A Dynamical Model of the Speech Perception-Production Link

Kevin D. Roon (kdroon@nyu.edu)
Department of Linguistics, New York University, 10 Washington Place
New York, NY 10003 USA

Adamantios I. Gafos (gafos@uni-potsdam.de)
Linguistics Department and Center of Cognitive Sciences, Universität Potsdam, Haus 14, Karl-Liebknecht-Straße 24-25
Potsdam 14476 Germany, *and*
Haskins Laboratories, 300 George Street
New Haven, CT 06511 USA

Abstract

How and to what extent the speech production and perception systems are linked is a question of longstanding debate (cf. Diehl, Lotto, & Holt, 2004; Galantucci, Fowler, & Turvey, 2006). Despite the long history of this debate and a vast number of studies providing experimental evidence indicating an intimate link between perception and production, formal proposals of this link have been conspicuously lacking in the literature. In this paper, we provide a computationally explicit, dynamical model of the process of phonological planning. In this model, the properties of a perceived utterance automatically serve as input to the ongoing planning of an intended utterance. Using tools from non-linear dynamics, we formalize how incoming inputs from perception influence the ongoing choice of phonological parameter values to be used in production. The use of a dynamical model enables establishing explicit bridges between phonological representations and response time data. Our model provides an account of response time modulations reported in independent experimental work, and makes additional concrete predictions that can be tested experimentally. In sum, our model provides a foundation for better understanding the cognition of speech perception, speech production, and the interaction between the two.

Keywords: Speech production; speech perception; dynamical modeling; perceptuo-motor effects; phonological planning.

Perceptuo-Motor Effects

Many studies have provided evidence for the influence of the speech production system during the process of speech perception. Yuen, Brysbaert, Davis, & Rastle (2010) showed that the articulations subjects produced could be modulated by stimuli they perceived immediately before producing a cued utterance. Specifically, subjects had increased alveolar closure in producing *s*- or *k*-initial utterances when they heard a *t*-initial distractor, compared to baseline cases (*t* is a sound produced with the tongue-tip making full contact at the alveolar ridge, but for fricatives like *s* the tongue-tip contact is not complete, and for *k* the contact is by a different articulator in a different location). D'Ausilio et al. (2009) administered transcranial magnetic stimulation to the areas of subjects' motor cortex that control lip or tongue movement and had subjects identify acoustic stimuli that were ambiguous as to place (labial vs. lingual). They found that subjects were more likely to mistakenly perceive stimuli as having the place whose corresponding motor cortex area was being stimulated. Kerzel & Bekkering (2000) and Galantucci, Fowler, & Goldstein (2009) found that subjects response times (RTs) can be modulated systematically and involuntarily by various stimuli they perceive while speaking. We refer to these effects broadly as "perceptuo-motor effects" (Galantucci et al., 2009) because they are effects that indicate an influence of speech motor plans during the process of speech perception.

Much of the debate in the literature on the speech perception-production link has centered on the claim of the Motor Theory of Speech Perception (Liberman & Mattingly, 1985) that motor codes are the sole object of speech perception. However, as Lotto, Hickok, & Holt (2009) point out, "there is no debate that speech production and perception interact in some manner [...] It is the 'nature' of the production-perception link that has not been established." The purpose of this study is not to disprove either side of the debate around that particular claim of the Motor Theory, but rather to address this latter point and provide a specific computationally explicit proposal regarding the nature of the perception-production link.

In this paper, we propose a specific formalization of the perception-production link within a computational model of the dynamics of phonological planning. To illustrate our model, we focus on the response time data from the response-distractor experimental task used by Galantucci et al. (2009). In this task, subjects learned visual stimulus-spoken syllable pairings (e.g., if you see && say *ba*, if you see ## say *da*). While subjects were preparing the required responses (either *ba* or *da*), distractors were presented at varying delays (i.e., Stimulus Onset Asynchronies, "SOAs") relative to the presentation of the visual cue indicating the required response. The distractors were either a short tone, the same syllable the subject was preparing to say (e.g., *da-da*), or another syllable that differed in place of articulation from the response (e.g., *ba-ga*). In the Kerzel & Bekkering (2000) study, video distractors were used instead of auditory distractors, with similar results.

