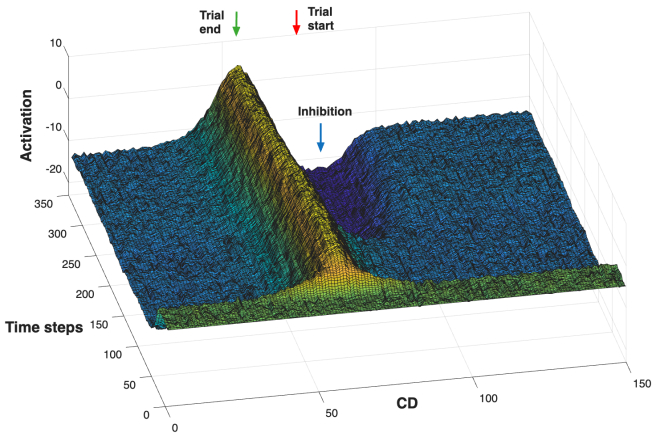

Figure 3: Production field activation over time.

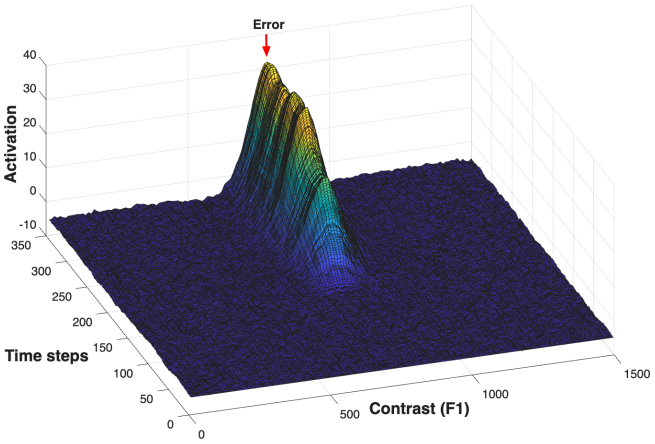

Figure 4: Contrast field activation over time.

Results

One trial

Before we discuss one-shot adaptation over trials, let us look at how online-compensation progresses within-trial. Below, we report figures from a one trial simulation of the model for **up-shift** perturbation, meaning the auditory feedback for /ɛ/ was shifted up (towards /æ/). The model should then account for perturbation by compensating in the opposite direction – a movement of CD to lower values, i.e., towards /ɪ/.

The crucial figure, Figure 3, is that of the production field over time. Changes in this field over time, representing a within-trial change in production, form the basis of testing whether the model is successfully compensating for feedback perturbation. We can see that the activation peak, shaded in yellow, moves over time under the influence of inhibition from the Contrast field, which has an activation peak at around 750Hz (Figure 4). This peak in the Contrast field represents detection of an unexpected percept – a result of Percept-LIEC mismatch. The inhibition of the 'incorrect' production via Contrast leads to a change in the production field activation peak over time ($y - axis$) in a direction that *opposes up-shift perturbation*, evidencing compensation.

Additionally, the phoneme node to field coupling is constantly being updated (not explicit in the figure) based on the error magnitude and the η parameter (equation 4), meaning that the input into the production field is updated independently of the inhibition from the Contrast field.

In summary, while the within-trial change in Figure 3 is only evident in the production field, it is a result of the workings of the full architecture in Figure 1: 1) Comparison of Percept to LIEC, leading to activation in the Contrast field, 2) Inhibition into the production field via this Contrast activation, and finally, 3) Update to phoneme node to production field coupling via equation 4.

Multiple trials

Below, we present the results of 100 simulations of the trial structure in Figure 2. Since these trajectories are a difference from baseline, a deviation from 0 evidences a change in production.

For both normalized compensation and normalized adaptation trajectories, we qualitatively reproduce the patterns in Hantzsch et al. (2022). For up-shift trials, we see a within-trial lowering of CD (towards /ɪ/) when F1 is shifted up (towards /æ/) (Figure 5), like we saw in the one trial simulation (Figure 3). We see the opposite pattern for the down-shift condition. The trajectories start out as different from 0 which is likely due to leftover learning from the shifted trials because of the repetition of the trial structure in Figure 2, leaving incomplete room for full 'unlearning' or 'deadaptation' (terms from K. S. Kim et al., 2023; Purcell & Munhall, 2006a).

