Opinion TRENDS in Cognitive Sciences Vol.10 No.8

Are there interactive processes in speech perception?

James L. McClelland[1], Daniel Mirman[2] and Lori L. Holt[1]

[1] Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA
[2] Department of Psychology, University of Connecticut, 406 Babbidge Rd., Unit 1020, Storrs, CT 06269-1020, USA

Lexical information facilitates speech perception, especially when sounds are ambiguous or degraded. The interactive approach to understanding this effect posits that this facilitation is accomplished through bi-directional flow of information, allowing lexical knowledge to influence pre-lexical processes. Alternative autonomous theories posit feed-forward processing with lexical influence restricted to post-perceptual decision processes. We review evidence supporting the prediction of interactive models that lexical influences can affect pre-lexical mechanisms, triggering compensation, adaptation and retuning of phonological processes generally taken to be pre-lexical. We argue that these and other findings point to interactive processing as a fundamental principle for perception of speech and other modalities.

Introduction

Identification of a stimulus is influenced by context, especially if the stimulus is ambiguous or degraded. For example, an ambiguous sound that might be a /g/ or a /k/ is more likely to be identified as a /g/ if followed by 'ift' and as a /k/ if followed by 'iss' [1]. Here we argue that this contextual effect is due to a lexical influence on pre-lexical representations, as predicted by interactive approaches to speech perception [2]. The interactive view predicts that lexical information actually reaches down and reshapes the mental representation of the sound that is heard. This view contrasts with other proposals, in which perceptual processing is seen as a strictly autonomous, bottom-up process, with the influence of lexical and semantic contextual information arising only at a later decision stage [3,4]. Figure 1 shows a schematic representation of information flow in interactive and autonomous models.

The debate about autonomous versus interactive approaches has continued for some time, with arguments and counter-arguments on both sides of the debate. The issues considered here are very general ones, and have a long history in our field. There are general theories in which top-down influences affect all levels of language processing [5], and similar proposals have been offered for the role of top-down or contextual influences in vision [6] and motor action selection [7]. Again this approach contrasts with other proposals, such as those of Marr [8] and Fodor [9], in which separate and impenetrable modules carry out automatic perceptual and motor processes. Although interactive processing remains a controversial proposal in neuroscience as well as psychology, the interactive approach is supported by several findings from neuroscience. We mention three that we think are particularly telling: (i) Bi-directional connections are the rule rather than the exception in connections between areas in the brain [10]; (ii) Inactivating 'downstream' motion sensitive cortex (area MT) reduces sensitivity to motion in 'upstream' visual areas (V1 and V2), as would be expected if MT integrates motion information across the visual field and feeds it back to lower areas [11]; (iii) Illusory contours in Kanizsa figures activate edge-sensitive neurons in low-level visual areas V1 and V2 [12]. This last effect is delayed relative to the direct bottom-up activation of these neurons that occurs for real edges, as though it were mediated by interactive processes distributed across several visual areas [12].

Within the domain of speech perception, which will be our focus here, much of the debate has centered around the interactive TRACE model of speech perception [2], described in Box 1. An early criticism centered on whether the interactive model could adequately characterize the quantitative form of the joint effects of context and stimulus information [13,14]. More recently, another wave of the debate emerged around a set of purported shortcomings of the TRACE model [4]. In our view the TRACE model has been strengthened as a result of this debate. The first wave led to an improved, intrinsically stochastic formulation [15,16]. The second wave has led to a better understanding of details of the model's predictions (e.g. [17]). Together with this, several purported shortcomings [4] of the TRACE model have been addressed (see Box 2). Yet most of the experimental data can be explained equally by interactive or autonomous approaches. This state of affairs raises this question: are there any unique predictions that follow from the interactive approach that would rule out autonomous approaches? The case for interactivity would naturally be further strengthened by evidence consistent with such predictions.

In this article, we focus on a unique prediction that arises from the interactive approach within the domain of speech. According to this approach, lexical influences should penetrate the mechanisms of perception, affecting processes at pre-lexical levels. By pre-lexical levels, we

Corresponding author: McClelland, J.L. (jlm@cnbc.cmu.edu).