Figure 1 summarizes the experimental results from Galantucci et al. (2009). First, the presence of any distractor resulted in longer RTs. Second, there was a monotonic effect of SOA on RTs. Both of these effects can be seen by looking at the Tone condition. RTs in the Tone condition (at both SOAs) were slower than on trials when there was no

distractor, and RTs in the Tone condition were longer at SOA 200 than at SOA 100. From these two effects, it is evident that the mere presence of any distractor (linguistic or not) results in a slow-down in RTs. The Identity and Mismatch conditions introduce effects of linguistic (in)congruency between the responses and distractors in addition to whatever process generates the non-linguistic effects seen in the slow down due to a distractor presence and SOA. Crucially, RTs in the Mismatch condition were longer than in the Identity condition.

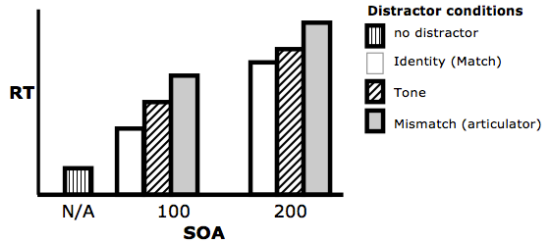

Figure 1: Result patterns from Galantucci et al. (2009).

These results motivate two broad computational principles, which in turn will inform the design of our model. These are the principles of excitation and inhibition. The fact that RTs were shorter in the Identity condition than in the Tone condition indicates the influence of excitation, since linguistic congruency offsets the slow-down introduced by the presence of a distractor. The longer RTs in the Mismatch condition compared to the Tone condition show the influence of inhibition due to linguistic incongruency, increasing the RTs beyond the effects of the mere presence of a distractor.

Dynamical Model of Phonological Planning

We propose a formal, dynamical, computational model of the perception-production link, situating it in the planning process by which phonological parameters are set in speech production. The components of the model are shown in Figure 2. The model includes four dynamical planning fields (shaded rectangles), Inputs to these planning fields (ovals) that determine the actual parameter values to be produced, and a Monitor function that decides when all of the required values have been determined.

Figure 2 also shows an Implementation system that executes the motor plans for the intended utterance based on the production parameter values determined by the model. This Implementation system is not part of our model. The focus in our modeling work is on planning, that is, on the process of choosing values for phonological parameters. This process unfolds in time and, in the schematic shown in Figure 2, takes place before articulatory movement initiation and control, which are the business of the Implementation system. The Implementation system could be, e.g., either the Task Dynamics Model (Saltzman & Munhall, 1989) or the DIVA model (Guenther, 1995).

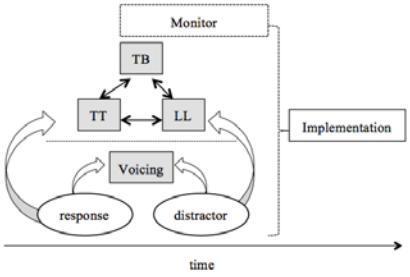

Figure 2: Components of the model.

Planning Fields A key concept in the model is that of the planning field. Each phonological parameter of an intended utterance is assigned a planning field. Planning fields evolve over time and determine the specific parameter settings of the phonological parameters in an intended utterance. A planning field is defined by three axes: an axis representing the possible phonological parameter values, an axis representing the activation level associated with each possible phonological parameter value, and an axis representing time (see Figure 3).

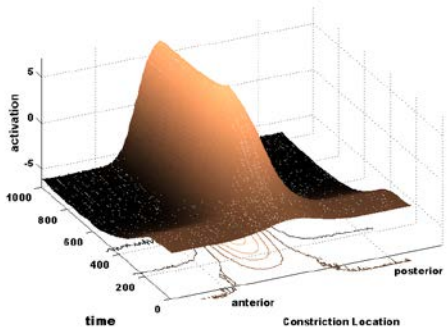

Figure 3: Tongue Tip (TT) articulator planning field.

