Psychonomic Bulletin & Review
2008, 15 (2), 458-462
doi: 10.3758/PBR.15.2.458

The FLMP STMPed

CAROL A. FOWLER
Haskins Laboratories, New Haven, Connecticut
and University of Connecticut, Storrs, Connecticut

Massaro and Chen (2008) offer a commentary ostensibly on a recent article by Galantucci, Fowler, and Turvey (2006). Our article provided an evaluation of Alvin Liberman's motor theory of speech perception. We considered it timely to evaluate the motor theory's different claims, and we hoped to understand why the theory has been better received outside the field of speech than within it. Accordingly, we evaluated its component claims and embedded our presentation of the theory within a wider scientific context, rather than restricting it to the field of speech. However, Massaro and Chen did not undertake to understand the motor theory or to evaluate the effort by Galantucci et al. Rather, they chose to evaluate a different theory of speech than the one on which Galantucci et al. focused. They evaluated direct-realist theory, erroneously referring to it as a motor theory. I take this opportunity to clarify the difference, staving off potential confusion, and to address other errors in their critique.

Galantucci, Fowler, and Turvey (2006) published a review of Alvin Liberman's motor theory of speech perception (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) with the goals of evaluating its central claims and of situating the theory in a broad scientific context. By doing the latter, we hoped to provide an understanding for ourselves and others why Liberman's theory has had a more positive reception outside the field of speech than within it.

The article by Massaro and Chen (2008) purports to be a commentary on Galantucci et al. (2006) and has the promising title, "The Motor Theory of Speech Perception Revisited." Two parts of the title, however, turn out to be misleading. They are "the motor theory of speech perception" and "revisited."

One can only revisit something that has been visited previously. Massaro and Chen (2008), however, commented on a different theory (my direct-realist theory; e.g., Fowler, 1986), rather than on the subject of Galantucci et al.'s (2006) review—Liberman's motor theory. Revisiting did not occur. Indeed, not only did Massaro and Chen fail to comment on Liberman's motor theory, they failed to comment on any motor theory at all.

How did this error come about? In the recently published *The Oxford Handbook of Psycholinguistics* (Gaskell, 2007), there is an article by Massaro and Jesse that permits rejection of one obvious hypothesis—namely, that Massaro was, in fact, confused. Following the lead of philosophers who might refer, for example, to the early writings of Heidegger as those of Heidegger_1 and the later writings as those by Heidegger_2 (to indicate a change or development in the writer's thought), I will refer to the first author of the *Handbook* article as Massaro_1 and the first author of the commentary as Massaro_2. It is clear that substantial intellectual morphing has occurred in the (approximately) 6 months that separated the writing of these two articles. In their article on audiovisual speech perception, Massaro_1 and Jesse (2007) include the following two headings: "2.3.2 Motor Theory" and "2.3.3 Direct Perception." (I include the heading numbering to indicate that "Direct Perception" is not a subhead under "Motor Theory.") Under the first heading, Massaro_1 and Jesse cite Liberman and colleagues and describe Liberman's motor theory; they do not cite my theory. Under the second heading, they cite me and fellow direct realist Catherine Best (e.g., 1995). They describe direct realism, not confusing it with a motor theory. But in the present commentary, titled "The Motor Theory of Speech Perception Revisited," Massaro_2 and Chen do not discuss Liberman's theory. Instead, they discuss direct realism, referring to it not as "direct realism," but as "the motor theory." What happened? Whatever the answer may be, because Massaro_2 and Chen's ostensible commentary can only spread confusion, I take on the task of damage control.

The Motor Theory Versus Direct Realism

As reviewed by Galantucci et al. (2006), Liberman and colleagues obtained findings (e.g, Liberman, Delattre, & Cooper, 1952; Liberman, Delattre, Cooper, & Gerstman, 1954) showing that listeners' speech percepts track articulation more closely than they do the acoustic signal. To accommodate these results, Liberman (e.g., Liberman et al., 1967) proposed that the speech motor system intervenes in the achievement of gestural speech percepts. Those are the bare bones of the motor theory of speech perception.

By contrast, the direct-realist theory of speech perception (e.g., Fowler, 1986), derived from Gibson's (e.g., 1966) more general theory of direct perception, is almost allied with the motor theory in just one respect; otherwise, the theories are not allied. Proponents of the theories agree that listeners to speech perceive linguistically significant articulatory actions. There are disagreements even there, however. Liberman and Mattingly (1985) concluded that the destructive consequences of coarticulation ensure that perceived gestures are "intended," not actual. In direct-realist theory, coarticulation is not

C. A. Fowler, carol.fowler@haskins.yale.edu

destructive, and perceived gestures, therefore, are both intended and actual. Relatedly, the theories disagree regarding whether acoustic information fully specifies speech gestures.

