Journal of Experimental Psychology:
Human Perception and Performance
2009, Vol. 35, No. 2, 588–594

0096-1523/09/$12.00 DOI: 10.1037/0096-1523.35.2.588

Trajectories Emerging From Discrete Versus Continuous Processing Models in Phonological Competitor Tasks: A Commentary on Spivey, Grosjean, and Knoblich (2005)

Robrecht P. R. D. van der Wel, Jeffrey R. Eder, and Aaron D. Mitchel
Pennsylvania State University

Matthew M. Walsh
Carnegie Mellon University

David A. Rosenbaum
Pennsylvania State University

M. J. Spivey, M. Grosjean, and G. Knoblich (2005) showed that in a phonological competitor task, participants' mouse cursor movements showed more curvature toward the competitor item when the competitor and target were phonologically similar than when the competitor and target were phonologically dissimilar. Spivey et al. interpreted this result as evidence for continuous cascading of information during the processing of spoken words. Here we show that the results of Spivey et al. need not be ascribed to continuous speech processing. Instead, their results can be ascribed to discrete processing of speech, provided one appeals to an already supported model of motor control that asserts that switching movements from 1 target to another relies on superposition of the 2nd movement onto the 1st. The latter process is a continuous cascade, a fact that indirectly strengthens the plausibility of continuous cascade models. However, the fact that we can simulate the results of Spivey et al. with a continuous motor output model and a discrete perceptual model shows that the implications of Spivey et al.'s experiment are less clear than these authors supposed.

Keywords: phonological processing, motor control, movement trajectories, discrete versus continuous, psycholinguistics

A central issue in human information processing research is whether processing occurs in discrete, discontinuous stages or in cascading, continuous stages. In research on speech processing, this issue translates to the question of whether lexical selection occurs in continuous or in discontinuous steps as the speech stream is being parsed. Proponents of modular models (Fodor, 1983; Forster, 1979; Frazier & Clifton, 1996) have argued that processing occurs in discrete stages such that a lexical decision is made before further processing occurs. By contrast, proponents of connectionist (McClelland & Elman, 1986; Norris, 1994) or distributed (Gaskell & Marslen-Wilson, 1997) models of speech perception have argued that spoken language processing consists of simultaneous, continuous activation of competitor items. In this latter view, multiple lexical representations remain partially active throughout later stages of processing (Marslen-Wilson, 1987). Thus, in discrete models, only a single lexical item is available at any stage of speech processing, whereas in distributed or continuous models, multiple lexical items remain available throughout speech processing.

Much of the research concerning this debate has focused on resolving lexical ambiguity. Relevant experimental tasks include masked lexical priming and gating (for a review, see Jusczyk & Luce, 2002). Recently, however, researchers have employed eye-movement recordings to investigate this issue further. They have recorded people's eye movements as participants reach for one of (typically) two items, where the names of the items begin with the same or different phonemes (e.g., *candy* vs. *candle* or *candy* vs. *pickle*). How the listeners look at the items as they hear the words gives clues to the dynamics of speech perception. Thus, Magnuson, Tanenhaus, Aslin, and Dahan (2003) found that participants momentarily fixated on a phonologically related distracter (e.g., candy when the target was called candle) when those participants were asked to reach for the target. Magnuson et al. argued from this result that lexical activation begins before the speech signal is fully processed. Spivey and Marian (1999) reached a similar conclusion using a bilingual analog of the look-and-listen procedure.

More recently, Spivey, Grosjean, and Knoblich (2005) expressed concern about a particular aspect of eye-movement data when studying lexical selection. The concern they expressed was rooted in the fact that eye movements from one static object to another are saccadic, meaning that the eyes move rapidly, ballis-

Robrecht P. R. D. van der Wel, Jeffrey R. Eder, Aaron D. Mitchel, and David A. Rosenbaum, Department of Psychology, Pennsylvania State University; Matthew M. Walsh, Department of Psychology, Carnegie Mellon University.

The authors thank Paola Giuli Dussias, Digby Elliott, Judith F. Kroll, James Magnuson, Claire Michaels, and Michael Spivey for helpful comments on this article. The authors also thank John Trueswell for useful comments and James McClelland for encouragement.

Correspondence concerning this article should be addressed to David A. Rosenbaum, Department of Psychology, Pennsylvania State University, University Park, PA 16802. E-mail: dar12@psu.edu

tically, and almost always in a straight line. These features of saccadic eye movements make it difficult to measure graded responses to phonological competitors in individual trials. Gradations in eye movement curvature that would be predicted from a continuous/cascade model can only arise from averaging over trials.