The phonological parameter relevant to the TT field is that of the constriction location for the tongue-tip articulator. Therefore, the phonological parameter axis in this field is represented by a continuum of constriction locations from dental (most anterior) to post-alveolar (most posterior). The use of a discrete planning field for each parameter is motivated by the desire to have our model be maximally compatible with extant models of phonological representation. The planning fields here correspond closely to the parameters used in Articulatory Phonology (Browman & Goldstein, 1986, *et seq.*), with a field for each "tract variable", though our model could be applied to any appropriate system. As with the tongue-tip articulator, there are also planning fields for the other two primary oral articulators used in producing the syllables relevant to the experimental setting at hand—the lower lip (LL) and tongue body (TB)—and one field for voicing. The parameter axis for the Voicing planning field is represented by the well-known continuum of Voice Onset Time ("VOT").

The planning fields evolve based on inputs to these fields. As we make explicit below, the dynamics of that evolution

are formalized within the computational framework of Dynamic Field Theory ("DFT", e.g., Erlhagen & Schöner, 2002). Each field evolves such that after sufficient input(s), a peak of activation builds up and eventually stabilizes, with one parameter value having a maximum activation level. In Figure 3, following increasing values on the time axis, we can see the gradual evolution of a localized peak in activation at some value of constriction location intermediate between anterior and posterior.

Representing each articulator as its own field in the model with voicing as one separate field reflects the purpose of a planning field, which is to compute a single production value based on one or more potentially conflicting inputs. Our model assumes that these planning fields are the mechanism by which phonological planning of any utterance is achieved, that is, they are not specific to this experimental task. The design of the planning fields therefore reflects the general demands of speech production.

Input There are two sources of input to the evolving planning fields. One corresponds to the parameter values for the required response, and the other corresponds to the parameter values for the auditory distractor perceived during the planning of the utterance. Inputs are represented as two-dimensional distributions of activation levels across the spectrum of possible values for a given parameter. Although not required by the framework, each given input in the present model is a normal distribution defined by the equation:

$$activation_{input} = e^{-(x-val+noise)^2 / 2\sigma^2}$$

val indicates the mean of the distribution, and was varied from trial to trial by adding a small noise term. Since constriction location does not vary in the examples used in the model of this task, the input values for constriction location did not materially change in the simulations. The standard deviation of the distribution (σ) defines its width. Both responses and distractors in the task modeled here are voiced, so the input to the Voicing field was always the same.

Dynamics The purpose of the planning fields is to determine the phonological parameter values to be sent to implementation. Planning fields have two possible stable states. They either stay flat at their resting value, or they can have a single, sustained peak of activation. The value sent to implementation for a given field is the parameter value that has the maximum amount of activation when the field stabilizes in this second stable state. The fields in the model evolve based on the mechanisms of DFT. The dynamics of each of the three articulator planning fields (LL, TT, and TB) are controlled by the equation:

$$\tau dA(x,t) = -A(x,t) + h + r(input_{RESPONSE}(x,t)) + d(input_{DISTRACTOR}(x,t)) - inhibition_{CROSS\text{-}FIELD}(x,t) + interaction(x,t) + noise$$

$dA(x, t)$ is the change in activation level A of x at time step t. The rate of evolution of the field is controlled by τ, with larger values of τ resulting in slower evolution of the field. h is the resting level of the field. The inputs are added to the field, when appropriate, by the terms $input_{RESPONSE}(x, t)$ and $input_{DISTRACTOR}(x, t)$. r and d encode the relative strengths of the inputs. The cross-field inhibition (indicated in Figure 2 by the bidirectional arrows between the articulator planning fields) introduced by any other articulator field(s) is added by the term $inhibition_{CROSS\text{-}FIELD}(x, t)$ when the activation peak in another fields exceeds a threshold value (χ). The DFT dynamics (Erlhagen & Schöner, 2002) capture the case of parameter setting for one effector. In our case, we have several articulators, and a single one must be chosen for any given speech segment. This is the motivation for the cross-field inhibition. In our model, cross-field inhibition follows a basic property of DFT in which inhibition comes into play when some threshold is crossed (we illustrate this with simulations below). Noise is added to introduce stochastic behavior into the model evolutions. The equation that defines the evolution of the Voicing field differs from the one that defines the evolution of the articulator fields only in that it does not contain a term for cross-field inhibition, because the Voicing field neither inhibits nor is inhibited by any other planning field. This design reflects the fact that voicing and articulator are cross-classifying parameters for English consonants (Chomsky & Halle, 1968).