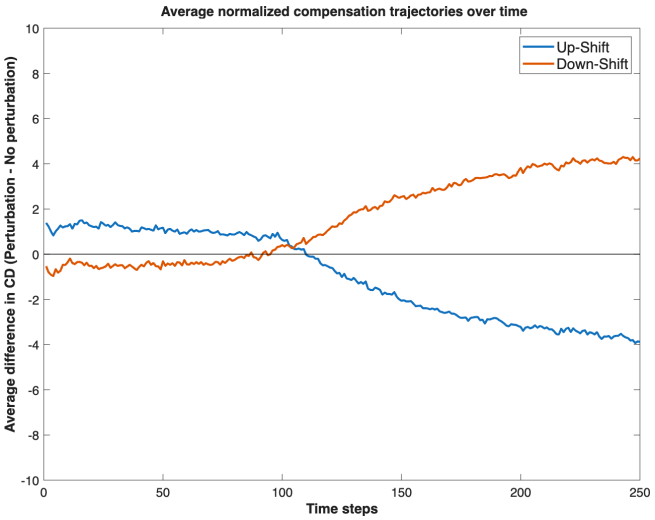

Figure 5: Normalized compensation trajectories over time.

For one-shot adaptation, we see a more stable line, different from 0, indicating a change in articulation from baseline under the presence of perturbation in the previous trials. For post up-shift adaptation, production changes to lower CD values, and for post-down shift trials, production changes to higher CD

values, evidencing learning as a result of expectation violation (Percept-LIEC mismatch; equation 4).

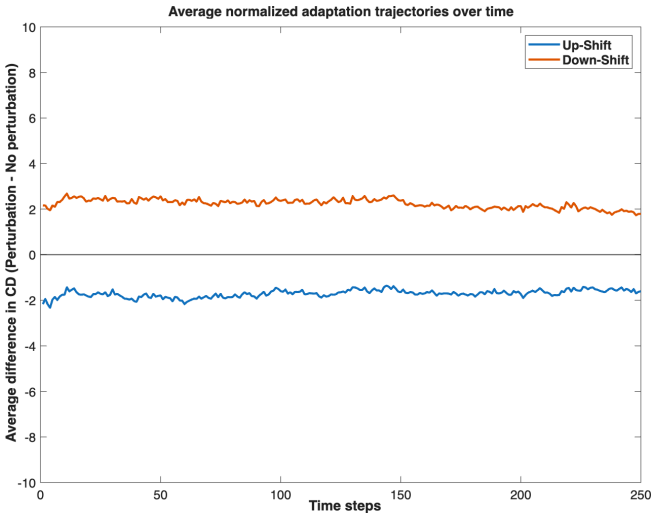

Figure 6: Normalized adaptation trajectories over time.

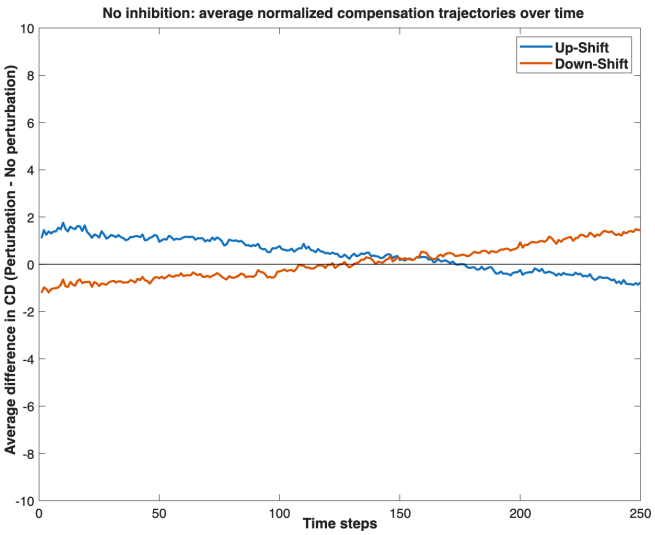

Figure 7: Normalized compensation trajectories without field inhibition (compare to Figure 5).

Discussion

We have outlined a speech production model that integrates the speech motor system with linguistic representations and is fully formalized as a dynamical system. We showed that online compensation and one-shot adaptation can be derived in this model in a trial structure with up- and down-shift perturbations that are interleaved by no perturbation trials, similar to Hantzsch et al. (2022).

In this model, two mechanisms work in tandem to achieve compensation: (1) inhibition into the production field (from field-to-field coupling with the Contrast field) and (2) updates to node to production field coupling, a type of implicit learning. Only the second of these two mechanisms contributes to adaptation. To demonstrate this, we turned off the production field's inhibitory coupling with Contrast and re-ran the trial structure in Figure 2 (100 runs). The results are shown in Figure 7. In the absence of inhibition, compensation is reduced relative to the full model, while adaptation (not shown) remains the same as in Figure 6. This behavior of the model is consistent with findings showing that adaptation occurs even in the absence of online compensation (O. A. Kim et al., 2022; Parrell et al., 2024), results also accounted for by existing models that do not rely on overt movement to derive adaptation (e.g., SFC, FACTS; Houde & Nagarajan, 2011; K. S. Kim et al., 2023).