To circumvent this problem, Spivey et al. (2005) recorded manual aiming movements. As shown in Figure 1, they used a computer to register the continuous displacement of a computer mouse from a start location to either of two end locations where two pictures appeared (one at each location). The names of the pictures were either phonologically similar, as in candle and candy or as in picture and pickle, or the names of the pictures were phonologically dissimilar, as in candle and picture or as in candy and pickle. Spivey et al. referred to the former case as the cohort condition and the latter case as the control condition.

In each trial, the participant used the mouse to position the cursor at the start location, which was indicated by a rectangular box near the bottom of the computer screen. The participant then initiated the trial by clicking the mouse, which brought up the two pictures. After the two pictures appeared, participants began moving away from the start position, and 500 ms after the pictures appeared the computer played an audio file with the name of one of the targets. The participant was supposed to move to the named picture and click on it (or the box around it) with the mouse.

Spivey et al.'s (2005) main finding was that hand movements to the targets were more curved in the cohort conditions than in the control conditions, as shown in Figure 1. The authors took this as

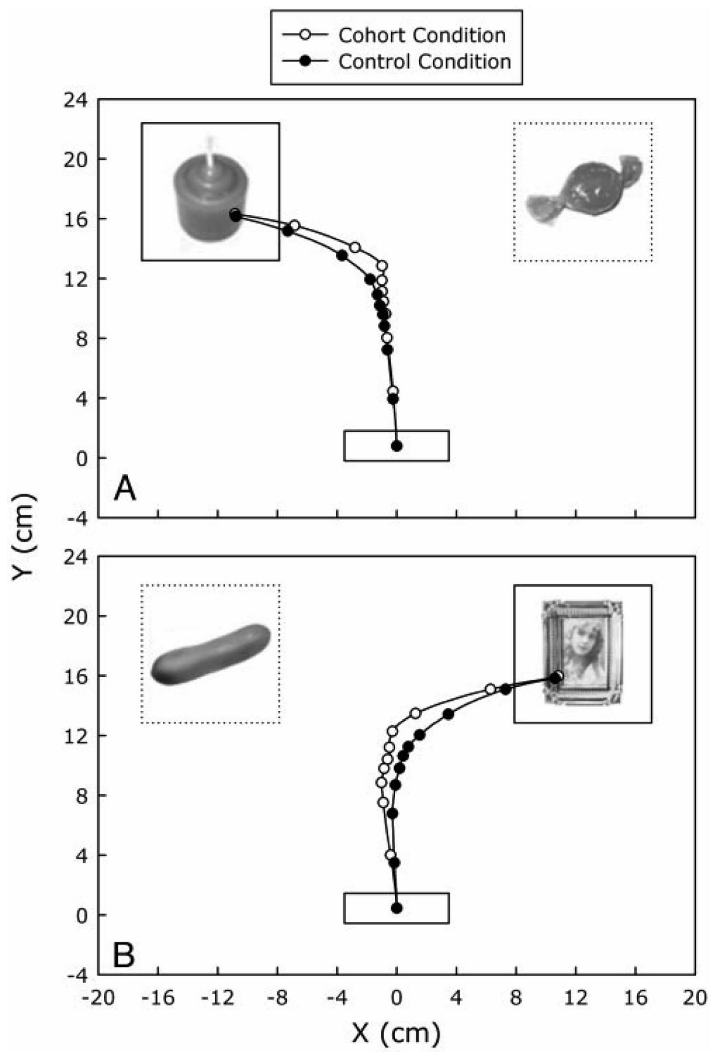

Figure 1. Mean hand trajectories directed to a target on the left (A) or on the right (B) when the names of the left and right targets were phonologically similar (cohort condition) or phonologically dissimilar (control condition). Cohorts *candy* and *candle* are shown in the top panel. Cohorts *pickle* and *picture* are shown in the bottom panel. Symbols are plotted every 10 normalized time slices (e.g., at the 10th, 20th, etc., percentile of the mean movement duration). From "Continuous Attraction Toward Phonological Competitors," by M. J. Spivey, M. Grosjean, and G. Knoblich, 2005, *Proceedings of the National Academy of Sciences*, *102*(29), 10393–10398. Copyright 2005 by National Academy of Sciences, U.S.A. Reprinted with permission.