More relevant in the present context, the theories disagree on the claim that makes the motor theory, in fact, a motor theory. Liberman's theory explains perception of gestures by invoking speech–motor involvement; I have rejected that claim for direct realism (e.g., Fowler, 1996). Accordingly, direct-realist theory is not a motor theory. The article by Galantucci et al. (2006) was about Liberman's motor theory. Massaro and Chen's (2008) commentary was not about any motor theory at all.

The following analogy may clarify the distinction between theories that are, or are not, motor theories. Imagine a theory of visual perception that makes the uncontroversial claim that an observer, looking at a person walking, sees a person walking. That is, the theory's claim is that viewers perceive motor action when there is motor action in their line of sight. They do so because they intercept structure in reflected light over time that informs about the motor actions that caused that structure. This theoretical claim is exactly analogous to a claim of the direct-realist theory of speech perception: Listeners perceive speech actions when there are actions to be perceived, because the actions structure acoustic signals (and optic arrays, oftentimes), which, in turn, inform about the actions.

Neither theory is a motor theory. The visual theory would be a motor theory only if there were an additional claim that viewers see locomotion because their own locomotor systems intervene in the perception of walking. Likewise, the direct-realist theory of speech perception would be a motor theory only if it claimed that the speech motor system is involved in speech perception, but it is explicit (and likely wrong) in rejecting that idea. Indeed, if a claim that speech gestures are perceived were to mean that a theory of speech perception was a motor theory, then Massaro's fuzzy logic model of perception (FLMP) would be a "half motor" theory. In the FLMP, for example, a visual "cue" for the syllable /ba/ is identified as "lips closed," gestural information. But is the FLMP a half motor theory? Of course not.

Massaro and Chen (2008) attempt to camouflage a discussion of direct realism in the guise of discussing motor theory. Consider the following: "The second claim [of the motor theory], that perceiving speech is perceiving gestures, is grounded in Gibson's view of direct perception" (p. 453). As written, this remark is false, for three reasons. First, Liberman and colleagues first hinted at a motor theory in 1952:

> Our results relate to the assumption that the perception of speech depends ultimately on the proprioceptive return from the articulatory movements which are made in speaking.
>
> (Liberman et al., 1952, pp. 512–513)

Gibson, however, did not present his theory of direct perception until 1966.

Second, Liberman and colleagues finally settled on analysis by synthesis as the mechanism by which motor involvement occurs in speech perception (e.g., Liberman et al., 1967; Liberman & Mattingly, 1985). Analysis by synthesis is not direct perception.

Third, Liberman and Mattingly (1985) openly rejected Gibson's theory, at least as applied to speech:

> Unlike Gibson, we do not think articulatory movements (let alone phonetic structures) are given directly (that is, without computation) by 'higher order invariants' that would be plain if only we had a biologically appropriate science of physical acoustics. . . . No higher-order invariants have thus far been proposed, and we doubt that any will be forthcoming.
>
> (p. 26)

Again, Massaro and Chen's (2008) remark, as written, is false. However, if their "motor theory" is replaced with "direct-realist theory," it is true.

Another instance of camouflage from Massaro and Chen (2008) appears on page 456: "One of our biggest concerns about the motor theory is that seldom do its theorists describe how gesture and motor processing actually solve perceptual outcomes." This statement, if directed, in fact, at motor theorists, is wrong.

I do not know what it means to "solve . . . outcomes," but Liberman and his colleagues *did* offer explanations of "how gesture and motor processing" are involved in speech perception. As Galantucci et al. (2006) noted, they proposed analysis by synthesis, and they proposed correlated neural networks supporting perception and production (see Liberman et al., 1967; Liberman, Cooper, Studdert-Kennedy, Harris, & Shankweiler, 1968; Liberman & Mattingly, 1985). Liberman and colleagues also offered a rationale for motor involvement in speech perception. In their view, efficient speech production requires a specialization for coarticulation, because human auditory systems cannot handle acoustic alphabets when their component sounds are sequenced at practically useful rates. Perception of speech requires a specialization for decoding coarticulated speech. Liberman et al. (1967; Liberman & Mattingly, 1985; Liberman & Whalen, 2000) inferred from evidence that listeners perceive articulation that the specializations for talking and listening are one and the same. One can disagree with those proposals—I do, mostly—but it is inaccurate to suggest they were not provided (often enough).

