

## Perceptuomotor compatibility effects in speech

**BRUNO GALANTUCCI**

*Haskins Laboratories, New Haven, Connecticut  
and Yeshiva University, New York, New York*

**CAROL A. FOWLER**

*Haskins Laboratories, New Haven, Connecticut  
and University of Connecticut, Storrs, Connecticut*

AND

**LOUIS GOLDSTEIN**

*Haskins Laboratories, New Haven, Connecticut  
and University of Southern California, Los Angeles, California*

Kerzel and Bekkering (2000) found perceptuomotor compatibility effects between spoken syllables and visible speech gestures and interpreted them as evidence in favor of the distinctive claim of the motor theory of speech perception that the motor system is recruited for perceiving speech. We present three experiments aimed at testing this interpretation. In Experiment 1, we replicated the original findings by Kerzel and Bekkering but with audible syllables. In Experiments 2 and 3, we tested the results of Experiment 1 under more stringent conditions, with different materials and different experimental designs.

In all of our experiments, we found the same result: Perceiving syllables affects uttering syllables. The result is consistent both with the results of a number of other behavioral and neural studies related to speech and with more general findings of perceptuomotor interactions. Taken together, these studies provide evidence in support of the motor theory claim that the motor system is recruited for perceiving speech.

---

Throughout a number of revisions, the motor theory of speech perception (Liberman, 1957; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985; Liberman & Whalen, 2000) has maintained its distinctive claim that the motor system is recruited for perceiving speech. (For a historical perspective on the development of the theory, see Liberman, 1996, chap. 1.) Although the claim has often been challenged by speech researchers (e.g., Sussman, 1989), in the last 15 years it has gained new credibility, thanks to evidence collected by researchers not traditionally connected to the field of speech (see Galantucci, Fowler, & Turvey, 2006, for a review).

Most notably, the discovery that some neurons in the premotor cortex of a monkey (henceforth, collectively termed the *mirror neuron system*) are active both when the monkey performs an action (e.g., grasping a piece of food) and when the monkey sees someone else performing that action (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992) has boosted the general credibility of motor theories of perception.<sup>1</sup> Moreover, the link between perception and action manifest in the mirror neuron system has been explicitly proposed to be one of the key ingredients for the development of human linguistic abilities (Rizzolatti & Arbib, 1998).

The relevance of such a proposal specifically for the motor theory of speech perception has been recently enhanced by

two sets of findings that more closely relate the mirror neuron system to speech.

The first set of findings connects the mirror neuron system to the vocal-auditory channel in two different species. In monkeys, Ferrari, Gallese, Rizzolatti, and Fogassi (2003) demonstrated that the mirror neuron system responds to communicative actions of the mouth (e.g., lip-smacking), whereas Kohler et al. (2002) demonstrated that some neurons in the mirror neuron system are active when the monkey hears the sound characteristic of the action coded by the neuron (e.g., the cracking noise of a peanut shell for the open-a-peanut action). In birds, Prather, Peters, Nowicki, and Mooney (2008) demonstrated that mirror-like neurons in the higher vocal center of swamp sparrows display nearly identical patterns of activity when the bird either hears a birdsong or sings it in the absence of auditory feedback.

The second set of findings connects portions of the human nervous system that are specifically motoric to speech perception. At a cortical level, two fMRI studies indicated that the same motor areas that are active during speech production are active during speech perception (Pulvermüller et al., 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004). At a more peripheral level, two TMS studies demonstrated that the muscles of the tongue (Fadiga, Craighero, Buccino, & Rizzolatti, 2002) and those of the

---

B. Galantucci, galantuc@yu.edu

---

lips (Watkins, Strafella, & Paus, 2003) are active during perception of speech sequences that include lingual and labial phones, respectively. Moreover, when such motor resonance is altered by TMS, phonetic discrimination is affected (D'Ausilio et al., 2009; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007), suggesting that the speech motor system plays a functional role in speech perception.

In light of these findings, the claim that the motor system is recruited for perceiving speech is no longer a suggestive speculation but, rather, a straightforward explanation for a growing body of empirical evidence. Of course, other theories of speech perception (Diehl & Kluender, 1989; Fowler, 1986; Kuhl, 1981; Massaro & Oden, 1980) might explain findings that the motor system is active during perception by postulating some form of linkage between perceptual and motoric representations of speech. However, the motor theory of speech perception is the only theory that specifically predicts such findings; all other theories of speech perception can offer only post hoc explanations for them.

Our intention in the present study is to test behaviorally a specific prediction that originates from the claim that the motor system is recruited for perceiving speech. Speech production might be selectively affected by the concurrent activation of the speech perception apparatus because, by recruiting the speech motor system, the perceptual system affects the actions brought about by the activation of the motor system. In particular, we follow up on earlier findings of Kerzel and Bekkering (2000).

Explicitly to test the motor theory, Kerzel and Bekkering (2000) ran four experiments in which they varied the stimulus–response compatibility among a task-irrelevant distractor, a video presentation of a speaker mouthing /ba/ or /da/, and a required choice response. In Experiment 1, the task was to utter the syllable /ba/ if “Ba” was printed on the mouth of the speaker on the video and to utter the syllable /da/ if “Da” was printed there. The visible speech gesture either was simultaneous with the onset of the printed syllable or preceded it by 167, 333, or 500 msec. Responses were faster if the visible gesture matched the response that participants were to utter than if the visible gesture and the response did not match. Response times were slower the closer in time the onset of the visible gesture was to the printed syllable, but the patterning of response times remained the same. Kerzel and Bekkering tentatively inferred that the perceived visible gesture activated the motor routines to produce that gesture, thereby facilitating responses if the required response was the one activated perceptually. They recognized, however, that the patterning of response times might be due to *stimulus–stimulus compatibility*, with the visible gestures facilitating reading of the corresponding printed syllable, rather than the *stimulus–response compatibility* effect that they were looking for (i.e., the effect of the visible gesture on the produced gesture). In a second experiment, they substituted symbols (## and &&) that had no preexperimental associations to /ba/ and /da/ gestures, and replicated the findings of their Experiment 1. In case the symbols had acquired associations with /ba/ and /da/ in the course of the experiment, in their Experiment 3, Kerzel and Bekkering went back to “Ba” and “Da” prompts. However, they presented them at least 1 sec before presenting the videos, so that perception of

the printed syllables could not be affected by the videos. The visible gestures now served as “go” signals for the response cued by the printed syllables. The gestures affected response times in nearly the same way as they had in the earlier experiments. In a fourth experiment, effects of the video were eliminated when the mouth was made invisible and lines moved on the screen—closer together, then farther apart (for /ba/) or only increasingly farther apart (/da/).

These findings do suggest that stimulus–response compatibility effects serve as behavioral indices of motor system activation during perception. However, a stronger test of the motor theory would be one in which the video displays are replaced with acoustic speech signals. Although the prediction that a visible gesture affects production of the same gesture is indeed compatible with the claim that the motor system is recruited for perceiving speech, it may or may not be a prediction unique to the motor theory. Most speech researchers agree that perceivers of speech see speech gestures. Without elaboration, no theory of speech perception, other than the motor theory, predicts any consequence of that for speech production. However, given the generality of findings of stimulus–response compatibility (Proctor & Reeve, 1990), it is not difficult to generate an elaboration that would do the job. Replacing the video displays with acoustic speech signals provides a more specific test: No theory, other than the motor theory, predicts that perceiving gestures from acoustic speech signals necessarily affects speech production.

The present study is aimed at providing such a test. The study is composed of three experiments. In Experiment 1, we replicated the findings of Kerzel and Bekkering (2000), with acoustic material. In Experiments 2 and 3, we further explore perceptuomotor compatibility effects in speech, testing them in ever more stringent conditions.