The neural dynamic framework of the model removes the need for any explicit buffers or delays. For example, since adaptation happens at each time-step, we do not need a memory buffer to store error driven changes that are used only after the perturbation trial (K. S. Kim et al., 2023, cf. FACTS:). Further, the timing of the compensation response, which is similar to reported values (deviation from baseline around 100ms) (e.g. Tourville et al., 2008), falls out from the neural dynamics – the detection of a percept, comparison to expectation, and inhibition of the production field – without the need for built-in delays.

We integrate linguistic representations into the model in two ways. First, adaptation is modeled as changes to the coupling between a phoneme-level representation and the production field. Second, the input into the expectation field (LIEC) is dictated by another linguistic representation, the lexical item. Each lexical item (LI) is coupled to the LIEC. The parameters of this coupling relation can, in principle, vary across lexical items, predicting lexical-item-specific responses to perturbation. That is, the same perturbation applied to different words may elicit varying magnitudes and timing of compensation, depending on LI-to-LIEC coupling parameters. Here, we illustrate the impact of varying the width, w, parameter (see (2)). We simulated 100 runs of the same trial structure with 4 different LI-to-LIEC coupling widths (varying from 15 to 21, in steps of 2). Figure 8 shows the influence of this parameter on compensation magnitude. We see that as the width of LI to LIEC coupling increases, reflecting a more variable mapping, the magnitude of compensation decreases, and it occurs more slowly.

Modeling variation in compensation as LI-to-LIEC coupling width could capture, for example, differences in compensation to high vs. low frequency words. Lexical frequency effects are ubiquitous in speech processing – in shadowing tasks, for example, which are designed to test how speaker pronunciation changes in response to a model talker, lower-frequency words elicit more change (Goldinger, 1998). In our model, lower frequency words can be represented with narrower LI-to-LIEC coupling, predicting more, and earlier, within-trial compensation for these words. To our knowledge, such frequency effects are hitherto unexplored in the auditory perturbation literature. Although the mechanism is differ-

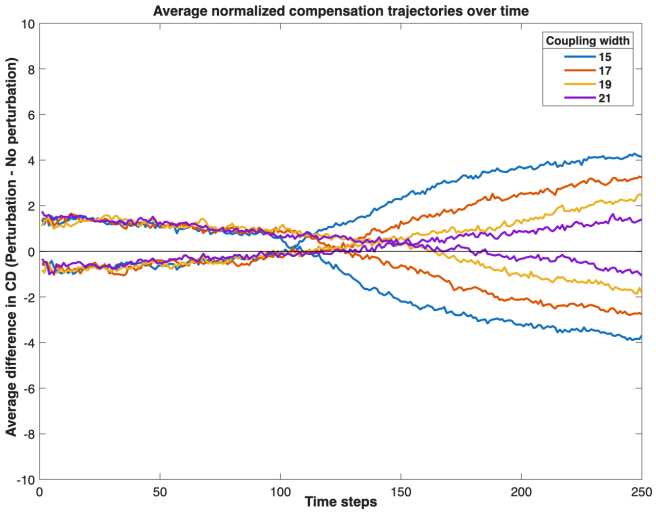

Figure 8: Normalized compensation trajectories by LI to LIEC coupling width.

ent, the intuition behind the LI-to-LIEC coupling is similar to Hickok (2014), who proposes that "familiar" words do not require much reliance on sensory targets, i.e., they can have stronger relations to the motor commands needed to produce them. A wider LI node to LIEC field coupling can be interpreted as a weaker reliance on sensory expectation, allowing motor commands to proceed relatively unaffected even in the presence of perturbations.

In the current form of the model, the predicted effects of LI-to-LIEC coupling width only apply to within-trial compensation and not adaptation. This is because, in our model, adaptation, i.e., adjusting the coupling between phoneme nodes and production fields, is based only on the position, p, of activation peaks in the Percept and LIEC fields (4). The factors that do affect adaptation are the η parameter, which may vary across individuals, and the distance between the expectation (LIEC) and Percept peak positions, without regard for their distribution width. Whether the same lexical factors that influence compensation also influence adaptation is a topic for future research that can inform this aspect of the model.