Although Massaro and Chen's (2008) remark would still be far off the mark if "direct-realist theory" replaced "motor theory" in the quotation above, it would be perhaps more understandable. Substitution in place, Massaro and Chen should not have "big concerns" here, but it is true not only that direct realists "seldom" explain how motor processing is involved in speech perception; they never do it. That is, in part, because direct-realist accounts are not about mechanism or mental processing. They are about *public perceptual function*—that is, about the functions that perception serves in the life of an animal (e.g., for Gibson, 1966, in the general case,

perception of affordances, perceptual guidance of action; cf. Fowler, 1986, for speech). It is also, in part, because direct-realist Fowler has denied that motor system involvement in perception likely occurs (Fowler, 1996). So it would be surprising if, either frequently or seldom, she offered an account about how motor system involvement were to take place, if only it did.

Gesture Perception in Speech

There is a domain in which motor theorists and direct realists almost agree—namely, the set of findings showing that listeners perceive speech gestures, intended and/or actual. When Massaro and Chen (2008) address this domain, their comments are at least relevant to the article by Galantucci et al. (2006), but they reflect misunderstandings and misrepresentations of the evidence.

Most notably: "The first [piece of evidence for gesture perception invoked by Galantucci et al., 2006] is the lack of signal–phoneme invariance in auditory speech, illustrated by the well-known /di/–/du/ schematic spectrograms" (p. 453). However, the import of /di/–/du/ for Liberman and colleagues (e.g., Liberman et al., 1954) went well beyond the lack of invariance in the synthetic acoustic signals. That bit of negative evidence would not lead a scientist of Liberman's caliber to a motor theory. It was the *lack* of acoustic invariance, coupled with the *presence* of invariance in the gesture for /d/ across coarticulatory contexts, *with perception tracking articulation.* As Galantucci et al. point out, Liberman did not set out to be a motor theorist. His data led him there.

Massaro and Chen (2008) comment that, because articulation causes the acoustic speech signal, articulations cannot be more invariant than the signals they cause. However, this is a (surprising) mistake. The tongue tip gesture for /d/ in /di/ and /du/ is the same in the two syllables. The second-formant transitions (for example) are not the same, because of coarticulation of the tip gesture with the tongue body and lip gestures for /i/ and /u/. Because the vowels are different, their tongue body and lip gestures are different, and the consequent acoustic signals are different. There is gestural and perceptual invariance for /d/, but there is acoustic context sensitivity.

As I have commented elsewhere (Fowler, 1999), Massaro's alternative proposal (i.e., that syllable prototypes solve the invariance problem) has to be wrong. For example, in this account, why do perceivers judge that /di/ and /du/ share their initial consonants? Liberman and fellow motor theorists had an answer: Their consonantal gestures are the same, and listeners perceive those gestures. In the FLMP, however, the prototypes for /di/ and /du/ are no more closely linked than are the prototypes for /di/ and /gu/. On the related question of how the FLMP's prototypes get their names (that is, e.g., how a particular collection of acoustic and visual cues get the name /di/), the FLMP is STMPed.

Next, consider /r/. In one place in their article, Massaro and Chen (2008) claim that there can be no dissociation between articulatory and acoustic variability. But in their discussion of /r/, in contrast, they suggest that, whereas there are variable tongue shapes for /r/, there is a reliable acoustic signature, a low *F*3. This is one of many tangents in Massaro and Chen's commentary ("How does the FLMP account for language acquisition?" [p. 457]). Their comments about /r/ concern the targets of speech *production.* The article by Galantucci et al. (2006), on which they purport to be commenting, is about speech *perception.*

As I have noted elsewhere (Fowler, 2003), remarkably, /r/ is a consonant on which two opposing classes of speech production theory pay special attention, the proponents of each considering it to provide strong evidence in favor of their account. The classes of theory are the one espoused by Massaro and Chen (2008), that talkers aim for acoustic targets, and the alternative, that they aim to produce gestural signatures. In the theory of articulatory phonology (e.g., Browman & Goldstein, 1986, 1992; Goldstein & Fowler, 2003), the relevant observation is that, regardless of tongue shape, /r/ has three constriction gestures—at pharyngeal, palatal, and labial locations—and these are the goals of /r/ production. So there is gestural invariance for /r/. There is much more that can be written about the merits of a gestural account of /r/ (summarized in Fowler, 2003). However, again, the discussion of /r/ is a tangent, because it does not relate to the *motor theory of speech perception*.