The interaction term, $interaction(x, t)$, the DFT "engine" that drives the evolution of the activation field through local excitation and global inhibition, is defined by:

$$w(x) = w_{excite} e^{-(x^2/2\sigma_w^2)} - w_{inhibit}$$

The interaction term induces changes in the field as some value(s) of x approach a "soft" threshold (θ), which is determined by a sigmoid threshold function, defined by:

$$f(u) = \frac{1}{1 + \exp[-\beta(u-\theta)]}$$

The use of a soft threshold means that some x values below θ do engage the interaction term, but the contribution to the interaction of activation values less than θ diminishes with distance from θ. The system is non-linear due to this soft threshold, in that incremental changes in activation levels have a non-uniform effect on the field's evolution.[1]

[1] The variable values used were: τ = 150 and h = –3.25. The noise term added/subtracted a random amount of activation averaging approximately 1.25 activation units to every x value at each time step in the evolution. The resting level of an activation field was therefore about –2 activation units, equal to the resting level h plus noise. The response input weight (r) was 2.7, and was the same for inputs to both the articulator and Voicing field of the required response. d was 4.5. The cross-field inhibition threshold (χ) was 0. The amount of cross-field inhibition subtracted on each step from other fields when an articulator field was above (χ) was 0.75. The values for the interaction term were the same in all four

Since the required response and the perceived distractor both serve as input to the model, the evolution of the fields is driven by a combination of excitation and inhibition, depending on whether the two inputs have congruent parameter values. Congruent inputs to the model introduce excitation, while incongruent inputs inhibit each other.

Monitor The Monitor determines when activation has built up in required fields to a level that is sufficient to send to Implementation, based on a criterion value (κ), which is the same across all four planning fields. The decision criteria for the Monitor are straightforward. The Monitor waits until the activation level for some x value in both the Voicing field and one articulator field reach criterion. At that point it chooses the parameter values from those two fields with the highest activation level to be sent to Implementation. This has the effect that sometimes it is the Voicing field and sometimes an articulator field—whichever field evolves more slowly—that determines the RT on the trial.

Simulations

Figure 4 illustrates the model dynamics by showing how the planning fields evolve during a single trial in three different conditions of the experimental study from Galantucci et al. (2009). The figures show the maximum activation level over time for each of the four planning fields. The dot-dashed red line shows the TT field evolution, the dashed blue line shows the LL field, the solid pink line shows the TB field, and the solid black line shows the Voicing field. Differences in the rate of rise of the maximum activation level of the fields predict differences in experimental RTs.

Figure 4A shows the evolution of the fields in the Tone condition. Since the tone distractor has no linguistic content, it serves as a baseline reference of how the planning fields evolve in the unperturbed case. The vertical dotted lines at time steps 100 and 500 indicate the duration of the required response input to the fields. Thus, the activation levels of the TT and Voicing fields start to rise at time step 100, the point at which the subject begins planning the required utterance based on the visual cue on that trial (here # # instructing the subject to say *da*). The horizontal dot-dashed black line drawn at activation level 0.7 indicates the soft threshold (θ) that determines the engagement of the within-field interaction term. The effects of the interaction term can be seen in that the rate of increase in the activation level of the TT and Voicing fields is not linear: as the activation level of each field approaches θ, the steepness of the curve increases due to the local excitation being generated in those fields by the interaction term. The TB and LL fields receive no input, and there is no change in their activation levels until around time step 200. At that point their activation levels start to drop due to the TT field reaching the cross-field inhibition threshold (χ), indicated by the horizontal dot-dashed teal line drawn at activation level 0. The TT and Voicing activation levels continue to rise until they both have passed the criterion value (κ), indicated by the solid line drawn at activation level 5. The time step at which the second field passes κ (minus 100, since that is the time step at which the cue is presented) is marked as the RT on that trial (the solid vertical line at about time step 425). The Monitor takes the maximum parameter values from the Voicing and TT fields and passes them to Implementation.