The inclusion of phoneme and LI representations is motivated by findings showing that compensation and adaptation are subject to phonological and lexical category effects (Bourguignon et al., 2014; Niziolek & Guenther, 2013). For example, perturbation crossing phoneme category boundaries leads to a faster and larger compensation response than the same magnitude of perturbation within a category (Niziolek & Guenther, 2013). In the current model, this may be implemented via the perceptual activation of a competitor phoneme representation (via the Percept field) that sends additional inhibitory input into the production field, cf. hyperarticulation effects (Stern & Shaw, 2023).

The parameters of our model may also be appropriate for capturing some dimensions of individual variation. In addition to the width of LI-to-LIEC coupling, the rate of evolution of both fields will play a role in within-trial compensation. Specifically, the parameter τ in equation 1 dictates how fast peaks stabilize. A higher τ means a slower field evolution rate, leading to slower compensation. In the current simulations, we set the τ of the Percept field to be higher than the LIEC field, but this is a potential dimension of individual variation in the timing of within-trial compensation.

To end, we note two prospects for extending this type of model. First, the production targets obtained from the production field can be used to drive articulatory dynamics (e.g. Saltzman & Munhall, 1989; Stern & Shaw, 2025), such that the fully dynamic nature of the model extends down to even lower levels of motor control (Chaturvedi & Shaw, 2025; Shaw, 2025, 2026). Second, the model has potential to account for other phenomena, beyond one-shot learning, that tightly link speech production and perception. Adaptation has been shown to lead to changes in perception (Lametti et al., 2014; Schuerman et al., 2017; Shiller et al., 2009), which suggests that the synaptic coupling from phoneme node to production field is bidirectional. Since the model can also learn from perception alone, it is well-suited to account for adaptation without overt movement (O. A. Kim et al., 2022; Parrell et al., 2024) as well as offline-transfer effects in perceptual learning studies, i.e., a change in production after passive perception (e.g., Murphy et al., 2026; Sato et al., 2013). Thus, while our model derives one-shot learning, it also offers the potential to generalize to other speech behaviors, enabled by the integration of linguistic representations with speech motor control and perception in a fully dynamic framework.

Conclusion

In this paper, we presented a dynamic neural field model of compensation and adaptation, featuring a dissociated compensation-adaptation mechanism and linguistically-conditioned expectations. Compensation arises via inhibition of production fields while adaptation arises as a result of changes in phoneme node-to-production field coupling. The model captures the basic facts of one-shot learning (Hantzsch et al., 2022) and makes new predictions about how linguistic categories, i.e., lexical items, may condition feedback-based correction in speech production.

Acknowledgments

We thank the anonymous reviewers, whose suggestions greatly improved this paper. Thank you also to members of the DYNAMICS group at Yale, and the attendees at Dynamical Models of Speech (DYMOS) 2025 for helpful comments and suggestions.

References

Bourguignon, N. J., Baum, S. R., & Shiller, D. M. (2014). Lexical-perceptual integration influences sensorimotor adaptation in speech. *Frontiers in Human Neuroscience*, *8*, 208.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*(2), 201–251.

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature neuroscience*, *13*(11), 1428–1432.

Chaturvedi, M., & Shaw, J. A. (2025). A dynamic neural model of tonal downstep. *Proceedings of the Annual Meetings on Phonology*, *1*(1).

Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, *233*(4771), 1416–1419.

Goldinger, S. D. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychological review*, *105*(2), 251.

Guenther, F. H. (2016). *Neural control of speech*. MIT Press.

Hantzsch, L., Parrell, B., & Niziolek, C. A. (2022). A single exposure to altered auditory feedback causes observable sensorimotor adaptation in speech. *eLife*, *11*, e73694. https://doi.org/10.7554/eLife.73694

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature reviews neuroscience*, *13*(2), 135–145.

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience*, *29*(1), 2–20.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, *69*(3), 407–422.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*(5354), 1213–1216.

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in human neuroscience*, *5*, 82.

Jenkins, G. W., & Tupper, P. (2016). A dynamic neural field model of speech cue compensation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *38*.

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of f 0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, *108*(3), 1246–1251.

Kati, L., Sabinasz, D., Schöner, G., & Kaup, B. (2024). Interaction of polarity and truth value-a neural dynamic architecture of negation processing. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*.

Kim, K. S., Gaines, J. L., Parrell, B., Ramanarayanan, V., Nagarajan, S. S., & Houde, J. F. (2023). Mechanisms of sensorimotor adaptation in a hierarchical state feedback control model of speech. *PLoS computational biology*, *19*(7), e1011244.