evidence for a cascade/continuous process model, according to which lexical competitors become activated before the speech processing is completed. Spivey et al. summarized their results this way:

> Following from other measures of dynamic motor output revealing temporally continuous perceptual-motor processes, our present findings do more than contribute to evidence for a cascaded flow of information from perceptual, to cognitive, to motor systems. Our present findings virtually project the ongoing output of the language comprehension process onto a two-dimensional action space in which the potential goal objects act like attractor points and the manual movement serves as a record of the mental trajectory traversed as a result of the continuously updated interpretation of the linguistic input. (p. 10398)

Cascade/Continuous Motor Output

Do the data of Spivey et al. (2005) necessarily support the model they endorsed? In the remainder of this commentary, we argue that they do not. It should be noted that we do not intend to take a stance in the debate over discrete versus continuous speech processing. Our aim in presenting this argument is not to impugn cascade/continuous processing but rather to suggest that the evidence presented by Spivey et al. for such processing is less ironclad than these authors supposed. We show that a discrete processing model can also account for Spivey et al.'s data, provided one ascribes cascade/continuous processing to motor output. As reviewed below, such cascade/continuous motor processing has been demonstrated for targeted hand movements (Flash & Henis, 1991; Henis & Flash, 1995). Of central importance for our argument, the main features of the data reported by Spivey et al. can be simulated with a model that relies on cascade/continuous motor output and discrete rather than cascade/continuous perceptual input. The motor control model used for our simulations (Flash & Henis, 1991; Henis & Flash, 1995) is just one motor control model that could serve this purpose. Other models that share features with the motor control model we used might work equally well.

The Movement Superposition Model of Henis and Flash (1995)

Evidence for cascade/continuous motor processing was provided by Henis and Flash (1995). These authors studied hand movements to one or two visually presented targets. In their experiment, Henis and Flash had participants make horizontal planar arm movements with the right hand, displacing a stylus from a start location to a target location. In the control trials (40% of the trials), a single target location appeared and participants were supposed to make direct movements to that target. In the experimental trials (60% of the trials), the first target was extinguished and was replaced by a different target shown at either of two equally likely locations after an interstimulus interval of 10–300 ms. When the second target appeared, the participants' task was to move to that target only. Whether one target or two targets would appear was unpredictable.

Henis and Flash (1995) found that the curvature of movements to the second target depended on when the second target appeared relative to the hand's motion toward the first target. We will not review the findings of Henis and Flash (1995) in detail here. Instead, we emphasize that their main conclusion was that they could best account for their findings with a movement superposition model. According to the model, two independent movements simply add together in a way that depends on the timing of the second target relative to the hand's movement away from the start position. One movement corresponds to the initially planned displacement from the start position (A) to the first target (B). The second movement corresponds to the displacement from the first target (B) to the second target (C). How the movements add—where in the movement from A to B the movement from B to C is added—depends on the timing of the second target relative to the motion of the hand away from the home position.[1]

Henis and Flash's (1995) model is a cascade/continuous model *par excellence*. Furthermore, it is important to note that Henis and Flash rejected an alternative *abort–replan* model according to which participants aborted the first movement if a second target came on and planned a second movement to the second target. Henis and Flash rejected this alternative model based on detailed, sophisticated analyses of their participants' hand movements. The rejected abort–replan model is a discrete processing model, whereas the accepted superposition model is a cascade/continuous model.

Simulating the Data of Spivey et al. (2005) With the Movement Superposition Model of Henis and Flash (1995)

We now ask whether the data of Spivey et al. (2005) can be simulated by relying on the superposition model of Henis and Flash (1995). If such a simulation can be achieved without invoking a continuous/cascade process for perception, the results of Spivey et al. need not be ascribed to continuous/cascade perceptual processing.