 1364-6613/$ – see front matter doi:10.1016/j.tics.2006.06.007

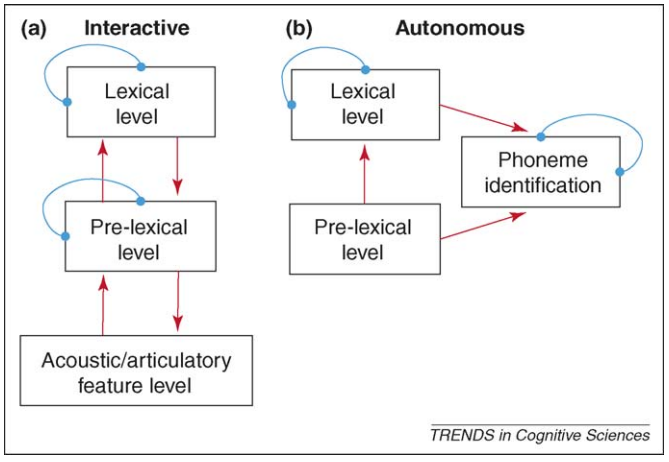

Figure 1. Schematic representations of information flow in interactive and autonomous models of speech perception. **(a)** An interactive model posits bi-directional excitatory connections between processing levels with phoneme-level responses produced at the pre-lexical processing level. Units within each layer compete through mutually inhibitory connections. (See Box 1 for architectural details of the interactive TRACE model.) **(b)** An autonomous model of the sort advocated in [4] posits strictly feedforward excitatory connections from pre-lexical processing to word processing and a separate phoneme identification layer that integrates inputs from phoneme and word processing layers. Units in the word processing and phoneme identification layers compete through mutually inhibitory connections. In both panels red arrows indicate excitatory connections and blue curves indicate inhibitory connections.

mean those levels thought to transform the raw acoustic signal that arrives at the ear into a form suitable for word identification. We will consider three such processes: compensation for local auditory context, selective adaptation and retuning of phonemic categories. We begin with compensation for local auditory context, as this is the case in which the prediction was first tested [18].

Compensation for auditory context

Auditory context influences the perception of both speech and non-speech sounds [19,20]. For example, when an ambiguous sound between /t/ and /k/ is preceded by /s/, it will tend to be heard as /k/; when preceded by /ʃ/, it will tend to be heard as /t/ [21]. This effect has been attributed to compensatory mechanisms embedded in the perceptual machinery that operate on speech sounds before lexical access. One possibility is that they arise from a contrast enhancement process that operates in both speech and non-speech [19]: activation of channels that are sensitive to a particular band of frequencies at one time point enhances responses to neighboring frequencies at neighboring time points. Consistent with this, both speech and non-speech can produce the compensatory effect if energy in appropriate frequency bands is present [19].

Using the interactive TRACE model, Elman and McClelland [18] presented a simulation demonstrating a counter-intuitive prediction of the interactive approach: lexical information about the identity of one speech sound could feed downward into perceptual mechanisms, triggering contrastive perception of neighboring speech sounds. Specifically, they simulated a shift in the perception of an ambiguous /t/-/k/ sound when the identity of the preceding sound was determined by lexical constraints: the same ambiguous /s/-/ʃ/ sound is heard as /s/ if the preceding context is 'Chirstma_' and /ʃ/ if the preceding context is 'fooli_'. As in the case where the /s/ or /ʃ/ was unambiguous, this had the consequence that the perception of the subsequent ambiguous /t/-/k/ sound was shifted towards /k/ in the first case and /t/ in the second (see Figure 2). Pursuant to this simulation, Elman and McClelland proceeded to demonstrate that the predicted effect could indeed be produced experimentally in human subjects. Just as in the simulation, they found a shift in the perception of an ambiguous /t/-/k/ sound towards /k/ when an ambiguous /s/-/ʃ/ sound was preceded by 'Chrisma_' and towards /t/ when it was preceeded by 'fooli_'. Although several controls were included in the original study [18], this result has been questioned by protagonists of autonomy in perception [4,22,23]. It has now been replicated, however, by two independent research groups using new materials that address concerns with the earlier studies [24,25] (Box 3). Overall, the evidence appears to us to support the prediction that lexical context can, indeed, penetrate the mechanisms of perception, as predicted by interactive, but not autonomous, approaches.

