Using dynamical representations to model perceptuo-motor interactions and phoneme categorization

NEW YORK UNIVERSITY

Kevin Roon[1], Adamantios I. Gafos[1,2]

[1]New York University, Department of Linguistics; [2]Haskins Laboratories. Authors gratefully acknowledge research supported by NSF Grant 0951831.

Haskins Laboratories

MAIN GOALS

1. What principles guide observed perceptuo-motor effects? What can those principles can tell us about how phonological representations mediate the perception-production loop?

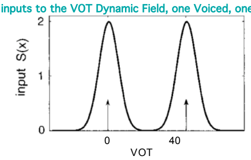
2. Develop a formal computational model to account for Response Time data.

3. Extend the computational modeling approach to other phonological phenomena (phoneme categorization).

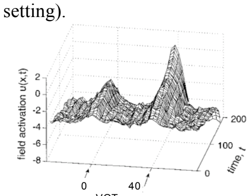
Dynamic Model

- Approach is to embed speech categories in a dynamical setting (Gafos 2006: *LPVIII*), specifically applying Dynamic Field Theory ("DFT": Kopecz & Schöner 1995, Erlhagen & Schöner 2002) to phonological representations.
- **Two key properties:**
  1. Representation of phonological form embraces phonetic detail, being parameterized in continuous dimensions.
  2. Representations evolve over time. Assigning values to these parameters is a time-dependent process, captured as the evolution of a dynamical system over time.
- Representation is the level of activation associated with each parameter value across a continuum of possible values.
- In DFT, given two inputs as in the cue-distractor task, within-category values (Voiceless) mutually raise each other's activation level through local excitation. Two inputs across categories (one Voiced and one Voiceless) reduce each other's activation level through lateral inhibition.
- The Dynamic Field through this interaction eventually stabilizes with one peak of activation.
- Parameter value (e.g., VOT) with the maximum activation serves as output from the Field (for, e.g., gestural parameter setting).

Two inputs to the VOT Dynamic Field, one Voiced, one Voiceless



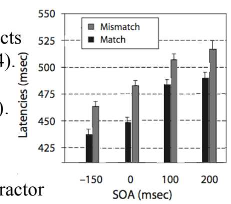
Adapted from Kopecz & Schöner (1995: Figure 3, p.548)



Evolution of the Field with two inputs that laterally inhibit each other, stabilizing to a Voiceless value

Perceptuo-motor effects

Galantucci et al. (2009: Fig 1, shown) designed a cue-distractor task to test for perceptuo-motor effects (Kerzel & Bekkering 2000, Gordon & Meyer 1984).

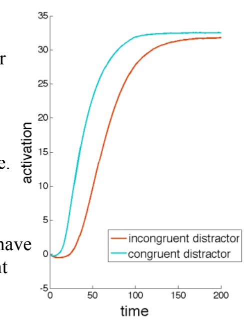


- Subjects saw one of two symbols, responded by saying syllable associated with that symbol (cue).
- Spoken distractor played while subject was responding, either the same as cue or different.
- Subjects' Response Times were faster when distractor was the same as the subject's utterance than when it was different.
- Authors' unformalized interpretation is that distractor activated articulatory plan of perceived utterance. Faster RTs result when that activation combines with activation resulting from planned utterance.

Goals: 1) test for perceptuo-motor effects when cue and distractor share most but not all features, e.g., /da/-/ba/ which differ only in place.

2) develop a formal model of the perception-production link that derives Response Time data

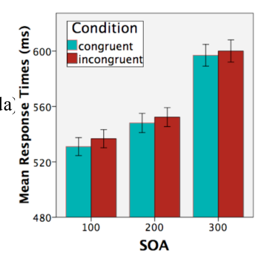
Single-field Dynamic Model



- One dynamic planning activation field gets input from visual cue and acoustic distractor
- VOT value with maximum activation when field stabilizes is sent to production
- Incongruent VOT values inhibit each other, therefore slow the rate of activation increase.
- Congruent VOT values excite each other, speed up the rate of activation increase
- Prediction: faster RTs when cue-distractor have congruent voicing, slower when incongruent