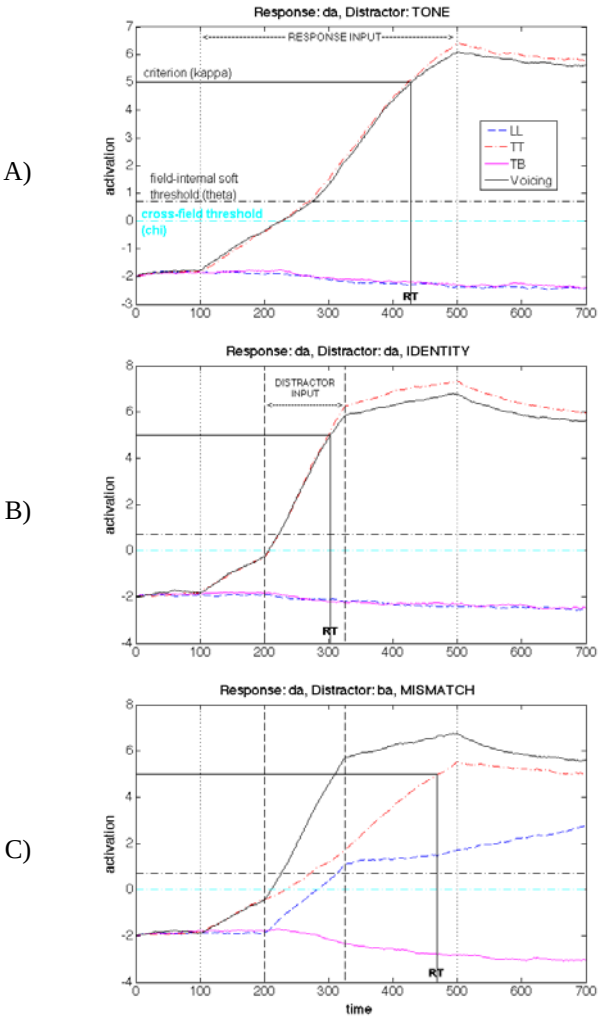

Figure 4: Evolution of planning fields in (A) the Tone, (B) Identity, and (C) Mismatch conditions. Dashed blue lines show the maximum activation level of the LL field, dot-dashed red of the TT field, solid pink of the TB field, solid black of the Voicing field. The horizontal black line at activation = 5 shows the threshold κ at which the Monitor chooses values for production, and the vertical black line perpendicular to it shows the simulated RT on the trial.

activation fields: $\theta = 0.7$, $w_{excite} = 0.45$, $w_{inhibit} = 0.1$, $\sigma = 1$. For the sigmoid function, β was always 1.5. The constriction location input distribution for all articulator fields had a mean (*val*) of 0 and SD = 2, defined on an arbitrary scale of constriction locations that ranged from –10 to 10. For the Voicing parameter, distributions for all voiced stimuli input had a mean of 5 ms VOT and SD = 45 ms. The criterion value (κ) was 5. The specific values of the variables in the above equations are not meaningful in and of themselves. Their values relative to each other are more informative.

Figure 4B shows the evolution of the fields in the Identity case on a trial with a *da* response and *da* distractor. From time step 0 to 200, all fields evolve in the same way as in the Tone condition. The vertical dashed lines at time steps 200 and 325 indicate the duration of the input from the distractor. Since the distractor inputs are qualitatively the same as those for the response, the activation level for the TT and Voicing fields rises at a much greater rate than in the Tone condition because both inputs add activation to the same range of parameter values, in addition to the local excitation being generated by the interaction term. The fields therefore cross κ earlier than in the Tone condition, and the simulated RT is shorter.