Selective adaptation

Repeated presentation of a particular speech sound, for example, /s/, causes a selective adaptation effect, so that identification of an ambiguous sound – for example, one between /s/ and /ʃ/ – shifts away from the repeatedly presented sound /s/ towards the alternative /ʃ/ [26]. Once again, this effect is thought to reflect adaptation of pre-lexical processes. Indeed, non-speech stimuli that share frequency components with a speech sound can produce selective adaptation of speech, supporting the view that selective adaptation affects early, probably pre-lexical, processing stages [27].

Because the interactive approach holds that lexical influences can directly influence pre-lexical processing stages, the prediction follows immediately that it should be possible to influence selective adaptation by varying lexical context. This prediction has been confirmed: selective adaptation can be produced by repeated presentation of an acoustically neutral sound in a disambiguating lexical context (e.g. 'bronchiti(?)', 'arthriti(?)'). In one study [28], the neutral sound was ambiguous between two alternatives (/s/ or /ʃ/), in another study [29] the neutral sound was a noise burst. The findings were that lexical information determined whether the neutral sound selectively adapted the representation of /s/ (when it was presented in /s/-biased lexical contexts such as 'arthriti(?)') or /ʃ/ (when it was presented in /ʃ/-biased lexical contexts such as 'aboli(?)'). These findings support the prediction of the interactive view that lexical information can influence pre-lexical processing in the form of selective adaptation.

If the interactive account is correct, selective adaptation should influence not only phoneme identification, but lexical processes as well. This could be tested as follows: after adaptation with /s/-biasing stimuli like 'arthriti(?)' the ambiguous sound could be placed in a context such as '(?)ip' where either /s/ or /ʃ/ can occur to form a word ('ship' or 'sip'). If indeed the adaptation has affected pre-lexical processes, then there should be an increased tendency to

Box 1. The TRACE model of speech perception

The TRACE model of speech perception is described fully in [2] and a fully-documented implementation is available on the web (http://maglab.psy.uconn.edu/jtrace/) and is described in [45]. The model (Figure I) consists of a feature layer, a phonemic layer and a lexical layer. Each layer consists of a set of simple processing units each corresponding to the possible occurrence of a particular linguistic unit (feature, phoneme, or word) at a particular time within a spoken input. Activation of a processing unit reflects the state of combined evidence within the system for the presence of that linguistic unit. Mutually consistent units on different levels (/k/ as the first phoneme in a spoken word, 'kiss' as the identity of the word) activate each other via excitatory connections, whereas mutually inconsistent units within the same level (/k/ and /g/ as the first phoneme) compete through mutually inhibitory connections. When input is presented at the feature layer, it is propagated to the phoneme layer and then to the lexical layer. Processing proceeds incrementally with between-layer excitation and within-layer competitive inhibition (Figure I). Crucially, excitation flow is bi-directional: both bottom-up (features to phonemes to words) and top-down (words to phonemes to features).

Featural information relevant to speech perception is represented by seven banks of units corresponding to values along each of seven feature dimensions. For example, one feature bank represents the degree of voicing, which is low for unvoiced sounds such as /t/ and /s/ and higher for voiced sounds such as /d/ and /z/. At the phoneme and lexical levels, one unit stands for each possible phoneme or word interpretation of the input. These sets of units and the connections between them are duplicated for as many time slices as necessary to represent the input to the model. Excitatory between-layer connections and inhibitory within-layer connections apply only to units representing speech elements that overlap in time.

The activation level of a unit is a function of its current activation state relative to its maximum or minimum activation level and the net input to the unit. Negative net input drives the unit towards its minimum activation level, positive net input drives the unit towards its maximum activation level and unit activation tends to decay to its baseline rest activation level.

Compensation for co-articulation in the TRACE model was simulated by assuming that activation of phoneme units in one time slice modulated connections from feature to phoneme units in adjacent time slices [2,18]. Recent evidence of cross-influences between speech and non-speech [19,20] suggest that this effect could occur through contrast enhancement across neighboring time points at a processing level shared by speech and non-speech. Such interactions could be implemented by allowing lateral interactions across time slices within the feature level of the TRACE model, and by allowing activation there to be produced by both speech and non-speech input.