## EXPERIMENT 1

Experiment 1 was designed to replicate the findings of Kerzel and Bekkering (2000) but with acoustic material. In particular, we replicated their Experiment 2, replacing the visible distractors (video displays of a mouth uttering /ba/ or /da/) with the acoustic syllables /ba/ or /da/. We also used a wider set of responses than did Kerzel and Bekkering. In different blocks, responses to the response cues were /ba/ and /da/ (henceforth, *voiced responses*), /pa/ and /ta/ (henceforth, *voiceless responses*), or /ma/ and /na/ (henceforth, *nasal responses*). This manipulation allowed us to explore a contrast that could not be realized with video displays. In fact, when the responses are /pa/ and /ta/ or /ma/ and /na/, there is information in the distractor syllables (/ba/ and /da/) that is consistent with the required response gestures, but, in addition, there is inconsistent information. That is, whereas the visible syllables /ba/, /pa/, and /ma/ all look the same, they differ in voicing or nasality, and the same is true for /pa/, /ta/, and /na/. There is more than one way in which these inconsistencies may affect any influence that the distractors will have on response latencies. From the perspective of the motor theory, the manipulation of the response consonant type should have an effect. However, it is not clear what this effect might be.

If perceiving speech gestures involves access to the speech motor system, and specifically to control structures responsible for producing those speech gestures (Lieberman & Mattingly, 1985), and if that access underlies the effect of the distractor syllable, then perception of a gesture incompatible with a gesture required for the response should decrease or eliminate the facilitating effect of the distractor that matches the constricting organ for the response. For example, if the distractor is /ba/ and the required response is /pa/, the distractor primes adduction of the vocal folds for voicing, but the response requires abduction. If the distractor is /ba/ and the required response is /ma/, the distractor primes velum raising to seal the nasal cavity, but the response requires velum lowering. In both examples, listeners will also perceive a lip constriction gesture, and that should prime the lip constriction gesture required for the response. However, any facilitatory effect of that may be offset by the interfering effect of the inconsistent voicing or velum gesture information. In other words, the outcome of Experiment 1 may differ from that of Kerzel and Bekkering's (2000) Experiment 2, showing a differential effect of the distractor, depending on the response type.

Alternatively, the outcome of Experiment 1 might be exactly like that of Kerzel and Bekkering's (2000) Experiment 2. Perhaps any shared organ (e.g., a lip) or gesture (e.g., constriction at the lips) will have a facilitatory effect on the response, regardless of other mismatches in organs or gestures between the distractor syllable and the cued response, particularly relative to effects of mismatching distractors that differ both in constricting organ and in voicing or nasality.

## Method

**Participants.** Twenty-four participants were included in the experiment. Participants were self-reported native speakers of American English with no known speech or hearing disorders. Each received \$8/h for approximately 1 h of participation.

**Stimuli.** A model speaker (author C.A.F.) was recorded saying two consonant-vowel syllables, /ba/ and /da/, which were used as distractors. The graphic symbols ## and && were used as cues for the response of the participants. Sixteen combinations of experimental stimuli were created by combining the auditory presentation of the two distractors with the visual presentation of the two cues at one of four latencies: 0, 165, 330, and 495 msec.

Participants were exposed to three blocks of the experimental stimuli. Each block consisted of 14 repetitions of the 16 stimuli, yielding 224 trials per block. Participants were instructed to say one CV syllable when presented with one symbol type and a second CV syllable when presented with the second symbol type. For example, they might be instructed to say /ba/ in response to seeing ## and /da/ in response to &&. Because half of the ## symbols were associated with occasions on which the model produced /ba/ and half were associated with productions of /da/, this meant that participants heard a syllable that was consistent with their own responses on half the trials (e.g., they heard the model producing /ba/ and responded /ba/), and heard inconsistent acoustic information on the other half.

A different response pair was used in each of the three blocks. The consonants in each pair differed in constricting organ (the lips for /ba/, /pa/, and /ma/ vs. the tongue tip for /da/, /ta/, and /na/) but shared gestures of the velum and glottis (i.e., voiced, voiceless, or nasal, as indicated above). The order in which participants were exposed to these blocks was counterbalanced in a Latin square design. Additionally, the symbols used to represent each constriction type were counterbalanced across participants. Half of the participants

gave lip responses to ## and tongue tip responses to &&, and half gave lip responses to && and tongue tip responses to ##.

Before beginning the experiment, participants were exposed to eight practice trials without the distractor syllables and eight additional practice trials with them.

**Procedure.** Participants sat in front of a computer monitor. They were given a card that matched each symbol type to the associated response syllable (example: ##, ba; &&, da). They were instructed to respond as quickly as possible when the symbol appeared on the screen but not to respond so quickly that they made numerous errors. They were informed that their responses would be recorded by two microphones, one of which would record how fast they made their responses and the other of which would record their speech onto a DAT player for subsequent analysis. It was explained to participants that as soon as the reaction time microphone registered a response, it would initiate the next trial. However, if they did not respond loudly enough, the next trial would not initiate, and they would see only a blank screen. They were told to repeat themselves more loudly if this occurred, and to speak a little louder in general if this occurred often. Finally, they were asked not to make any extraneous movements during a trial (e.g., moving their chair) so that the reaction time microphone would not be triggered accidentally.

Participants were run in a sound-isolation booth. Stimuli were presented in PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993), and a microphone connected to a voice-activated trigger enabled the PsyScope program to detect responses. (It also recorded latencies, but we used latencies from hand measurements.) A second omnidirectional microphone with a 40 Hz-to-40 kHz frequency response was used to record participant responses. All audio recordings were sampled at 44.1 kHz and recorded directly to one channel of a DAT recorder. An 80-Hz hardware high pass filter (Presonus M80) was applied at the time of recording. A tone that occurred simultaneously with the presentation of the symbol was recorded onto the second channel of the DAT tape. The tone was not heard by the participants and allowed the experimenters to make reaction time measurements of the speech recording by hand.

## Results

Latencies on trials in which an erroneous response was produced were excluded from the analyses. This restriction excluded 212 responses out of 16,128 (1.3%). Also, for each participant and each condition, latencies that were outside a range of 5 standard deviations centered on the mean for that condition were excluded from the analysis. This restriction excluded 197 responses out of 16,128 (1.22%).

Figure 1 displays the latencies. In the leftmost display, latencies of /ba/ and /da/ responses are shown. They closely replicate the findings of Kerzel and Bekkering (2000); when distractor syllables matched the required response, latencies were consistently faster than latencies in corresponding mismatch conditions. However, the results in the middle (/pa/ and /ta/ responses) and rightmost (/ma/ and /na/) panels are different. In the middle panel, two comparisons yielded a numerical match advantage, one yielded a match disadvantage, and one yielded no difference. In the rightmost panel, one showed a numerical match advantage, one a disadvantage, and two showed no difference.

In an ANOVA with stimulus onset asynchrony (SOA), response type and distractor as within-subjects factors, the main effects of SOA and response type were significant, whereas the main effect of distractor was marginally significant (see Table 1 for details). The interaction between distractor and response type was significant, and so was the three-way interaction among the factors. The remaining interactions were not significant.

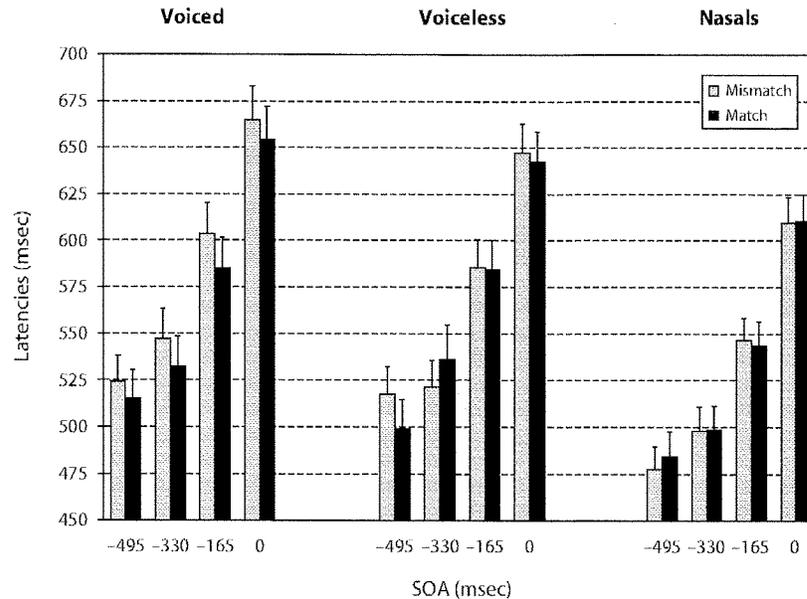


Figure 1. Experiment 1: Mean latencies for each of the experimental conditions. Error bars represent half of 1 standard error of the mean. SOA, stimulus onset asynchrony.