Figure 4C shows the evolution of the fields in the Mismatch case on a trial with a *da* response and *ba* distractor. Since the response and distractor share the same value of voicing, the evolution of the Voicing field in this condition is qualitatively the same as in the Identity case. The evolution of the TT field, however, is different. When the distractor input starts at time step 200, the activation level of the LL field begins to rise, and eventually crosses χ, introducing cross-field inhibition to the TT field. The distractor input ends at time step 325, but by that time the LL field maximum is well above θ, so it maintains a peak of activation for some time due to the interaction term, and the cross-field inhibition of the TT field by the LL field therefore persists. As a result, the rate of rise of the TT field activation level slows down compared to its rise in the Tone condition. The Monitor has to wait longer for the TT field to cross κ, and thus the RT on this trial is longer than in the Tone condition.

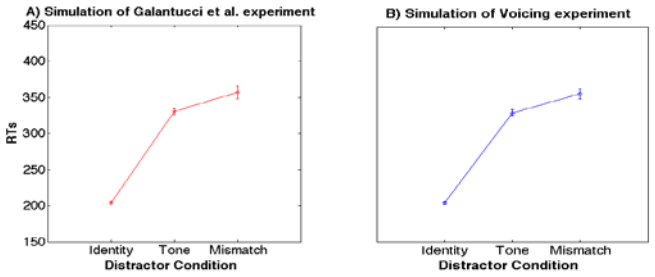

Figure 5: (A) Model simulations of the Galantucci et al. experiment. (B) Predictions for a scenario where distractor and response differ in Voicing rather than articulator.

Using our model, we simulated 150 trials for each of the three conditions (Identity, Tone, and Mismatch) at an SOA of 100 time steps for a total of 450 trials. On each trial, the time step at which the Monitor determined the RT was recorded. The activation level of each planning field was reset to its trial-initial state at the beginning of each trial. The simulated results are shown in Figure 5A. The model qualitatively replicates the experimental results from Galantucci et al. (2009). RTs in the Identity condition were shorter than the Tone condition, due to the reinforcing inputs and lack of any inhibition. On the other hand, RTs were longer in the Mismatch condition than in the Tone condition (and than the Identity condition). This is because the distractor and response differed in articulator. As explained above, in this case cross-field inhibition slows down the evolution of the articulator field for the required response.

Discussion

Our model fills a gap in the speech planning literature. Models of speech motor implementation (Saltzman & Munhall, 1989; Guenther, 1995) explicitly capture how articulators move through space and over time to achieve their linguistic targets, but existing models of the sources of those target values either do not address the timecourse of the planning process (Chomsky & Halle, 1968; Browman & Goldstein, 1986), or assign little to no role to representations at the level of phonological features (Dell, 1986; Roelofs, 2000). Our results show the benefit of a model that addresses timecourse and phonological features explicitly.

Our model makes additional predictions for the cue-distractor task that can be tested experimentally. The model predicts that similar RT effects should be obtained when the Mismatch condition is such that the distractor and response differ in voicing rather than in articulator. The model predicts longer RTs in this Mismatch condition (e.g., *da-ta*) than in the Tone or Identity condition (e.g., *da-da*). In our model, the source of this difference in RTs is the within-field inhibition that arises from the introduction of two incompatible inputs to the same field. This within-field inhibition is an inherent property of the DFT computational framework. Galantucci et al. (2009) did not test this condition, but results reported by Roon (2013) show perceptuo-motor effects of voicing in the response-distractor task that are independent of articulator. This prediction is thus borne out. Figure 5B shows the model predictions for an experiment where the Identity condition is the same as the one reported in Figure 5A (*da-da*), but where the Mismatch condition is *da-ta*. The model predicts slower RTs in the Mismatch condition than in the Tone condition. Future work will involve expanding the model to accommodate these new experimental results.

A second set of predictions concerns variation and phonetic detail in representations and processes. Since categories like voicing are defined as distributions on a phonetic continuum like VOT, compatible inputs need not be exactly the same in order to mutually excite each other: it is sufficient for the maximum activation peaks of two inputs to be near enough to each other. This excitation happens automatically, without any need to classify inputs categorically by defining category ranges. We plan to pursue this set of predictions in future work as well.