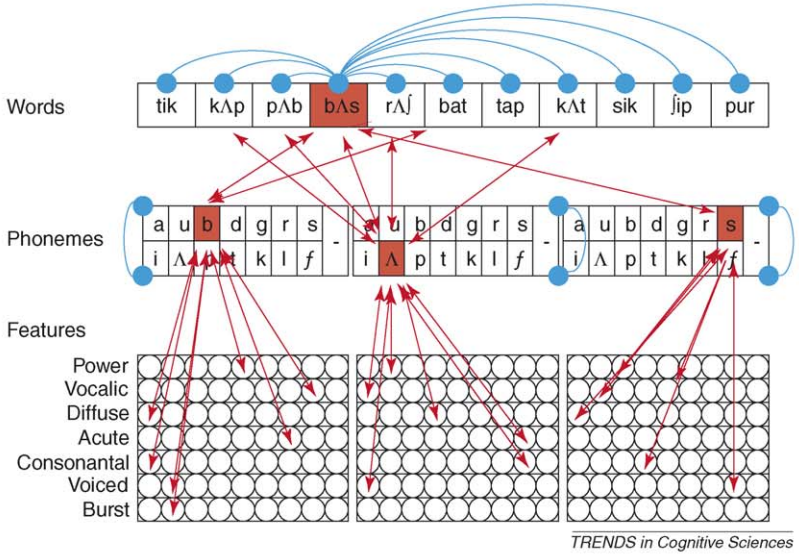

Figure I. Architecture of the TRACE model. Bi-directional excitatory connections are shown in red: mutually consistent elements at adjacent levels support each other through excitation. Units within a layer compete through inhibitory connections (blue; the full set is shown for the lexical layer, for clarity only a schematic connection is shown at the phoneme level).

identify the word as 'ship' in this case. To our knowledge this prediction has not yet been tested.

Tuning of speech perception

The mechanisms of speech perception must be tuned to dialectical and speaker differences. For example, speakers of French and Spanish place the boundary between /b/ and /p/ at a different place from native English speakers. When listening to a native French or Spanish speaker producing English words like 'pore' or 'pier', adjustment of the boundary allows the listener to avoid perceiving these utterances incorrectly as 'bore' or 'beer'. According to the interactive view, we would naturally predict that lexical influences would help guide the retuning of the pre-lexical mechanisms that mediate boundary adjustment. In fact, just such a role for lexical context was explicitly suggested by McClelland and Elman [2]. And indeed, in accordance with this, several recent experiments [30–35] have demonstrated that lexical influences can also guide tuning of speech perception. When listeners heard a perceptually ambiguous /s/-/f/ sound at the end of an utterance that would be a word if completed with /s/, they identified the sound as /s/. Repeated exposure to this sound in /s/-biased lexical contexts retuned perception so that subsequently the sound tended to be heard as /s/ even in lexically neutral contexts. Furthermore, consistent with a pre-lexical locus for this effect, subsequent word identification processes are also affected: the ambiguous sound, when placed in a context where either /s/ or /f/ could make a word, results in lexical activation of the /s/-consistent alternative [35]. These findings follow directly from the interactive framework. Indeed, in simulations, we have shown that the

Box 2. Recent challenges to the TRACE model of speech perception

In a recent critique of interactive processing in general and the TRACE model in particular, Norris *et al.* [4] argued that findings on several specific topics were inconsistent with predictions of the TRACE model. One of these topics – lexically triggered compensation for acoustic context – is discussed extensively in the main text and in Box 3, where we argue that the balance of evidence is consistent with the TRACE model. He we review recent research that has shown that the TRACE model is also consistent with the findings from experiments on several other topics raised by Norris *et al.*

Lexical inhibition of phoneme recognition

Intuitively, the interactive view predicts that phonemes will be recognized faster when they are consistent with their context and slower when they are inconsistent with their context. Two studies [46,47] failed to find any slowing effect for contextually inconsistent phonemes, casting doubt on interactive processing. However, subsequent simulations showed that the TRACE model was consistent with previous failures to demonstrate this effect and correctly predicted the conditions required to show lexical inhibition [17].