As for the main effect of response type, response times were fastest when responses were nasals ( $M = 534$  msec) and slowest when they were voiced ( $M = 578$  msec); response times when responses were voiceless were closer to the response times to voiced than those to the nasal consonants ( $M = 567$  msec). These differences may be artifacts of the acoustic measurement of latencies. Nasal consonant onset was defined as the onset of the nasal murmur, at the onset of the consonant closure. In contrast, voiced and voiceless stop onsets were defined at the stop burst, at the onset of consonant release following closure.

As for the main effect of SOA, it reflected the finding, as in Kerzel and Bekkering's (2000) Experiment 2, that response latencies decreased the earlier the distractor syllable occurred relative to the response cue.

The interaction of distractor with response type reflected the fact that the effect of distractor was significant for voiced responses [ $M = 13$  msec;  $F(1,23) = 17.1, p < .001, \eta^2 = .43$ ] but not for voiceless or nasal responses ( $M = 3$  msec,  $F < 1$ , and  $M = -1$  msec,  $F < 1$ , respectively). The effect of distractor for voiced responses was

significant at SOAs  $-165$  and  $-330$  and had a numerical trend in the same direction at SOAs  $0$  and  $-495$  (see Table 2 for details).

### Discussion

A major purpose of Experiment 1 was to provide a more direct test of the distinctive claim of the motor theory than did the experiments by Kerzel and Bekkering (2000), by replacing the video distractors with acoustic speech syllables. The results of Experiment 1 replicated the findings by Kerzel and Bekkering, showing that perceiving distractor syllables affects spoken responses. This is consistent with the prediction that perceiving speech gestures in acoustic speech signals affects the speech motor system.

Experiment 1 also provides additional information. Priming was eliminated in match conditions when responses were voiceless or nasal, in contrast to the voiced condition. This is interesting because, as we noted above, when responses were voiceless or nasal and the distractors matched the responses, one gesture (the lip or tongue tip constriction gesture) was shared by responses and distractors; however, another (the voicing/devoicing gesture or the velum raising/lowering gesture) was not. In other words, the results of Experiment 1 support the hypothesis that, when two vocal gestures are concurrently primed, one mismatching gesture eliminates the facilitatory effects that the other gesture may otherwise produce.

Experiment 1 is not a conclusive test of the predictions of the motor theory, however, because the effects that we found cannot be unequivocally attributed to a stimulus-response compatibility effect. On the one hand, as we mentioned above, the symbols ## and && may have acquired an association with the syllables /ba/ and /da/ over the course

Table 1  
Experiment 1: Overall ANOVA

Source	$F$	$p$	$\eta^2$
Response type	$F(2,46) = 8.19$	$<.01$	.26
Distractor	$F(1,23) = 3.73$	.066	.14
SOA	$F(3,69) = 140.89$	$<.001$	.86
Response type $\times$ distractor	$F(2,46) = 3.97$	$<.05$	.15
Response type $\times$ SOA	$F(6,138) = 0.827$	.551	.03
Distractor $\times$ SOA	$F(3,69) = 1.2$	.317	.05
Response type $\times$ distractor $\times$ SOA	$F(6,138) = 2.34$	$<.05$	.09

Note—SOA, stimulus onset asynchrony.

**Table 2**  
**Experiment 1: Comparisons Among Distractor Conditions**  
**Across Stimulus Onset Asynchronies (SOAs)**

Comparison Conditions	SOA	Mean Difference in Latencies (msec)	<i>t</i> (23)	<i>p</i>	<i>d</i>
Voiceless	0	-6	-1.08	.292	-.03
(Match - mismatch)	-165	-1	-0.19	.848	-.01
	-330	14	1.18	.251	.08
	-495	-19	-2.53	<.050	-.12
Nasal	0	1	0.16	.870	.01
(Match - mismatch)	-165	-3	-0.73	.470	-.02
	-330	1	0.18	.855	.01
	-495	7	1.47	.154	.05
Voiced	0	-10	-1.64	.115	-.06
(Match - mismatch)	-165	-18	-3.80	<.010	-.10
	-330	-14	-3.10	<.010	-.09
	-495	-8	-1.21	.240	-.05

of the experiment, leading to a stimulus–stimulus compatibility effect. On the other hand, latencies on mismatch trials are difficult to interpret, because the task-irrelevant syllable corresponds to a syllable in the response set. When distractors and response cues correspond to different acceptable responses, slower latencies in mismatch trials might be due to a problem in selecting the correct response (henceforth, *selection effect*) rather than to a perceptuomotor effect. That is, as suggested by the literature on the Stroop effect (MacLeod, 1991), the mismatch effect may reflect a decision bias at a cognitive level. The irrelevant cue might prime the selection of the wrong response in the response set, slowing down the response process. (In the remaining part of the article, we will refer to stimulus–response compatibility effects that are independent from selection effects as *pure perceptuomotor effects*.)

Moreover, although the method used for the experiment fit the goal of replicating Kerzel and Bekkering's (2000) Experiment 2 with acoustic syllables, the method may not be ideal for investigating perceptuomotor compatibility effects, for two more reasons.

The first reason concerns the SOAs chosen for the experiment. On average, the onset of the responses occurred quite late after the onset of the distractor: 1,014 msec later for the -495 SOA, 870 msec later for the -330 SOA, 759 msec later for the -165 SOA, and 659 msec later for the 0 SOA. Considering that the matching/mismatching information was within the first 100 msec of the syllables that we used as distractors, there was a considerable temporal gap between the perceptual activation that was relevant for our test and the onset of the responses. In other words, the perceptuomotor effects that we detected must have derived either from activations of the motor system that lasted a relatively long time or from responses that had particularly short latencies. This might not be the best strategy to detect pure perceptuomotor effects. Pure perceptuomotor effects might be easier to detect when the distractor occurs right before the motor system is engaged by the response cue.

The second reason that the method used for Experiment 1 is not ideal for investigating perceptuomotor com-

patibility effects concerns the exact nature of the effect that we found. Although we have referred to facilitatory priming effects of the distractors, most likely, there is more going on. In picture naming experiments, for example, acoustic distractors that share phonological properties with the picture names lead, at appropriate SOAs, to faster responses than do phonologically unrelated distractors (Levelt et al., 1991). However, all distractors slow responding relative to conditions in which no distractors are presented. Our finding that response times increased at shorter SOAs (i.e., with decreased temporal separation of the distractors from the response cues to the response) may reflect such an interference effect. However, in absence of proper baseline conditions, we cannot assess the interfering (or facilitatory) effect that the distractors might have on the response process.

Experiment 2 was designed to address the interpretative and methodological issues of Experiment 1.

## EXPERIMENT 2

As we described above, Kerzel and Bekkering (2000) addressed the issues of stimulus–stimulus compatibility and selection effects by presenting the distractors as go signals 1 sec after the presentation of the response cue. We addressed the issue of stimulus–stimulus compatibility in a different way, by introducing an additional task, performed manually. That is, participants learned to respond by pressing one key for one symbol and a different key for the other. The distractor conditions and SOAs used in the manual-response task were the same as those in the vocal-response task. Considering that each participant was exposed to three blocks of manual trials alternated with three blocks of vocal trials, the contrast between the manual and the vocal tasks offers a robust test for a possible stimulus–stimulus compatibility effect. In fact, if we observe the same pattern of results in the manual task as in the vocal task, the effects found for the vocal task cannot be attributed with certainty to a perceptuomotor effect, as required by a motor theoretical account of the result, but may be attributed to a stimulus–stimulus compatibility effect. By the same token, if we find different patterns of results in the two tasks and the difference indicates a higher sensitivity to distractors in the vocal task, we can conclude that there is an interaction between the spoken stimulus (i.e., the distractor syllable) and the vocal response, as would be predicted by the motor theory.