In considering an alternative to gestures, Massaro and Chen (2008) note that "to solve the invariance problem between acoustic signal and phoneme . . . Massaro (1972) proposed the open [*sic*] syllable V, CV, or VC as the perceptual unit" (p. 454). That may be, but why has Massaro not, in the ensuing three and a half decades, ever noticed the implausibility of this proposal? Why has he not addressed the question of where those alleged perceptual units are in upstanding English words such as *strength* and *script*, both CCCVCC monosyllables? The FLMP is STMPed again. Across the range of real languages, V, CV, and ("or"?) VC syllables cannot be perceptual units.

In any case, the proposal of these units does not "solve" the invariance problem. Vowel-to-vowel coarticulation occurs widely and extensively in languages (e.g., Öhman, 1966; Recasens, 1984), and it is sufficiently perceptually salient to have supported development of vowel harmony in the phonologies of several of them. How does this occur? The FLMP is STMPed. Moreover, listeners use context sensitivity, as is expected of gesture perceivers (see Fowler, 2006, for a review). Context sensitivity due to coarticulation does not constitute noise to be tolerated; it constitutes information to be exploited (see also Elman & McClelland, 1986).

All of that aside, listeners are also talkers and users of public language more generally. Talkers show by their spontaneous errors of speech production (e.g., Dell, 1986) that consonants and vowels are (permutable) planning units. Syllables, including the "perceptual units" of Massaro (1972), are not. Moreover, as Abler (1989) and Studdert-Kennedy (e.g., 1998) have noted, these "particulate" units of language, consonants and vowels, are hallmarks of its generativity at the level of lexical form. It would be astonishing if listeners did not use units that (1) persist in language, (2) persist because they do important work, and (3) are units for talkers. The FLMP implies

that language speakers become wholly other people when they are language perceivers. In contrast, motor theorists recognize the necessity of a "parity" constraint on successful communication by speech (e.g., Liberman & Whalen, 2000): What count as language forms for talkers and listeners must coincide.

I stand by my demonstration of a tactile McGurk effect (Fowler & Dekle, 1991), cited by Galantucci et al. (2006) as evidence for gesture perception. Our results are real; readers can count on them. I leave the "more valid" use of a fake speaker to Massaro and Chen. Their experiment will certainly be less icky than ours was.

Another kind of evidence in favor of gesture perception, cited by Galantucci et al. (2006), is Fowler, Brown, Sabadini, and Weihing's (2003) finding of an unusually small latency difference between simple and choice speech tasks. Contrary to Massaro and Chen's (2008) suggestion, in the choice response task, there is no chance that the small latency increment over the simple response times occurred because talkers initiated what looked like acoustic evidence of consonant closure without making a real closure. We measured formant transitions into the closure, and the transitions were specific to the /p/, /t/, or /k/ response that the participants were heard to produce. That information is in the published article. Moreover, I reminded the authors about it in a review of their commentary. If Massaro and Chen cannot find sincere ways to contest evidence in favor of gesture perception, perhaps they need to reevaluate their opposition to the claim.

One would think, too, that if Massaro and Chen (2008) found our account of the different latencies as measured from closure and release "unconvincing," they would have presented it. But they chose not to let readers decide, electing instead to eliminate the results from their consideration. (We explained the relatively slow releases on choice trials by participants' lack of confidence in the consonants they had committed to. As a pilot participant, I can testify to that.) In short, none of the evidence favoring the motor theory's claim of gesture perception is challenged effectively by Massaro and Chen.

Other Points

Beyond their comments on (and off) the topics of motor theory and gesture perception, Massaro and Chen (2008) offer several erroneous arguments concerning the role of motor systems in perception. Their comments on the "degrees-of-freedom problem" getting in the way of motor system intervention in speech perception are just puzzling: Massaro and Chen knew about the article that I cite at the end of this paragraph. I told them about it in two reviews of their commentary. The degrees-of-freedom problem does not apply to skilled actors. They have solved that problem. Indeed, they exploit degrees of freedom (e.g., Bernstein, 1935/1967). In any case, the evidence *shows* that there is online motor system activation in speech perception that is specific to the speech gestures being heard (e.g., Fadiga, Craighero, Buccino, & Rizzolatti, 2002). As such, there is not much point in speculating that such activation would impede perception.

The authors' comments on mirror neurons are equally puzzling: "Our understanding, however, is that mirror neurons cannot account for perception" (p. 456). Did Galantucci et al. (2006) suggest that mirror neurons account for perception? Not even almost. (Also, cf. Goldstein & Fowler, 2003, where we address limits on the explanatory power of mirror neurons.) However, mirror neurons *are* active during perception, including speech perception (Fadiga et al., 2002). I know of no theory of speech perception, other than the motor theory, that predicts this outcome. Even so, it is not the burden of mirror neurons, in anyone's view as far as I know, to "account for perception."