Subcategorical mismatch

Initial experiments and TRACE simulations investigating the influence of lexical status on subcategorical mismatch found that TRACE did not fit the behavioral data [48,49]. However, subsequent experiments using eye-movement analysis (a finer grained on-line method) and subsequent TRACE simulations using standard parameters found that the TRACE model was consistent with the behavioral data [50]. Analysis of global model behavior [51] confirmed that TRACE produces the correct behavioral pattern; in fact this analysis showed that TRACE provides a more robust fit to the behavioral data in this case than the autonomous Merge model [4].

Attentional modulation of lexical effects

The impact of lexical information on phoneme processing appears to be modulated by the degree of attention to lexical information ([52–54] and similar effects have been found on speech production [55] and reading [56]). Norris *et al.* [4] argued that to account for this attentional modulation, interactive models would have to turn off feedback, thereby making them autonomous. However, attentional modulation of lexical influences can also be accomplished by modulation of lexical *activity* [54], leaving the interactive architecture in place. In fact, modulation of lexical activity rather than lexical feedback is consistent with findings suggesting that attentional modulation operates by affecting neural responsiveness to input [57–59].

In summary, the behavioral phenomena that led Norris *et al.* [4] to reject the TRACE model specifically, and interactive processing in general, have turned out to be consistent with TRACE and thus with the overall interactive perspective. Together with evidence for direct lexical influence on pre-lexical processes reviewed in the main text and Box 3, we argue that these data show interactive processing to be the most complete and parsimonious account.

inclusion of Hebbian learning [36] in the interactive TRACE model (as previously suggested [2]) produces lexically guided tuning of speech perception [37]. Here the same lexical feedback that influences identification of ambiguous speech sounds provides the guidance for tuning the mapping between acoustic and speech sound representations. The TRACE model also accounts for the pattern of generalization seen in several other studies [33], based on the idea that generalization of the tuning effect will be determined by the acoustic similarity between the learned sounds and novel sounds [37].

Ironically, the experiments demonstrating lexically guided tuning of speech perception were carried out by proponents of the autonomous perspective [30]. Based on other arguments against the interactive approach presented in their earlier work [4], these researchers have proposed that lexical information is propagated to pre-lexical levels for learning, but not for perception. These authors are resourceful in defense of their views, and it seems likely that research and discussion will continue for some time before the matter is finally resolved. Box 2 summarizes our response (presented more fully in other papers) to their earlier arguments against interactive approaches, and Box 3 summarizes our response to their arguments against the accounts offered above for lexically mediated compensation and adaptation effects. Here we simply stress the following point: the interactive approach, by virtue of its inherent assumption that lexical influences penetrate pre-lexical processes, inherently predicts that such effects should occur. Thus it would count as counter-evidence to the approach if such lexical effects did not occur. The situation is quite different for those who still wish to argue that lexical context does not penetrate perception. Here the effect is not independently predicted and it is necessary to add a special mechanism providing lexical information to pre-lexical processes that affects only learning and not processing. To us this is a step that seems to render their approach unparsimonious.

In summary, three very different lines of investigation support the view that lexical information can trigger effects thought to arise in the mechanisms that provide the input to word identification. All these effects follow naturally and simply from the interactive perspective, as implemented in the TRACE model of speech perception.

The status of the TRACE model of speech perception

In the research described above, the interactive TRACE model of speech perception has played a central role, providing a concrete instantiation of the interactive perspective. In closing, we wish to comment briefly on our own view of the model's status. Our views can be cast within the context of Marr's [8] distinction between computational, algorithmic and implementational levels of analysis and Smolensky's [38] distinction between neural and conceptual levels of processing and representation.

At the computational level, one can frame the idea of interactive processing in terms of the goal of finding the optimal perceptual interpretation of a spoken input in terms of linguistic units at several levels of granularity. According to such a framing, the brain is seen as seeking to settle on the most probable ensemble of feature, phoneme, word and larger linguistic units given the perceptual input and knowledge of the probabilistic relations between them. We suggest that TRACE provides an algorithm that, at least to an approximation, allows us to characterize this inference process [16]: the units represent the hypotheses, the connections represent the relations among them, and the units' activations represent the state of the evidence as it unfolds over time during processing. Importantly, however, we see the TRACE model not as the brain's implementation of this process, but as characterizing at a conceptual level a process that is implemented at a neural level using a much more distributed form of representation.