To control for selection effects, we added the distractor syllable /ga/. Given that /ga/ is not in the response set that the participants used during the experiment, this condition provides an opportunity to separate the selection and perceptuomotor effects. The syllable /ga/ constitutes a mismatch trial whether the response is /ba/ or /da/. However, because /ga/ is not a response option, it cannot bias participants toward either of the two possible responses in the response set, and no selection effect should occur. (In consequence, the term *mismatch trials* will be used in Experiment 2 exclusively to indicate trials in which the distractor corresponds to a wrong response in the response set.) If response latencies on trials with /ga/ as a distractor are close to the latencies on mismatch trials, and both latencies are longer than latencies on match trials, we can conclude that

the result reflects a pure perceptuomotor effect. However, if response latencies on trials with /ga/ as a distractor are shorter than latencies on mismatch trials, we need to consider the comparison with match trials. If response latencies on trials with /ga/ as a distractor are longer than latencies on match trials, both selection and perceptuomotor effects are likely present. If response latencies on trials with /ga/ as a distractor are close to the latencies on match trials, the results likely reflect a pure selection effect (i.e., a selection effect that occurs independently of perceptuomotor effects).

The remaining changes in design between Experiments 1 and 2 concern the methodological issues with Experiment 1. In particular, we used a different set of SOAs and added two new baseline conditions.

In Experiment 1, distractors were either simultaneous with the response cues or preceded them. In Experiment 2, we chose just one SOA to represent those SOAs of Experiment 1 in which the distractor preceded the response cues. We chose an SOA of 150 msec. (We will refer to this as SOA -150 msec.) As in Experiment 1, we also used an SOA of 0 msec. Then, to explore SOAs in which the distractor syllables followed the response cues, we chose SOAs of 100 msec and 200 msec. (We will refer to these SOAs as SOA 100 and SOA 200.) These two SOAs were chosen to investigate the effect of distractors that were presented in close temporal proximity to the response, enhancing the likelihood of detecting pure perceptuomotor effects.

As for the new baseline conditions, in one of them, no distractor was presented, and, in the other, a tone was presented. The no-distractor condition provides a baseline for the response latencies, allowing us to assess the general facilitatory and/or interfering effects of the distractors. The tone condition allows us to assess the distracting effect (if any) of a sound that, according to the motor theory, should not activate the speech motor system. This condition is particularly useful, because it allows us to assess the *relative* facilitatory and/or interfering effects of the distracting syllables with respect to a neutral (i.e., nonspeech) distractor.

## Method

**Participants.** Forty-two undergraduate students at the University of Connecticut participated in the study.<sup>2</sup> Participants were self-reported native speakers of American English with no known speech or hearing disorders. Each received course credit for approximately a half hour of participation.

**Stimuli.** The syllables /ba/, /da/, and /ga/ were recorded by a male native speaker of American English and then edited to reduce their overall duration to 150 msec (in Experiment 1, the duration for /ba/ was 386 msec; that for /da/ was 411 msec). This was done because we wanted to minimize the likelihood of a temporal overlap between the distractors and responses. The edited syllables were clearly and easily recognized by a small sample of native speakers of American English.

Participants were exposed to six blocks of the experimental stimuli, three blocks of manual responses alternating with three blocks of vocal responses. Each block consisted of four repetitions of the combinations of the five distractors and four SOAs, yielding 80 trials per block. In the vocal task, participants were instructed to say /ba/ when presented with == and /da/ when presented with ##. In the manual task, participants were instructed to press a green key on the left side of a response box when presented with == and to press a yellow key to the right of the green key when presented with ##. Half of the participants began with a block of manual responses, half with a block of vocal responses.

Prior to beginning the experiment, participants were exposed to 20 practice trials with the manual task and 20 practice trials with the vocal task. During the practice trials, there were no distractors.

**Procedure.** Participants sat in front of a computer monitor. They were given a card that matched symbol type to the associated response syllable (==, ba; ##, da). They were instructed to respond as quickly as possible when the symbol appeared, but not to respond so quickly that they made numerous errors. In the vocal task, they were informed that their responses would be detected by a microphone that would record how fast they made their responses. It was explained to participants that, as soon as the reaction time microphone registered a response, the next trial would be initiated. However, if they did not respond loudly enough, the next trial would not initiate. They were told to repeat themselves more loudly if this occurred, and to speak a little louder in general if this occurred often. Finally, they were asked not to make any extraneous movements during a trial (e.g., moving their chair) so that the reaction time microphone would not be triggered accidentally.

Stimuli were presented in E-prime (Psychology Software Tools, Inc., Pittsburgh, PA), and a microphone connected to a voice-key-activated trigger recorded reaction times.<sup>3</sup> The distractors were presented via headphones. The participants were told to ignore what they heard over the headphones because it was irrelevant for the experimental task.

## Results

Latencies for trials on which an erroneous response was produced were excluded from the analyses.<sup>4</sup> This restriction excluded 476 responses out of 20,160 (2.4%). Also, for each participant and each condition, latencies that were outside a range of 5 standard deviations centered on the mean for that condition were excluded from the analysis. This restriction excluded 275 responses out of 20,160 (1.4%).

Figure 2 displays the absolute mean latencies for the two tasks, whereas Figure 3 displays the same latencies relative to the no-distractor condition. Visual inspection of the figures shows that (1) manual responses were faster than were vocal responses, (2) latencies increased with SOA, and (3) distractors had an impact on latencies. Separate ANOVAs were performed for the manual and vocal tasks, with SOA (-150, 0, 100, 200) and distractor (match, mismatch, tone, /ga/, no distractor<sup>5</sup>) as within-subjects factors. These analyses highlighted different patterns of results for the two tasks.

**Manual task.** The purpose of the manual task was to test for the presence of stimulus-stimulus compatibility effects. A motor theoretical account of our results requires ruling out such effects, which implies finding no differential effect of distractors on manual responding. That is what we found.

There was a significant main effect of SOA [ $F(3,123) = 16.9, p < .001, \eta^2 = .29$ ], but distractor was not significant ( $F < 1$ ), and there was no significant interaction between the two factors [ $F(12,492) = 1.3, p = .21, \eta^2 = .03$ ]. The significant effect of SOA was due to the fact that, at SOA -150, all distractors led to shorter latencies than at SOA 0, and both of these SOAs led to shorter latencies than at SOAs 100 and 200. This effect, which is compatible with the main effect of SOA found in Experiment 1 and in the experiments by Kerzel and Bekkering (2000), is most likely due to the fact that the distractors functioned as alerting signals, indicating that the response cue was about to appear. Alerting effects of this kind are common in reaction time experiments (Posner, 1978), and our interpreta-

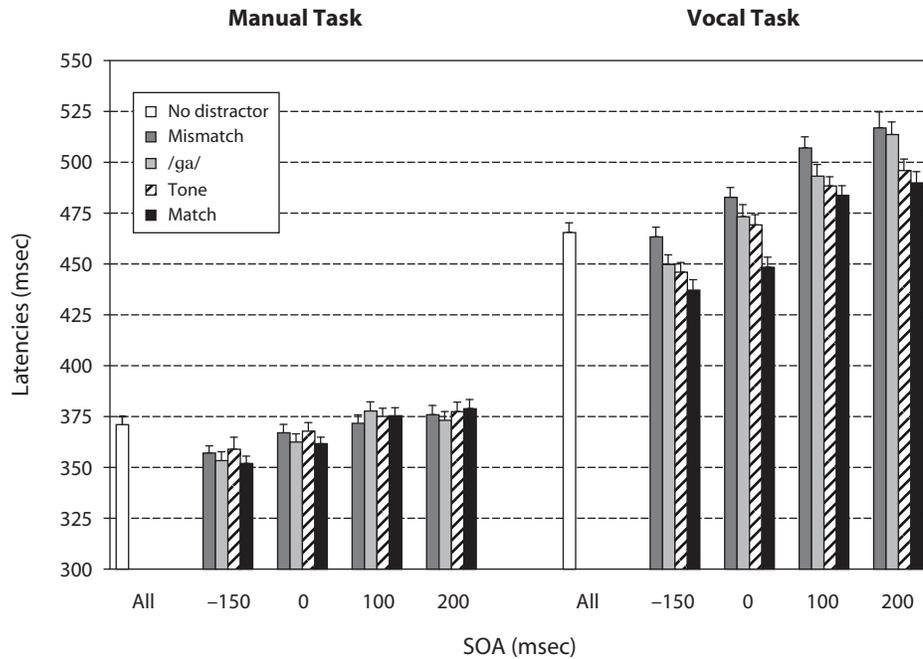


Figure 2. Experiment 2: Mean latencies for each of the experimental conditions (for the no-distractor condition, there was no stimulus onset asynchrony [SOA] manipulation).

tion is supported by the fact that, at SOA -150, all trials with distractors were associated with faster responses than were trials with no distractor. Latencies on the trials with distractors were also shorter at SOA 0 than they were at SOAs 100 and 200<sup>6</sup> (see the left side of Figure 2). These results are compatible with the well-known facilitatory effect of concurrent auditory stimulation for tasks that depend on visual stimuli (e.g., Doyle & Snowden, 2001).