Kim, O. A., Forrence, A. D., & McDougle, S. D. (2022). Motor learning without movement. *Proceedings of the National Academy of Sciences*, *119*(30), e2204379119.

Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., & Ostry, D. J. (2014). Plasticity in the human speech motor system drives changes in speech perception. *Journal of Neuroscience*, *34*(31), 10339–10346.

Lee, C., Rohrer, W. H., & Sparks, D. L. (1988). Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature*, *332*(6162), 357–360.

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

Mitsuya, T., MacDonald, E. N., Purcell, D. W., & Munhall, K. G. (2011). A cross-language study of compensation in response to real-time formant perturbation. *The Journal of the Acoustical Society of America*, *130*(5), 2978–2986.

Murphy, T. K., Holt, L. L., & Nozari, N. (2026). Exposure to an accent transfers to speech production in a single shot. *Cognition*, *270*, 106412.

Niziolek, C. A., & Guenther, F. H. (2013). Vowel Category Boundaries Enhance Cortical and Behavioral Responses to Speech Feedback Alterations. *Journal of Neuroscience*, *33*(29), 12090–12098. https://doi.org/10.1523/JNEUROSCI.1008-13.2013

Oganian, Y., Bhaya-Grossman, I., Johnson, K., & Chang, E. F. (2023). Vowel and formant representation in the human auditory speech cortex. *Neuron*, *111*(13), 2105–2118.

Parrell, B., Naber, C., Kim, O. A., Nizolek, C. A., & McDougle, S. D. (2024). Audiomotor prediction errors drive speech adaptation even in the absence of overt movement. *bioRxiv*.

Pintado-Urbanc, A. (2025). A dynamic neural field model of asymmetric interference effects in code-switching. *Proceedings of the Linguistic Society of America*, *10*(1), 5909. https://doi.org/10.3765/plsa.v10i1.5909

Purcell, D. W., & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, *120*(2), 966–977.

Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, *119*(4), 2288–2297.

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

Sabinasz, D., Richter, M., & Schöner, G. (2024). Neural dynamic foundations of a theory of higher cognition: The case of grounding nested phrases. *Cognitive Neurodynamics*, *18*(2), 557–579.

Saltzman, E. L., & Munhall, K. G. (1989). A Dynamical Approach to Gestural Patterning in Speech Production. *Ecological Psychology*, *1*(4), 333–382. https://doi.org/10.1207/s15326969eco0104_2

Sato, M., Grabski, K., Garnier, M., Granjon, L., Schwartz, J.-L., & Nguyen, N. (2013). Converging toward a common speech code: Imitative and perceptuo-motor recalibration processes in speech production. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00422

Schneegans, S. (2012). *Cosivina* (Version 1.4). https://github.com/sschneegans/cosivina

Schöner, G., & Spencer, J. P. (2016). *Dynamic thinking: A primer on dynamic field theory*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199300563.001.0001

Schuerman, W. L., Nagarajan, S., McQueen, J. M., & Houde, J. (2017). Sensorimotor adaptation affects perceptual compensation for coarticulation. *The Journal of the Acoustical Society of America*, *141*(4), 2693–2704.

Shaw, J. A. (2025). Unifying phonological and phonetic aspects of speech in dynamic neural fields: The case of laryngeal patterns in japanese. *Phonological Studies*, *28*, 57–68.

Shaw, J. A. (2026). Locating phonological explanation in the neural dynamics of speech: Hysteresis and coupling. *Proceedings of the Annual Meetings on Phonology*, *2*(1).

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, *125*(2), 1103–1113.

Stern, M. C. (2025). Dynamic Field Theory unifies discrete and continuous aspects of linguistic representations. *Proceedings of the Linguistic Society of America*, *10*(1), 5934. https://doi.org/10.3765/plsa.v10i1.5934

Stern, M. C., Chaturvedi, M., & Shaw, J. A. (2022). A dynamic neural field model of phonetic trace effects in speech errors. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44).

Stern, M. C., & Piñango, M. M. (2026). Contextual modulation of language comprehension in a dynamic neural model of lexical meaning. *Cognition*, *266*, 106336.

Stern, M. C., & Shaw, J. A. (2023). Neural inhibition during speech planning contributes to contrastive hyperarticulation. *Journal of Memory and Language*, *132*, 104443. https://doi.org/https://doi.org/10.1016/j.jml.2023.104443

Stern, M. C., & Shaw, J. A. (2025). Nonlinear second-order dynamics describe labial constriction trajectories across languages and contexts. *Journal of Phonetics*, *111*, 101427.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3), 1429–1443. https://doi.org/10.1016/j.neuroimage.2007.09.054

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, *423*(6942), 866–869.