To apply the superposition model to Spivey et al.'s (2005) phonological competitor task, we used a well-known equation in motor control research, developed by Flash and Hogan (1985). This equation defines a "minimum jerk" trajectory from one point to another. Such a trajectory is characterized by a bell-shaped speed profile whose precise form minimizes the sum of the squared rates of change of acceleration (jerk) over the duration of the movement. Using this equation, we defined the in–out component of the first movement as

$$y(t) = y_s + (y_m - y_s)(\alpha_1 T^3 - \alpha_2 T^4 + \alpha_3 T^5), \quad (1)$$

where y_s corresponds to the y value of the start position, y_m corresponds to the y value of the midpoint between the two targets in the Spivey et al. display, $T = t/t_f$ is the ratio of each time t divided by the final time, t_f, and the three coefficients of the polynomial have values used by Henis and Flash (1995) based on earlier work by Flash and Hogan (1985): $\alpha_1 = 10$, $\alpha_2 = 15$, and

[1] Henis and Flash (1995) also distinguished between averaged and nonaveraged initial movement trajectories. Averaged trajectories, which result from rapid target displacement, initially approach a point between the location of the first and the second target. Conversely, nonaveraged trajectories, which result from later target displacement, initially approach the first target. The late spoken-word onset used in Spivey et al.'s task would produce nonaveraged trajectories. These are the type we simulated.

$\alpha_3 = 6$. We assumed that the horizontal component of the first movement was based on the same equation as (1), replacing x for y, but with $x_m - x_s = 0$, so that $x(t) = 0$ for all t. For the second movement, there was no y displacement because this movement, by hypothesis, was from the midpoint between the targets to the selected target. We defined the second movement's x displacement analogously to equation (1):

$$x(t) = x_m + (x_r - x_m)(\alpha_1 T^3 - \alpha_2 T^4 + \alpha_3 T^5), \quad (2)$$

where x_m corresponds to the x value of the midpoint between the two targets, x_r corresponds to the x value of the selected target, and the other terms are the same as before. We added the x vectors of the first and second moves and the y vectors of the first and second moves, allowing for different delays between the two moves. When the first and second moves overlapped, we summed the corresponding values of each. When the moves did not overlap, the first or the second movement vector contributed uniquely to the superposed trajectory.

We simulated 1,200 trials for the cohort condition and 1,200 trials for the control condition, assuming that the second move started later in the cohort condition than in the control condition. This assumption was compatible with any discrete perceptual model that allows longer discrimination times for dissimilar as opposed to similar alternatives. We selected 33% and 55% as the normalized times at which the second movement was initiated in the control and cohort conditions, respectively. These values follow from features of Spivey et al.'s task and data. In their task, the spoken word came on 500 ms after the picture pair was presented, and the spoken word duration was 532 ms. Word discrimination in the control condition occurred around 800 ms after word onset. Word discrimination in the cohort condition occurred around 1,150 ms after word onset. By subtracting the observed movement initiation time of 345 ms from these estimated word discrimination times, we could calculate discrimination time with respect to movement initiation time. These values were 455 ms and 805 ms in the control and cohort conditions, respectively. Finally, by dividing the latter times by the observed movement durations of 1,360 ms and 1,477 ms in the control condition and cohort condition, respectively, we arrived at the normalized times at which the second movement was initiated: 455 ms/1,360 ms $\approx$ .33 for the control condition, and 805 ms/1,477 ms $\approx$ .55 in the cohort condition.

We applied Gaussian noise ($\mu = 0$, $\sigma^2 = 3.5$) to the start time of the second movement, as well as independent vectors of Gaussian noise ($\mu = 0$, $\sigma^2 = .1$) to the x and y components of the first and second movements. The purpose of this noise, which did not affect the mean x and y movement components, was to generate a continuous distribution of movement curvatures. The way we calculated movement curvatures for our simulated trajectories was the same as that used by Spivey et al. We calculated the area between each individual simulated trajectory and a straight line connecting the individual trajectory's start and endpoint.

The resulting trajectory simulations, shown in Figure 2, are similar to the data of Spivey et al. (Figure 1). We did not try to identify parameters that would give us the best fit of the model data to the actual data. The comparisons we sought were only qualitative.

We also studied the Euclidian proximity of the cursor to the ultimately selected target and to the distracter, just as Spivey et al. (2005) did. Their results are shown in the top panel of Figure 3, and our simulated results are shown in the bottom panel of Figure 3. For the simulated data, as for the real data, the Euclidian proximity to the distracter and to the target items diverged more quickly in the control condition than in the cohort condition. In the control condition, where the second move started early, the diver-