**Vocal task.** In an ANOVA with SOA and distractor as within-subjects factors, SOA, distractor, and their interaction were all statistically significant [SOA,  $F(3,123) = 42.1, p < .001, \eta^2 = .51$ ; distractor,  $F(4,164) = 19.7, p < .001, \eta^2 = .32$ ; SOA  $\times$  distractor interaction,  $F(12,492) = 7.2, p < .001, \eta^2 = .15$ ]. The results were further analyzed with four separate ANOVAs, one for each SOA. At all SOAs, distractor was statistically significant [SOA -150,  $F(4,164) = 6.5, p < .001, \eta^2 = .14$ ; SOA 0,  $F(4,164) = 8.2, p < .001, \eta^2 = .17$ ; SOA 100,  $F(4,164) = 13.9, p < .001, \eta^2 = .25$ ; SOA 200,  $F(4,164) = 15.9, p < .001, \eta^2 = .28$ ].

Replicating the results of Experiment 1, latencies in the match trials were faster than latencies in the mismatch trials at SOA 0. The same pattern was observed for SOAs -150, 100, and 200. The difference was statistically significant, collapsed across SOAs (see Table 3), as well as at each SOA [SOA -150,  $M = 27$  msec,  $t(41) = 3.54, p < .01, d = .38$ ; SOA 0,  $M = 34$  msec,  $t(41) = 5.88, p < .001, d = .5$ ; SOA 100,  $M = 23$  msec,  $t(41) = 3.9, p < .001, d = .32$ ; SOA 200,  $M = 27$  msec,  $t(41) = 3.09, p < .01, d = .29$ ].

We turn now to the effects of the different distractors. We introduced the /ga/ distractor to distinguish two interpretations of the match effect. One interpretation is that the effect may have a perceptuomotor origin. Perceiving a syllable primes a matching motor response. In that case, /ga/

should be as effective a distractor as any other mismatching syllable. It should slow responses. Alternatively or in addition, a selection effect might occur such that, when the distractor matches one of the response options, participants are disposed to respond with that option. This helps on match trials, but impairs performance on mismatch trials. Because /ga/ is not a response option, it cannot cause a selection effect. We find evidence for both interpretations.

Latencies on the match trials were faster than latencies with the /ga/ distractor at all SOAs. The difference was statistically significant collapsed across SOAs (Table 3), as well as at SOAs 0 [ $M = 25$  msec,  $t(41) = 3.17, p < .01, d = .32$ ] and 200 [ $M = 24$  msec,  $t(41) = 3.54, p < .01, d = .29$ ]. At SOAs -150 [ $M = 13$  msec,  $t(41) = 1.83, p = .074, d = .18$ ] and 100 [ $M = 9$  msec,  $t(41) = 1.69, p = .099, d = .13$ ], the difference was marginally significant.

Table 3  
Experiment 2: Comparisons Among Distractor Conditions Across Stimulus Onset Asynchronies (SOAs)

Comparison Conditions	Mean Difference in Latencies (msec)	$t(41)$	$p$	$d$
Mismatch - match	28	8.02	<.001	.52
/ga/ - match	18	4.81	<.001	.33
Mismatch - /ga/	-10	-2.95	<.010	-.18
Match - tone	-10	-3.66	<.001	-.20
No distractor - mismatch	-27	-6.75	<.001	-.50
No distractor - match	-1	-0.14	.884	-.01
No distractor - tone	-9	-2.49	<.050	-.18
No distractor - /ga/	-18	-4.81	<.001	-.33
/ga/ - tone	8	2.10	<.050	.14
Mismatch - tone	18	5.37	<.001	.33

These findings are consistent with the interpretation that the distractor effect has a perceptuomotor origin.

However, a selection effect occurred as well. Latencies when /ga/ was the distractor were faster than latencies on the mismatch trials at all SOAs. The difference was statistically significant collapsed across SOAs (Table 3), as well as at SOAs  $-150$  [ $M = 14$  msec,  $t(41) = 2.26$ ,  $p < .05$ ,  $d = .2$ ] and  $100$  [ $M = 14$  msec,  $t(41) = 2.35$ ,  $p < .05$ ,  $d = .18$ ].

We included the tone distractor and trials without a distractor to provide baseline conditions for our syllable distractors. To get an estimation of the general interference/facilitation effects of the distractors, we first looked at the difference between the tone condition and the no-distractor condition. The results indicate that different effects were possible. At SOA  $-150$ , the tone condition led to significantly shorter latencies than did the no-distractor condition [ $M = 19$  msec,  $t(41) = 2.99$ ,  $p < .01$ ,  $d = .28$ ]. At SOAs  $100$  and  $200$ , the tone condition led to significantly longer latencies than did the no-distractor condition [SOA  $100$ ,  $M = 23$  msec,  $t(41) = 4.25$ ,  $p < .001$ ,  $d = .35$ ; SOA  $200$ ,  $M = 30$  msec,  $t(41) = 5.15$ ,  $p < .001$ ]. At SOA  $0$ , the tone condition led to slightly longer latencies than did the no-distractor condition, but the difference was not significant [ $M = 4$  msec,  $t(41) = 0.77$ ,  $p = .45$ ,  $d = .05$ ]. These results are consistent with the results of the manual condition.<sup>7</sup> Not only do we find the same alerting effect at SOA  $-150$ , but we also find a similar overall linear trend over SOAs: The shorter the temporal gap between the tone and the response, the slower the responses [ $F(1,41) = 36.22$ ,  $p < .001$ ]. As for the three syllable conditions, they all differed from the no-distractor condition in a way similar to the tone condition, leading to a similar overall linear trend over SOAs [ $F(1,41) = 59.25$ ,  $p < .001$ ].

The next step of our analyses involved comparing tone trials to the match/mismatch trials, in order to detect relative interference and/or facilitation effects. The pattern of results that we found was consistent with the pattern predicted by the motor theory of speech perception. At all SOAs, latencies for the tone distractor were shorter than were latencies for the mismatch trials and longer than were latencies for the match trials. The former difference was statistically significant collapsed across SOAs (Table 3), as well as at each SOA [SOA  $-150$ ,  $M = 17$  msec,  $t(41) = 2.68$ ,  $p < .05$ ,  $d = .25$ ; SOA  $0$ ,  $M = 14$  msec,  $t(41) = 2.48$ ,  $p < .05$ ,  $d = .18$ ; SOA  $100$ ,  $M = 19$  msec,  $t(41) = 2.98$ ,  $p < .01$ ,  $d = .27$ ; SOA  $200$ ,  $M = 21$  msec,  $t(41) = 2.35$ ,  $p < .05$ ,  $d = .22$ ]. The latter difference was significant collapsed across SOAs (Table 3) and at SOA  $0$  [ $M = 21$  msec,  $t(41) = 4.23$ ,  $p < .01$ ,  $d = .29$ ].

## Discussion

The primary goal of Experiment 2 was to address two interpretative issues with Experiment 1. In particular, we wanted to test whether the match/mismatch effect that we found in Experiment 1 could be detected independently from both stimulus–stimulus compatibility effects and selection effects. Experiment 2 provided valuable results with regard to both issues.

As for stimulus–stimulus compatibility effects, Experiment 2 confirmed the conclusions of Kerzel and Bekker-

ing's (2000) Experiment 3. Responses in the manual task were affected by the distractors only when the distractor preceded the response cue (SOA  $-150$ ), and, in that case, all distractors had an equal effect. In contrast, the effect of the distractors was significant at all SOAs in the vocal task and was modulated by compatibility effects. Considering that participants performed the two tasks three times each in alternated blocks, this pattern of results indicates that the differences between distractors that we obtained in the vocal task are related specifically to the vocal response and not to the perceptual processes that the manual and vocal tasks have in common.