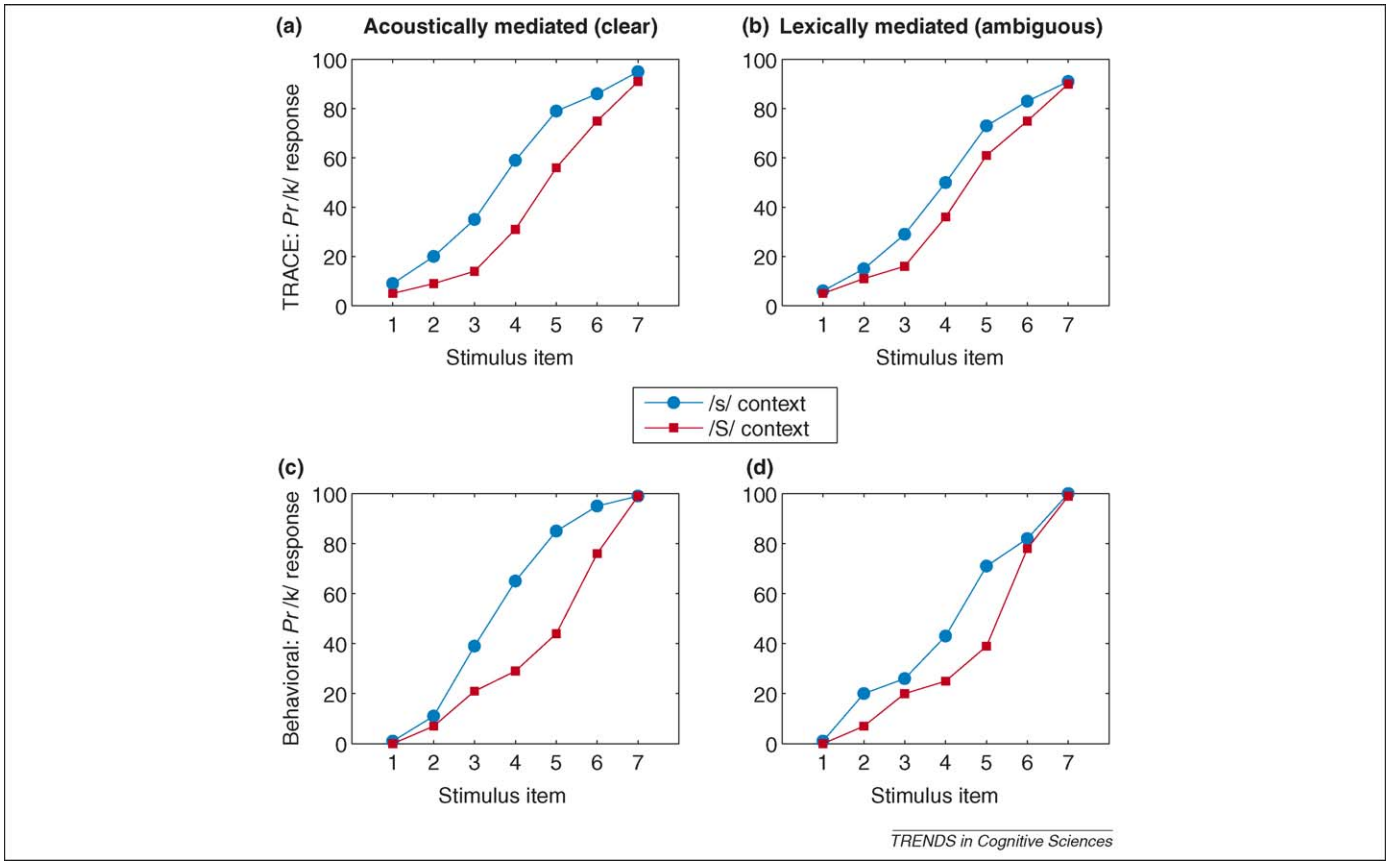

Figure 2. Acoustically and lexically mediated compensation for coarticulation. **(a)** and **(b)** show data from simulations of the TRACE model, and **(c)** and **(d)** show behavioral data from human listeners. Panels **(a)** and **(c)** show that an ambiguous stop will be identified as /k/ more often when it is preceded by a clear /s/ sound than when it is preceded by a /ʃ/ sound (acoustically mediated compensation for coarticulation). Panels **(b)** and **(d)** show that this effect persists (albeit more weakly) when the fricative is replaced by an ambiguous one and lexical context is manipulated. That is, a lexically defined fricative causes a shift in identification of neighboring sounds (lexically mediated compensation for coarticulation). This lexically mediated compensation for coarticulation effect requires that lexical influences act directly on phoneme processing rather than on a post-perceptual decision stage. Data replotted with permission from Ref. [18].

The TRACE model's explicit representation of the evolution over time of the evidence for the various hypotheses is useful, we argue, to the scientist who wishes to explore the implications of the interactive perspective and to generate predictions for perception of specific words, phonemes and features in a particular language. But the model should not be confused with the actual state of affairs existing within the mechanism of speech perception itself. For example, the separation of units in the TRACE model into featural, phonemic and lexical levels should not be seen as indicating the existence in the brain of an explicitly phonemic level of representation or indeed of an explicit lexical level. The actual mechanism might be more similar to recurrent models such as those of Elman [39] and of Gaskell and Marslen-Wilson [40]. The latter model, in particular, has the capacity to allow lexical context to influence pre-lexical representations without explicit representation of lexical or phoneme-level hypotheses. Lexical context directly influences pre-lexical perceptual processing, just as in TRACE, although the implementation forgoes explicit word and phoneme units or word-to-phoneme connections.

Within the context of these ideas, one important goal for future theoretical research is to understand more fully the relationship between the computational, algorithmic/conceptual and neural/implementational levels. In this context the following issues (summarized in Box 4) seem paramount.

Representation of time