Galantucci et al. (2006) reviewed the history and scientific standing of Liberman and colleagues' motor theory, to keep its potential contributions alive, despite the death of its main proponents. We welcome commentaries on our effort.

AUTHOR NOTE

Preparation of this article was supported by NICHD Grant HD-01994 and NIDCD Grant DC-03782 to Haskins Laboratories. I thank Bruno Galantucci for commenting on an early version of the manuscript. Thanks to the Arthurs, Glenberg and Samuel, for their helpful reviews. Correspondence concerning this article should be addressed to C. A. Fowler, Haskins Laboratories, 300 George Street, Suite 900, New Haven, CT 06511 (e-mail: carol.fowler@haskins.yale.edu).

REFERENCES

Abler, W. (1989). The particulate principle of self-diversifying systems. *Journal of Social & Biological Structures*, **12**, 1-13.

Bernstein, N. (1967). *The co-ordination and regulation of movements.* New York: Pergamon. (Original work published 1935)

Best, C. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language research* (pp. 167-200). Timonium, MD: York.

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

Browman, C., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, **49**, 155-180.

Dell, G. S. (1986). A spreading-activation theory of retrieval in speech production. *Psychological Review*, **93**, 283-291.

Elman, J., & McClelland, J. (1986). Exploiting lawful variability in the speech wave. In J. Perkell & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 360-385). Hillsdale, NJ: Erlbaum.

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*, **25**, 399-402.

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. (1996). Listeners do hear sounds not tongues. *Journal of the Acoustical Society of America*, **99**, 1730-1741.

Fowler, C. A. (1999). A universal algorithm for pattern recognition? *Contemporary Psychology*, **44**, 12-14.

Fowler, C. A. (2003). Speech production and perception. In A. Healy & R. Proctor (Eds.), *Handbook of psychology: Vol. 4. Experimental psychology* (pp. 237-266). New York: Wiley.

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics*, **68**, 161-177.

Fowler, C. A., Brown, J., 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**, 296-314.

Fowler, C. A., & Dekle, D. (1991). Listening with eye and hand:

Crossmodal contributions to speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 816-828.

GALANTUCCI, B., FOWLER, C. A., & TURVEY, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, **13**, 361-377.

GASKELL, G. (2007). *The Oxford handbook of psycholinguistics*. Oxford: Oxford University Press.

GIBSON, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

GOLDSTEIN, L., & FOWLER, C. A. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 159-207). Berlin: Mouton de Gruyter.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431-461.

LIBERMAN, A. M., COOPER, F. S., STUDDERT-KENNEDY, M., HARRIS, K. S., & SHANKWEILER, D. P. (1968). On the efficiency of speech sounds. *Zeitschrift für Phonetik, Sprachwissenschaft & Kommunikationsforschung*, **21**, 1-21.

LIBERMAN, A. M., DELATTRE, P., & COOPER, F. S. (1952). The role of selected stimulus variables in the perception of the unvoiced-stop consonants. *American Journal of Psychology*, **65**, 497-516.

LIBERMAN, A. M., DELATTRE, P., COOPER, F. S., & GERSTMAN, L. (1954). The role of consonant–vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General & Applied*, **68**, 1-13.

LIBERMAN, A. M., & MATTINGLY, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.

LIBERMAN, A. M., & WHALEN, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, **4**, 187-196.

MASSARO, D. (1972). Preperceptual images, processing time and perceptual units in auditory perception. *Psychological Review*, **79**, 124-145.

MASSARO, D. W., & CHEN, T. H. (2008). The motor theory of speech perception revisited. *Psychonomic Bulletin & Review*, **15**, 453-457.

MASSARO, D. W., & JESSE, A. (2007). Audiovisual speech perception and word recognition. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 19-35). Oxford: Oxford University Press.

ÖHMAN, S., (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, **39**, 151-168.

RECASENS, D. (1984). Vowel-to-vowel coarticulation in Catalan VCV sequences. *Journal of the Acoustical Society of America*, **76**, 1624-1635.

STUDDERT-KENNEDY, M. (1998). The particulate origins of language generativity: From syllable to gesture. In M. Studdert-Kennedy & C. Knight (Eds.), *Approaches to the evolution of language* (pp. 202-221). Cambridge: Cambridge University Press.

(Manuscript received August 29, 2007;
revision accepted for publication November 11, 2007.)