As for selection effects, the results of Experiment 2 support their existence; /ga/ distractors—that is, mismatching syllables that were not in the response set—led to shorter latencies than did mismatching trials in which the syllable was part of the response set. Although the results of Experiment 2 showed that selection effects occurred *together* with perceptuomotor effects (/ga/ distractors led to longer latencies than did distractor syllables that matched the response), the existence of selection effects affects the interpretation of our Experiment 1 and of Experiments 1 and 2 of Kerzel and Bekkering (2000), suggesting that results in these experiments may also reflect a combination of selection and perceptuomotor effects. We will return to this issue at the end of this section.

Also, the changes in method between Experiments 1 and 2 provided valuable results. First, Experiment 2 confirmed our expectation that distractors slow response times except when they predict the occurrence of the response cue, as they did at SOA  $-150$ . At SOAs  $100$  and  $200$ , the distracting effect was clearly observable for all distractors in the vocal task (Figure 3).

Second, the use of a tone distractor gave us the opportunity to further test the predictions of the motor theory. We found that the tone distractors led to shorter latencies than did mismatching syllables and to longer latencies than did matching syllables. This pattern of relative facilitation and interference is fully consistent with the predictions of the motor theory.

Finally, the use of different SOAs seems to have enhanced our ability to detect pure perceptuomotor effects. In fact, although the results of Experiment 2 indicate that, at earlier SOAs, the match/mismatch effect is not a pure indicator of a perceptuomotor effect, at SOA  $200$  the match/mismatch effect seems to be due solely to perceptuomotor effects. Mismatch trial response latencies were very close to those for /ga/ trials [difference = 3 msec;  $t(41) = 0.42$ ,  $p = .68$ ,  $d = .03$ ], and both trial types were associated with significantly slower response latencies than were match trials. These results indicate that pure perceptuomotor effects can be detected separately from other effects only when the distractor is presented very close in time to the moment at which the vocal response is about to be produced. This finding again affects the interpretation of our Experiment 1 and of Experiments 1 and 2 by Kerzel and Bekkering (2000), which all used SOAs far from SOA  $200$ .

Experiment 3 was designed to more directly and reliably detect the presence of pure perceptuomotor compatibility effects.

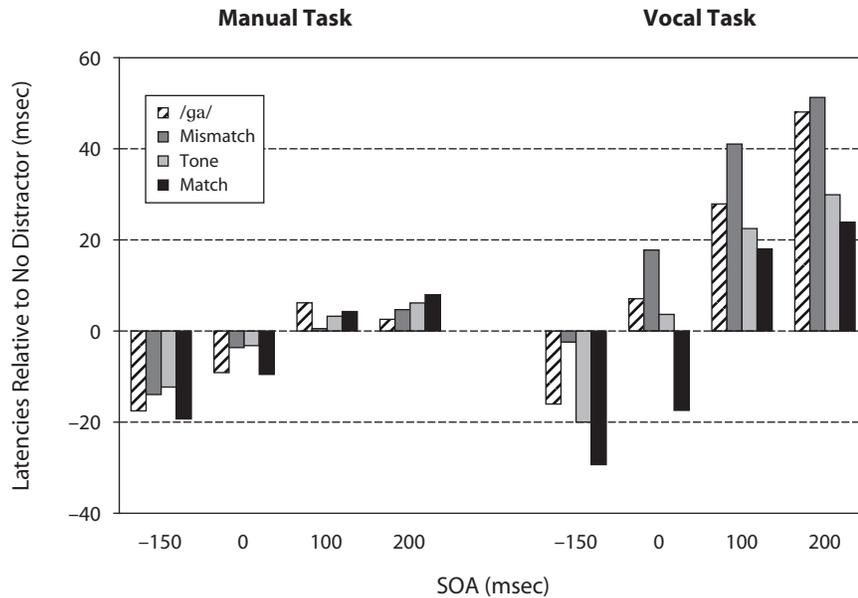


Figure 3. Experiment 2: Differences in latencies between the no-distractor condition and the other conditions. SOA, stimulus onset asynchrony.

### EXPERIMENT 3

Experiment 3 was designed to detect pure perceptuo-motor effects—that is, stimulus–response compatibility effects that occurred in isolation from selection effects. In particular, similar to Experiment 3 of Kerzel and Bekkering (2000), we eliminated the element of choice in the response process. Kerzel and Bekkering did so by presenting the response cues at least 1 sec before the go signals for the responses. We opted for a different procedure. During an experimental session, participants responded on every trial by *always* producing the same syllable—in separate sessions, either /ba/ or /da/. The distractors and SOAs were the same as those used in Experiment 2. The logic of Experiment 3 is straightforward. The effects of the distractors cannot be due to selection effects, because no response selection occurs in a simple reaction time task such as ours. Moreover, stimulus–stimulus compatibility effects were ruled out by Experiment 2. Therefore, if response latencies in Experiment 3 are shorter on match trials than those on mismatch trials, we can safely conclude that the difference is due to a pure perceptuo-motor effect (cf. Fowler, Brown, Sabadini, & Weihing, 2003), the effect specifically predicted by the motor theory of speech perception.

#### Method

**Participants.** Twenty-four undergraduate students at the University of Connecticut participated in the study. Participants were self-reported native speakers of American English with no known speech or hearing disorders. Each received course credit for approximately a half hour of participation.

**Stimuli.** The materials were the same as those used in Experiment 2.

**Design.** Participants took part in two experimental sessions. In each session, participants were exposed to 12 repetitions of the combinations of the five distractors and the four SOAs, yielding 240 tri-

als. Half of the participants began with a session in which they were instructed to say /ba/ when presented with the cue ##, followed by a session in which they were instructed to say /da/ when presented with the same cue. For the other half of the participants, the order of the two sessions was reversed.

**Procedure.** The procedure was the same as that in Experiment 2, except that participants were instructed to always respond with the same syllable. After completion of the first session, participants were told that they would participate in a second session with a different syllable. For both sessions, prior to the beginning of the experimental trials, participants were exposed to 24 practice trials. During these trials, there were no distractors.

#### Results

Latencies for trials in which an erroneous response was produced were excluded from the analyses.<sup>8</sup> This restriction excluded 231 responses out of 11,520 (2%). Also, for each participant and each condition, latencies that were outside a range of 5 standard deviations centered on the mean for that condition were excluded from the analysis. This restriction excluded 101 responses out of 11,520 (0.8%).

For consistency with previous analyses, a mismatch condition was created. For /ba/ responses, this condition was obtained by averaging latencies for responses with /da/ as a distractor and those for responses with /ga/ as a distractor. For /da/ responses, the condition was obtained by averaging latencies for responses with /ba/ as a distractor and those for responses with /ga/ as a distractor. Figure 4 presents the results of the experiment. The latencies were submitted to an ANOVA with two within-subjects factors, SOA (-150, 0, 100, or 200) and distractor (match, mismatch, tone, or no distractor).