One of the challenges for models of speech perception is that speech input unfolds across time. TRACE deals with this by duplicating all units and weights for every segment of time. This simplification allows for computational investigation of many issues in speech perception, but is almost certainly not the way the brain handles the temporal dimension in auditory processing. How can mechanisms that do not directly represent time in this way nevertheless effectively represent the state of evidence for the linguistic units occurring at present, past and future time points?

Representation of linguistic units

The TRACE model incorporates units for features, phonemes and words, corresponding to units that have historically played crucial roles in linguistic theory. There are many good arguments against the idea that any of these units are actually explicitly represented [39,41,42]. Just what then is the content of our mental representations? Are there representations that are specific to speech (as opposed to other sounds), and if so what relation do such representations bear to the feature, phoneme and word levels used in the TRACE model of speech perception?

Box 3. Arguments on lexically mediated compensation and adaptation

Proponents of autonomy have argued against the interactive account of lexically mediated compensation (LMC) and adaptation (LMA) [4,22,23]. We consider what we see as their main points.

Failure to replicate?
Proponents of autonomy note that in one study [22] subjects identified an ambiguous /s/-/ʃ/ stimulus in accord with lexical context, yet there was no compensatory effect on a subsequent ambiguous /k/-/t/ stimulus [22]. However, the words used to produce the lexical bias were short and there was a gap between the ambiguous /s/-/ʃ/ and the ambiguous /k/-/t/. A later study [25] indicates that both word length and the presence of a gap influence the compensation effect. With longer words and no gap, compensation is obtained. In [22] the gap may have weakened the ability of the lexically restored fricative to induce compensation in the subsequent stop.

Transitional probabilities?
A second argument is that compensation is triggered, not by lexical knowledge, but by knowledge of phoneme transitional probabilities (TPs, the probability of a phoneme given one or two preceding phonemes) [22]. This argument cannot apply to Experiment 4 in [18]; there the two phonemes preceding the ambiguous fricative were identical in both lexical contexts. Further, the claim of TP-bias elsewhere in [18] was based on TPs in British English texts. The relevant studies were all done in the US, however, and the apparent bias vanishes if American corpora are used [24]. Finally, the LMC effect was replicated [24] using materials in which the TP-bias ran opposite to lexical constraints. Reliance on higher-order TP biases has been proposed [23], but analyses in [60] suggest that there is no coherent higher-order TP account covering all the relevant effects.