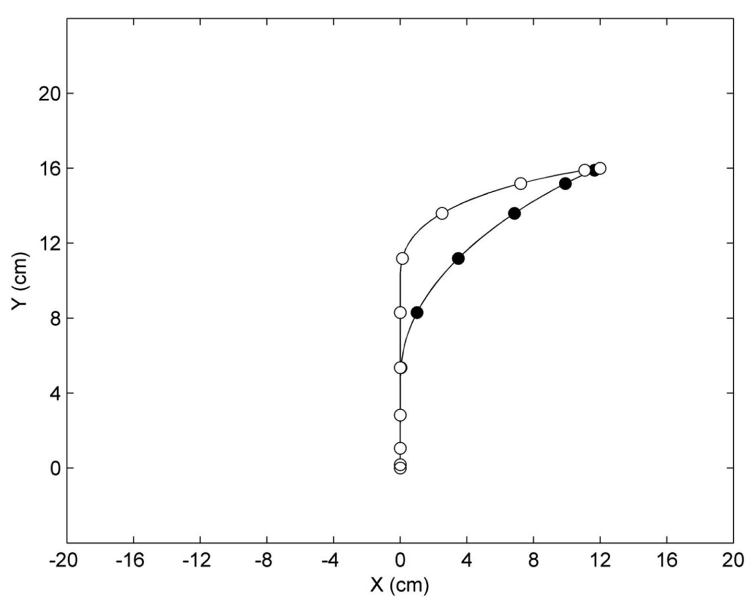

Figure 2. Simulated movement trajectories produced by the movement superposition model. Black circles correspond to simulations with early second move onsets (control trials in phonological competitor task), and white circles correspond to simulations with late second move onsets (cohort trials).

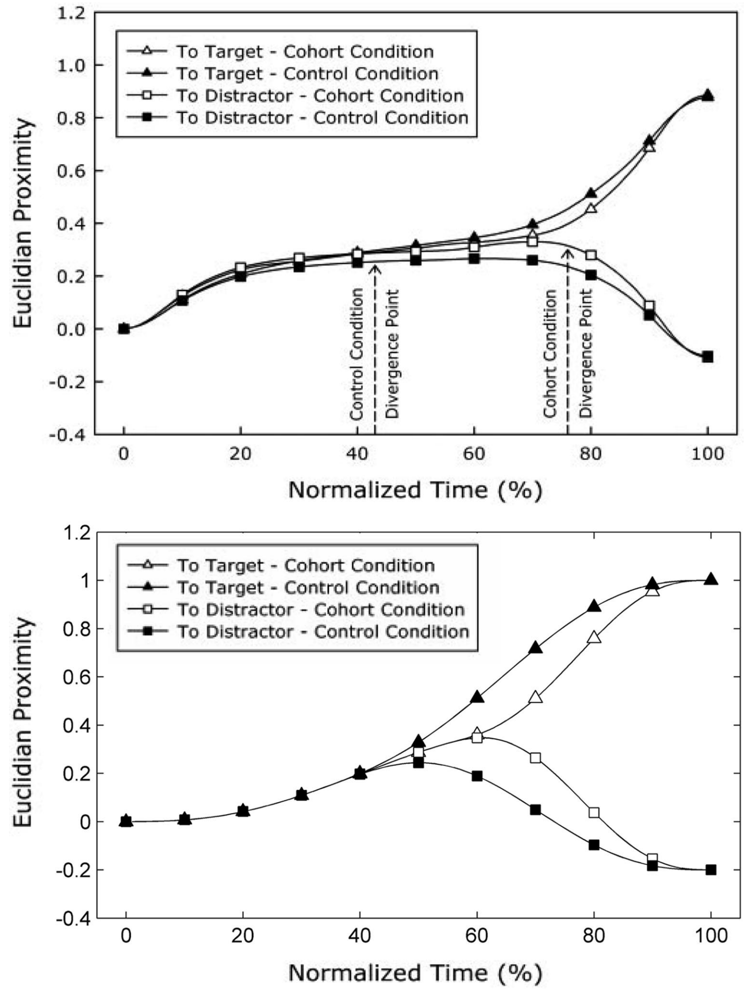

Figure 3. Euclidian proximity to distracter and target items, as observed by Spivey et al. (2005; top panel) and as simulated with the movement superposition model (bottom panel). Top panel and legend of graph in bottom panel from "Continuous Attraction Toward Phonological Competitors," by M. J. Spivey, M. Grosjean, and G. Knoblich, 2005, *Proceedings of the National Academy of Sciences*, *102*(29), 10393–10398. Copyright 2005 by National Academy of Sciences, U.S.A. Reprinted with permission.

gence of the Euclidian proximity points reached statistical significance 43% of the way into the total movement time. In the cohort condition, where the second move started late, the divergence of the Euclidian proximity points reached statistical significance 65% of the way into the total movement time. Qualitatively, then, our simulation data were similar to those of Spivey et al.