The main effects of SOA and distractor were significant, as was their interaction [SOA,  $F(3,69) = 116.6, p < .001, \eta^2 = .83$ ; distractor,  $F(3,69) = 64.1, p < .001, \eta^2 = .74$ ; SOA  $\times$  distractor interaction,  $F(9,207) = 51.2, p < .001$ ,

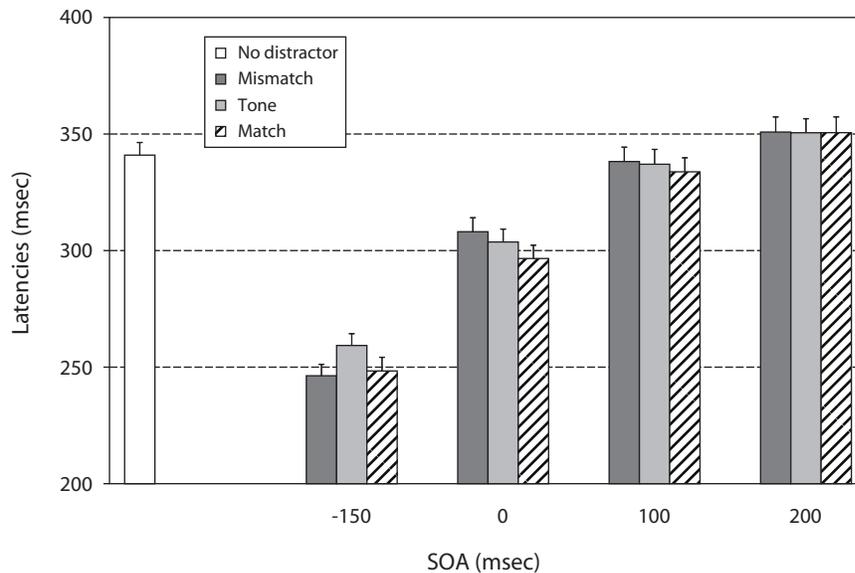


Figure 4. Experiment 3: Mean latencies for each of the experimental conditions (for the no-distractor condition, there was no stimulus onset asynchrony [SOA] manipulation).

$\eta^2 = .69$ ]. As for the main effect of SOA, performance at every SOA was significantly different from that at every other SOA following a pattern that we have come to know well: The closer in time the distractor and response were, the slower the latency of the response was.

As for the main effect of distractor, comparisons among the conditions revealed that response latencies in the match condition were significantly shorter than those in every other condition (see Table 4 for details), and latencies in the no-distractor condition were significantly longer than those in every other condition. Particularly interesting for us was the difference between latencies in the match condition and those in the mismatch condition, which was small but significant across SOAs (Table 4) and, when the data were split by SOA, at SOA 0 [ $M = 11$  msec,  $t(23) = 2.9$ ,  $p < .01$ ,  $d = .2$ ].

## Discussion

Experiment 3 confirms the presence of isolated perceptuomotor effects, as detected at SOA 200 in Experiment 2. In other words, Experiment 3 suggests that Experiment 2, as well as the previous experiments by us and

by Kerzel and Bekkering (2000), reflected the presence of pure perceptuomotor effects, although combined with selection effects.

However, in Experiment 3, the magnitude of the perceptuomotor effect is much smaller than it was in the previous experiments, and its temporal location seems to be shifted with respect to the outcome in Experiment 2. (Pure perceptuomotor effects appeared at SOA 200 in Experiment 2 and at SOA 0 in Experiment 3.) For both facts, we offer an explanation based on the observation that response times were much faster in the simple task of Experiment 3 (average latency =  $318 \pm 97$  msec) than were those in the choice task of Experiment 2 (average latency =  $476 \pm 120$  msec). When responses are fast in simple response tasks, effects of perceiving speech on producing it tend to be rather small (Fowler et al., 2003). And at SOA 0, the SOA at which perceptuomotor effects clearly occurred in Experiment 3, the magnitude of the effect (11 msec) is compatible with the results by Fowler et al.

As for the shift in location of the perceptuomotor effect, it is also explained by the faster latencies in Experiment 3. In Experiment 2, we have evidence for pure perceptuomotor effects only at SOA 200. The average response latency for that SOA was about 500 msec. Considering that the distractors lasted 150 msec, pure perceptuomotor effects occurred when the distractors were perceived in a time window ranging from 300 to 150 msec before the response. In Experiment 3, we have evidence for pure perceptuomotor effects only with an SOA of 0 msec. Considering that the average response latency for that SOA was about 300 msec and that the distractors again lasted 150 msec, perceptuomotor effects occurred when the distractors were perceived in the time window that ranged from 300 to 150 msec before the response, exactly as in Experiment 2. This time window is fully consistent with the timing of cortical activation. On the perceptual

Table 4  
Experiment 3: Comparisons Among Distractor Conditions  
Across Stimulus Onset Asynchronies (SOAs)

Comparison Conditions	Mean Difference in Latencies (msec)	$t(23)$	$p$	$d$
Match – mismatch	–4	–2.16	<.050	–.06
Match – tone	–5	–2.25	<.050	–.09
No distractor – mismatch	30	11.11	<.001	.55
No distractor – match	34	8.96	<.001	.61
No distractor – tone	28	8.48	<.001	.53
Mismatch – tone	–2	–0.88	.388	–.03

side, vocally induced activation of the motor cortex occurs about 110 msec after the onset of speech stimuli (Fadiga et al., 2002). On the motor side, the activation of the motor cortex precedes the actual production of the response by about 150 msec (Gunji, Kakigi, & Hoshiyama, 2000). Combined, these studies lead to the prediction that pure perceptuomotor effects should be stronger when the speech distractors occur about 260 msec before the onset of the response—that is, within the time window in which we found clear evidence for pure perceptuomotor effects in Experiments 2 and 3.

## CONCLUSIONS

Taken together, Experiments 1–3 provide evidence in favor of the distinctive claim of the motor theory of speech perception that the motor system is recruited for perceiving speech.

Experiment 1 confirmed that the perceptuomotor effects found by Kerzel and Bekkering (2000) extend to the auditory modality. It also demonstrated that a mismatch in one vocal gesture can neutralize the facilitatory effect of a match in another vocal gesture.

Experiments 2 and 3 confirmed, in ever more stringent conditions, the existence of perceptuomotor compatibility effects in speech. In particular, Experiment 2 demonstrated that perceptuomotor effects cannot be explained by stimulus–stimulus compatibility or selection effects, whereas Experiment 3 demonstrated that perceptuomotor effects can be detected in isolation from selection effects.

Across different experiments, with different materials and designs, Kerzel and Bekkering (2000) and we consistently found the same pattern of results: Perceiving syllables, by eye or by ear, affects the production of syllables, as predicted by the motor theory. The results of these two compatibility studies are consistent with the results of other studies with vocal tasks (Bell-Berti, Raphael, Pisoni, & Sawusch, 1978; Cooper, 1979; Fowler et al., 2003; Kerzel & Bekkering, 2000; Porter & Lubker, 1980). Moreover, the results of our Experiments 2 and 3 are fully compatible with the neural evidence about speech perception that we presented in the introduction (for a review, see Iacoboni, 2008). Not only did we find the same overall effects of perception on the motor system as in the neural studies, but the temporal course of such effects is also exactly the temporal course one would predict by looking at the neural activations.

The claim of the motor theory that the motor system is recruited for perceiving speech represents today a simple unifying explanation for a large number of empirical results related to speech perception. More generally, the simple explanation offered by the motor theory for perceptuomotor effects in speech is consistent with a general trend in contemporary cognitive science (Galantucci et al., 2006). In the last 20 years, the idea that the motor system plays a role in perception has been suggested by a number of scholars in different fields (e.g., Prinz, 1990; Rizzolatti & Craighero, 2004; Viviani, 2002) to explain perceptuomotor interactions that occur in non-speech-related tasks, at a behavioral level (e.g., Stürmer, Aschersleben, & Prinz,

2000; Viviani, Baud Bovy, & Redolfi, 1997), as well as at a physiological level (e.g., Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Grafton, Arbib, Fadiga, & Rizzolatti, 1996; Iacoboni et al., 1999; Strafella & Paus, 2000). In other words (and contrary to the expectations of motor theorists, e.g., Liberman & Mattingly, 1985), the claim of the motor theory of speech perception that perceiving speech implies the activation of the motor system appears to be an expected consequence of a much broader design feature of cognition.

## AUTHOR NOTE

We thank Julie Brown and Jeffrey Weihing for help in collecting the data. Preparation of the manuscript was supported by NICH Grant HD-01994 and NIDCD Grant DC-03782 to Haskins Laboratories. Correspondence concerning this article should be addressed to B. Galantucci, Department of Psychology, Yeshiva University, 2495 Amsterdam Avenue, New York, NY 10033 (e-mail: galantuc@yu.edu).