Most speech consists of utterances that are longer than monosyllables. Our present model does not address the planning field dynamics beyond CV syllables, which is what is required to account for reported perceptuo-motor effects. Future expansion of the model will address the dynamics involved in the planning of larger utterances.

Our model of the observed experimental effects bears directly on establishing the nature of the perception-production link. In our model, speech perception is linked to

speech production as part of the process by which parameter values are set. The link between perception and production is the obligatory input of the perceived distractor to the motor planning field shown in Figure 2. Given the facilitation and inhibition based on (in)congruency between distractors and responses, there must be some intersection between the motor codes activated during motor planning of the required response and the codes activated during the perception of the distractor. The term "codes" refers to parameters such as voicing and articulator, and more precisely to the parameter values represented in our model. Our claim is not that the codes activated by perceiving the distractor must exclusively be motor codes. Rather, it is that the codes activated in the perception of the distractor must minimally be motor codes. Our study was not designed to address whether non-motor codes are also activated. Our results are fully compatible with the Motor Theory (Liberman & Mattingly, 1985). Our results are also consistent with theories that do or could propose a link between auditory-acoustic (or other) codes that are activated during the perception of the distractor, and motor codes corresponding to these auditory-acoustic codes (cf. Viviani, 2002: Figure 21.12), as long as a link between these other codes and the motor codes is assumed.

In sum, the perception-production link must be specified at the level of setting motor parameter values, including articulator and voicing, that need to be activated either directly or via associated codes. The effects of the perception-production link are seen as the influence of a perceived distractor on the process of setting those parameters, i.e., on a production process, as seen in the reported RT modulations and their simulation by our model.

Conclusions

During speech production, a speaker must retrieve the phonological representations of the required utterances by assembling a set of parameter values that specify the vocal tract actions corresponding to these utterances. We have presented a formal, dynamical, computational model of this process. In the model, assigning values to these parameters is a time-dependent process, captured as the evolution of a dynamical system over time. The model accounts for experimental results that have been proposed as evidence for an intimate link between perception and production. In our model, the perception-production link consists of the phonological parameter values of a perceived stimulus obligatorily contributing to the evolution of the activation levels of the fields engaged with the ongoing phonological planning of a required response. The present model can explain reported effects on response times, and makes new, experimentally testable predictions about similar response time modulations. The model therefore provides a foundation for a better understanding of speech production, perception, and the link between the two.

Acknowledgments

KDR gratefully acknowledge research supported by NSF Grant 0951831. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. AIG gratefully acknowledges support from ERC AdG Grant 249440.

References

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*, 219–252.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology, 19*, 381–385.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psych. Review, 93*, 283–321.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology, 55*, 149–179.

Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psych. Review, 109*, 545–572.

Galantucci, B., Fowler, C. A., & Goldstein, L. M. (2009). Perceptuomotor compatibility effects in speech. *Attention, Perception, & Psychophysics, 71*(5), 1138–1149.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13*(3), 361–377.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psych. Review, 102*, 594–621.

Kerzel, D., & Bekkering, H. (2000). Motor activation from visible speech: Evidence from stimulus response compatibility. *Journal of Experimental Psychology: Human Perception and Performance, 26*(2), 634–647.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1–36.

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *TRENDS in Cognitive Science, 13*, 110–114.

Roelofs, A. (2000). WEAVER++ and other computational models of lemma retrievala and word-form encoding. In L. R. Wheeldon (Ed.), *Aspects of Language Production* (pp. 71–114). Philadelphia: Psychology Press.

Roon, K. D. (2013). *The dynamics of phonological planning.* Doctoral dissertation, Department of Linguistics, New York University, New York, NY.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*(4), 333–382.

Viviani, P. (2002). Motor competence in the perception of dynamic events: a tutorial. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action: Attention and performance XIX* (pp. 406–442). Oxford/New York: Oxford University Press.

Yuen, I., Brysbaert, M., Davis, M. H., & Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences (Social Sciences), 107*(2), 592–597.