The predicted Euclidian proximities differ from the observed data in one regard, however. In the observed data, Euclidian proximities appear to be concave down during the first 40% of movement. Conversely, the graphs of our simulated data are concave up during this portion of movement. This subtle difference likely relates to the shape of participants' velocity profiles. If participants reached peak velocity shortly after movement onset and then continued moving at a sustained, low velocity while waiting for word presentation, the observed patterns in Euclidian proximity would occur. Because Flash and Hogan's (1985) minimum jerk model uses a bell-shaped velocity profile, our simulations do not capture this nuance.

Finally, we turned to another measure from Spivey et al. (2005), the distribution of movement curvature values in the cohort condition. Spivey et al. were careful to check that the high movement curvature they observed when picture name alternatives were phonologically similar was not just an artifact of mixing very low-curve moves with very high-curve moves. Very low-curve moves could emerge if there were no competing lexical activation of the foil (nontarget item) to the target. Very high-curve moves could occur if participants arrived at, or nearly arrived at, the foil location and then turned to the target location, having temporarily misinterpreted the spoken word as referring to the foil. These very low-curve and very high-curve moves could arise from a discrete processing system, and if the two curvature distributions were mixed appropriately, they could give rise to an average trajectory whose curvature spuriously supported cascade/continuous processing.

As shown in the top panel of Figure 4, Spivey et al. (2005) could reject this mixture hypothesis. The distribution of curvatures ob-

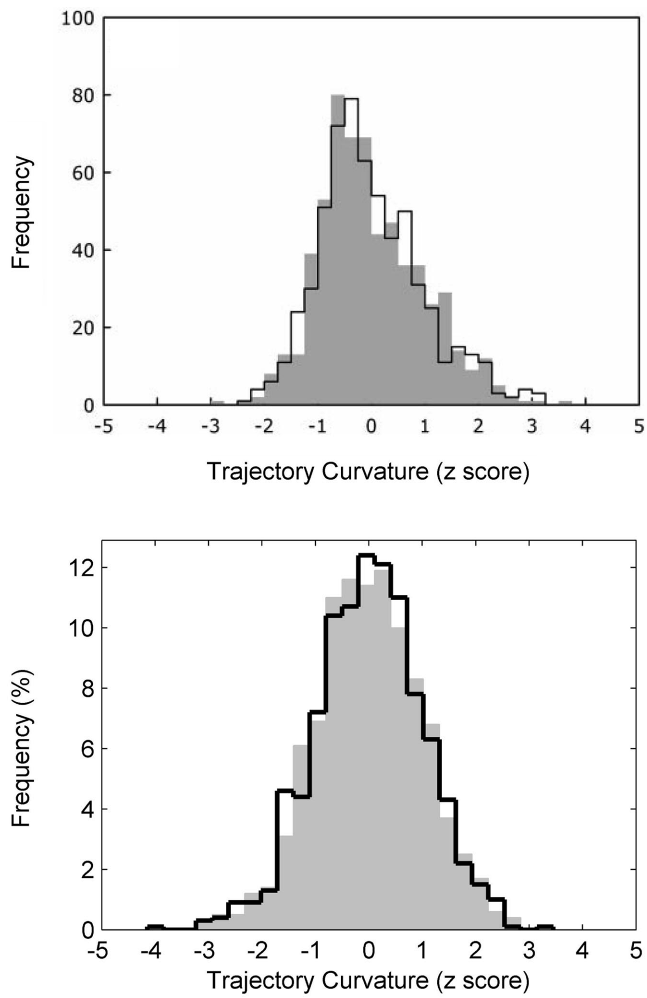

Figure 4. Overlaid histograms of movement curvatures observed by Spivey et al. (2005; top panel) and derived from our simulations (bottom panel). Black lines correspond to trials with early second move onsets. Gray bars and lines correspond to trials with late second move onsets. Top panel from "Continuous Attraction Toward Phonological Competitors," by M. J. Spivey, M. Grosjean, and G. Knoblich, 2005, *Proceedings of the National Academy of Sciences*, *102*(29), 10393–10398. Copyright 2005 by National Academy of Sciences, U.S.A. Reprinted with permission.

tained in their cohort condition was not bimodal, contrary to the mixture hypothesis. Rather, the distribution of curvatures obtained in their cohort condition was unimodal, as predicted by the cascade/continuous model. Moreover, the curvature distribution was normal in both the cohort and the control conditions.