## REFERENCES

- BELL-BERTI, F., RAPHAEL, L. J., PISONI, D. B., & SAWUSCH, J. R. (1978). Some relationships between speech production and perception. *Phonetica*, **36**, 373-383.
- COHEN, J. D., MACWHINNEY, B., FLATT, M., & PROVOST, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments, & Computers*, **25**, 257-271.
- COOPER, W. E. (1979). *Speech perception and production: Studies in selective adaptation*. Norwood, NJ: Ablex.
- 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.
- DIEHL, R. L., & KLUENDER, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, **1**, 121-144.
- DI PELLEGRINO, G., FADIGA, L., FOGASSI, L., GALLESE, V., & RIZZOLATTI, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, **91**, 176-180.
- DOYLE, M. C., & SNOWDEN, R. J. (2001). Identification of visual stimuli is improved by accompanying auditory stimuli: The role of eye movements and sound location. *Perception*, **30**, 795-810.
- FADIGA, L., CRAIGHERO, L., BUCCINO, G., & RIZZOLATTI, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, **15**, 399-402.
- FADIGA, L., FOGASSI, L., PAVESI, G., & RIZZOLATTI, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, **73**, 2608-2611.
- FERRARI, P. F., GALLESE, V., RIZZOLATTI, G., & FOGASSI, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, **17**, 1703-1714.
- FOWLER, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, **14**, 3-28.
- FOWLER, C. A., BROWN, J. M., SABADINI, L., & WEIHING, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory & Language*, **49**, 396-413.
- GALANTUCCI, B., FOWLER, C. A., & TURVEY, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, **13**, 361-377.
- GRAFTON, S. T., ARBIB, M. A., FADIGA, L., & RIZZOLATTI, G. (1996). Localization of grasp representations in humans by positron emission tomography. *Experimental Brain Research*, **112**, 103-111.
- GUNJI, A., KAKIGI, R., & HOSHIYAMA, M. (2000). Spatiotemporal source analysis of vocalization-associated magnetic fields. *Cognitive Brain Research*, **9**, 157-163.
- IACOBONI, M. (2008). The role of premotor cortex in speech perception: Evidence from fMRI and rTMS. *Journal of Physiology, Paris*, **102**, 31-34.

- IACOBONI, M., WOODS, R. P., BRASS, M., BEKKERING, H., MAZZIOTTA, J. C., & RIZZOLATTI, G. (1999). Cortical mechanisms of human imitation. *Science*, **286**, 2526-2528.
- KERZEL, D., & BEKKERING, H. (2000). Motor activation from visible speech: Evidence from stimulus response compatibility. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 634-647.
- KOHLER, E., KEYSERS, C., UMILTÀ, M. A., FOGASSI, L., GALLESE, V., & RIZZOLATTI, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, **297**, 846-848.
- KUHL, P. K. (1981). Discrimination of speech by nonhuman animals—Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America*, **70**, 340-349.
- LEVELT, W. J. M., SCHRIEFERS, H., VORBERG, D., MEYER, A. S., PECHMANN, T., & HAVINGA, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, **98**, 122-142.
- LIBERMAN, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, **29**, 117-123.
- LIBERMAN, A. M. (1996). *Speech: A special code*. Cambridge, MA: MIT Press.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. (1967). Perception of speech code. *Psychological Review*, **74**, 431-461.
- LIBERMAN, A. M., & MATTINGLY, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.
- LIBERMAN, A. M., & MATTINGLY, I. G. (1989). A specialization for speech perception. *Science*, **243**, 489-494.
- LIBERMAN, A. M., & WHALEN, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, **4**, 187-196.
- MACLEOD, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, **109**, 163-203.
- MASSARO, D. W., & ODEN, G. C. (1980). Evaluation and integration of acoustic features in speech-perception. *Journal of the Acoustical Society of America*, **67**, 996-1013.
- MEISTER, I. G., WILSON, S. M., DEBLIECK, C., WU, A. D., & IACOBONI, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, **17**, 1692-1696.
- PORTER, R. J., JR., & LUBKER, J. F. (1980). Rapid reproduction of vowel-vowel sequences: Evidence for a fast and direct acoustic-motoric linkage. *Journal of Speech & Hearing Research*, **23**, 593-602.
- POSNER, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum.
- PRATHER, J. F., PETERS, S., NOWICKI, S., & MOONEY, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature*, **451**, 305-310.
- PRINZ, W. (1990). A common coding approach to perception and action. In O. Neumann & W. Prinz (Eds.), *Relationships between perception and action: Current approaches* (pp. 167-201). New York: Springer.
- PROCTOR, R. W., & REEVE, T. G. (Eds.) (1990). *Stimulus-response compatibility: An integrated perspective*. Amsterdam: Elsevier.
- PULVERMÜLLER, F., HUSS, M., KHERIF, F., MOSCOSO DEL PRADO MARTIN, F., HAUKE, O., & SHTYROV, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, **103**, 7865-7870.
- RIZZOLATTI, G., & ARBIB, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, **21**, 188-194.
- RIZZOLATTI, G., & CRAIGHERO, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, **27**, 169-192.
- STRAFELLA, A. P., & PAUS, T. (2000). Modulation of cortical excitability during action observation: A transcranial magnetic stimulation study. *NeuroReport*, **11**, 2289-2292.
- STÜRMER, B., ASCHERSLEBEN, G., & PRINZ, W. (2000). Correspondence effects with manual gestures and postures: A study of imitation. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 1746-1759.
- SUSSMAN, H. M. (1989). Neural coding of relational invariance in speech: Human language analogs to the barn owl. *Psychological Review*, **96**, 631-642.
- VIVIANI, P. (2002). Motor competence in the perception of dynamic events: A tutorial. In W. Prinz & B. Hommel (Eds.), *Attention and performance XIX: Common mechanisms in perception and action*. Oxford: Oxford University Press.
- VIVIANI, P., BAUD BOVY, G., & REDOLFI, M. (1997). Perceiving and tracking kinesthetic stimuli: Further evidence of motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 1232-1252.
- WATKINS, K. E., STRAFELLA, A. P., & PAUS, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, **41**, 989-994.
- WELFORD, A. T. (1988). Reaction time, speed of performance, and age. *Annals of New York Academy of Science*, **515**, 1-17.
- WILSON, S. M., SAYGIN, A. P., SERENO, M. I., & IACOBONI, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neurosciences*, **7**, 701-702.

## NOTES

1. However, it is perhaps ironic that, in the case of the motor theory of speech perception, the finding is problematic for another claim of the theory: that, in its recruitment of the motor system, speech perception is a special mode of perceiving (Liberman & Mattingly, 1989; for further discussion of this point, see Galantucci et al., 2006).
2. In Experiment 1, participants were recruited from a general population, via fliers posted in public spaces of the city of New Haven. In Experiments 2 and 3, the participants were all undergraduate students—that is, people in their early twenties. The age difference between the two pools is the most likely reason for the sharp overall difference in latencies between Experiment 1 on one side and Experiments 2 and 3 on the other (Welford, 1988).
3. In Experiments 2 and 3, we resorted to voice-key measurements for practical reasons. Because of the large number of conditions, more than 31,000 hand measurements would have been necessary for the analyses of the two experiments.
4. These trials were detected and recorded by the experimenter during the experimental sessions.
5. The factorization of the no-distractor condition with SOA was implemented in the program that ran the experiment via the use of a silent sound file (for reasons of programming consistency) but is obviously meaningless. Hence, in the analyses that follow, the no-distractor condition will be analyzed averaged across the four SOAs.
6. Visual inspection of the results for the manual task suggests that the effect of the distractors continued when the distractors followed the response cue. This was confirmed by a trend analysis, which found a significant linear trend over SOAs [ $F(1,41) = 29.11, p < .001$ ]. The distractors sped up the response process when they occurred well in advance of the responses. When they occurred closer in time to the responses, they slowed the responses down.
7. Although the overall pattern of facilitation and interference between distractors and responses over SOAs was similar across the vocal and the manual tasks, there were two notable differences. First, in the manual task, the effects of the different distractors were undifferentiated, whereas in the vocal task, the effects of the distractors depended on the relation the distractors had with the responses. Second, the facilitatory effect of the distractors seems to extinguish itself at an earlier SOA for the vocal task than for the manual task. This could be due to the fact that manual responses were, on average, 108 msec faster than vocal responses. In other words, responses were considerably closer in time to the distractors in the manual task.
8. As for Experiment 3, these trials were detected and recorded by an experimenter during the experiment.

(Manuscript received February 17, 2008;  
revision accepted for publication January 4, 2009.)