Experiment

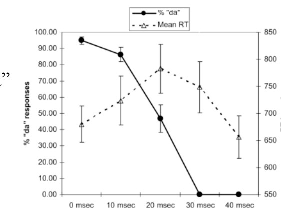
- Same cue-distractor task as Galantucci et al., but without identity condition: place controlled for and only voicing manipulated experimentally
- Cues (da, ta, ga, ka) and distractors (ba, da)
- Test for effect of matched voicing (congruent) vs. mismatched voicing (incongruent) on RTs
- On-going analysis of 38 subjects shows RTs trend as predicted (not significant)



- In progress: additional data on VOT and separate experiments to test for effects of place

Phonemic categorization

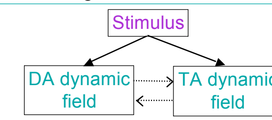
Many studies (e.g., Blumstein et al. 2005: Fig. 1 shown, Phillips et al. 2000, Liberman et al. 1957) have shown the categorical perception of phonemes across a phonetic continuum, e.g., perceiving 'ta' or 'da' over a range of VOT values. Narrow band of VOT values yield ambiguous categorizations at perceptual boundary.



- 0 and 10 ms stimuli are categorized as "da", 30 and 40 as "ta", 20 is ambiguous and categorized ~50% "ta", ~50% as "da"
- RTs are fastest for good stimuli, slower for marginal stimuli, slowest for ambiguous stimuli.
- No account of these very reliable RT results across studies.

Goal: develop one model that derives *both* the categorization and RT results

Coupled-field Dynamic Model



- Two coupled dynamic fields, one corresponding to DA, the other to TA, each with activation pre-shape corresponding to representation of appropriate VOT.
- Same stimulus input for each field.

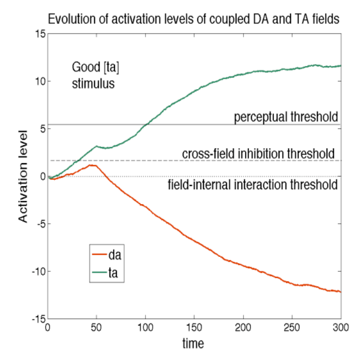
Field coupling is shown in the following set of coupled equations:
dA_ta(t) = –A_ta(t) + r + input(t) + interaction(t) + f(A_da(t)) + noise
dA_da(t) = –A_da(t) + r + input(t) + interaction(t) + f(A_ta(t)) + noise

- Field-internal interaction begins when threshold crossed.
- Cross-field inhibition begins when a higher threshold is crossed. Reduces activation of other field by amount of its maximum activation.
- Coupling represents inherent competition between categories.

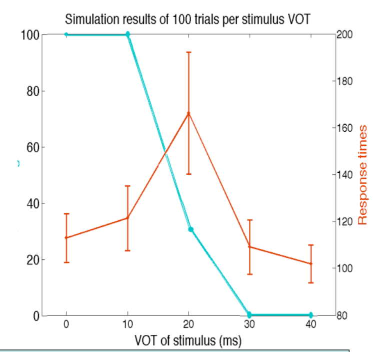
Simulations

- 500 repetitions of tokens with 5 VOTs.
- Field that crosses "perceptual threshold" indicates ta/da decision.
- Time winning field crosses "perceptual threshold" marked as RT.

Evolution of activation levels of coupled DA and TA fields



Simulation results of 100 trials per stimulus VOT



CONCLUSIONS

- Use of dynamical representations enables establishing formal links between representation and Response Time data.
- New experimental data suggests that sub-phonemic properties of speech may play a role in the perception-production link.
- Dynamical model can qualitatively account for *both* Response Time *and* categorization data in a classic phoneme-discrimination task.
- No separate decision-making mechanism necessary for phoneme categorization task due to the inherent stabilization property of the dynamical formalism.
- Models based on dynamical representations can account for different data from one task (e.g., Response Times and phoneme categorization) and the same kind of data across different tasks (RTs in phoneme categorization and in the cue-distractor task), using the same computational dynamical principles.