The bottom panel of Figure 4 shows that our simulated data look much the same. As expected, curvatures of our simulated trajectories were greater with late onset second moves than with early onset second moves ($\mu_1 = 54.3$, $\mu_2 = 78.6$, $p < .001$). When we normalized the distributions of movement curvatures separately for each condition, as Spivey et al. (2005) did (top panel of Figure 4), we found results like those of Spivey et al. Consistent with Spivey et al., the distribution of curvature in the early onset condition ($M = 0$, variance $= 1$, kurtosis $= .26$, skewness $= -.16$) did not differ significantly from the distribution of curvature in the late onset condition ($M = 0$, variance $= 1$, kurtosis $= .12$, skewness $= -.13$), and neither distribution deviated from normality as evaluated with the Kolmogorov-Smirnov test (both $p > .15$), as found by Spivey et al.

Conclusions

This commentary has been meant to show that a conclusion reached by Spivey et al. (2005) in their phonological competitor task may have been premature. Spivey et al. said their "findings virtually project the ongoing output of the language comprehension process . . . as a result of the continuously updated interpretation of the linguistic input" (p. 10398). We share the excitement over the prospect of an external projection of the language comprehension process, but we do not share Spivey et al.'s confidence that their data necessarily implied continuous updating of linguistic input. As shown here, Spivey et al.'s data can also be explained with a discrete perceptual process coupled with a continuous motor process.

Does this outcome imply that the measurement of hand movements in phonological competitor tasks can serve no useful purpose for drawing inferences about the dynamics of speech perception? We think not. One possibility for future research using cursor movement data would be to estimate the time at which the second movement is initiated for existing movement data as well as for data from future experiments in which differences between the cohort and control conditions are manipulated. Such an elaboration could provide a more fine-grained understanding of the dynamics of speech perception.

To prepare this commentary, we relied on a model developed by Henis and Flash (1995) to account for hand movements that those authors observed in an experiment that was very different from the experiment of Spivey et al. (2005). Henis and Flash used punctate visual stimuli (light-emitting diodes beneath a glass surface) to specify targets; Spivey et al. used spoken words. Henis and Flash had participants move a stylus from point to point; Spivey et al. had participants move a mouse to displace a cursor on a computer screen. Henis and Flash had clearly defined targets for first moves; Spivey et al. did not. All of these differences make it unclear whether Henis and Flash's model of movement superposition necessarily applies to Spivey et al.'s task. If it does not, this would mean that Spivey et al. could be right that their hand movement data implicate cascade/continuous processing. However, if the movement superposition model of Henis and Flash did not apply to the phonological competition task, one would wonder why.

Clearly, what is needed is some way of ascertaining how successive movements are coordinated in the phonological competitor task. Toward this aim, it will be useful in future research to clearly define the target of the first movement to be made as well as the time for that first movement to be completed. It will also be useful to vary the nature of the stimuli to be discriminated, letting their discriminability differ in graded ways over time so that, even if successive movements do superpose as Henis and Flash suggested, the dynamics of that superposition may depend on the nature of the perceptual discriminations to be made. The nature of that dependency might reveal whether and how the discriminations are made discretely or continuously. More work remains to be done, then, on this difficult problem for which Spivey et al. offered a promising new approach.

References

Flash, T., & Henis, E. (1991). Arm trajectory modifications during reaching towards visual targets. *Journal of Cognitive Neuroscience, 3*(3), 220–230.

Flash, T., & Hogan, N. (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of Cognitive Neuroscience, 5,* 1688–1703.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Forster, K. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27–85). Hillsdale, NJ: Erlbaum.

Frazier, L., & Clifton, C. (1996). *Construal.* Cambridge, MA: MIT Press.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12,* 613–656.

Henis, E. A., & Flash, T. (1995). Mechanisms underlying the generation of averaged modified trajectories. *Biological Cybernetics, 72,* 407–419.

Jusczyk, P., & Luce, P. A. (2002). Speech perception and spoken word recognition: Past and present. *Ear & Hearing, 23,* 1–40.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132,* 202–227.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition, 25,* 71–102.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* l–86.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52,* 189–234.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences, 102*(29), 10393–10398.

Spivey, M., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10,* 281–284.

Received October 22, 2007
Revision received February 28, 2008
Accepted March 9, 2008 ■