Learning?
Proponents of autonomy have proposed [23,30] that participants might learn the relation between the part of the word preceding the last phoneme and the last phoneme itself during the course of the experiment, rather than using prior lexical knowledge. This argument applies only to some LMC experiments. For the case where it was proposed [24], the size of the lexical compensation effect should have increased over time, but analysis [60] of data from [24] show that the effect was as strong early in the experiment as it was later.

Proponents of autonomy also propose that the post-perceptual learning mechanism they invoke to explain lexically mediated retuning might also explain the LMA effect [30]. However, retuning and adaptation have different time courses and go in opposite directions. An ambiguous sound is *less* likely to be heard as the lexically appropriate sound after adaptation, but *more* likely to be heard as the lexically appropriate sound after retuning. The arguments offered to explain how the same post-perceptual learning mechanism can account for these differences are post-hoc and appear strained to us. On the other hand, one naturally expects adaptation and tuning to go in opposite directions if, as the interactive view proposes, both are consequences of an increased activation of neurons associated with the contextually specified alternative. Adaptation would then follow as a short-lasting aftereffect of neural and/or synaptic activity [61], whereas retuning would reflect Hebbian long-term potentiation of synaptic connections.

Relation to optimal Bayesian inference

There is now a class of models based on the graphical models framework for optimal Bayesian Inference that have been developed for several domains [43,44], and it seems to us likely that there exists a fairly direct mapping between the TRACE model and models of the sort that would be naturally constructed within this framework.

Box 4. Questions for future research

Representation of time?
How can mechanisms in which no units are dedicated to specific moments nevertheless effectively represent the state of evidence for (and constraints among) the linguistic units occurring at present, past and future time points?

Representation of linguistic units?
The TRACE model incorporates units for features, phonemes and words, corresponding to units that have historically played crucial roles in linguistic theory. What sorts of representation actually comprise the brain's representation of spoken language? Are there representations that are specific to speech (as opposed to other sounds), and if so what relation do such representations bear to the feature, phoneme and word levels used in the TRACE model of speech perception?

Relation to optimal Bayesian inference?
There is now a class of models based on the graphical models framework for optimal Bayesian Inference that can be used to characterize optimal Bayesian inference at a computational level. What is the exact relationship between the interactive activation process in TRACE and the updating of explicit probability estimates in such models?

What is the exact relationship between the interactive activation process as instantiated either in the original TRACE model [2] or more recent stochastic versions of the TRACE model [15,16]?

Conclusion

We have argued here that the interactive approach predicts that the effects of context can penetrate the mechanisms of perception, but autonomous approaches make no such predictions. In our view the evidence is clear in supporting the interactive perspective. For us this motivates the effort to continue to develop the interactive framework. We see a future in which the interactive TRACE model continues to play a constructive role, providing a concrete framework within which to explore the implications of thinking explicitly in terms of interactive processes in perception. More generally we see a bright future at the interface between computational, psychological and neural investigations of the ways in which context can penetrate the mechanisms of perception.

Acknowledgements
Preparation of this article was supported by NIH Grants P50 MH64445 to J.L.M. and by NRSA F31DC0067 to D.M. We thank Nicole Landi and several anonymous reviewers for comments on earlier drafts.

References

1 Ganong, W.F. (1980) Phonetic categorization in auditory word perception. *J. Exp. Psychol. Hum. Percept. Perform.* 6, 110–125
2 McClelland, J.L. and Elman, J.L. (1986) The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86
3 Massaro, D.W. (1998) *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle,* MIT Press
4 Norris, D. *et al.* (2000) Merging information in speech recognition: Feedback is never necessary. *Behav. Brain Sci.* 23, 299–370
5 Rumelhart, D.E. (1977) Toward an interactive model of reading. In *Attention and Performance VI* (Dornic, S., ed.), pp. 573–603, Erlbaum
6 Lee, T.S. and Mumford, D. (2003) Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A* 20, 1434–1448
7 Cooper, R.P. *et al.* (2005) The simulation of action disorganization in complex activities of daily living. *Cogn. Neuropsychol.* 22, 959–1004
8 Marr, D. (1982) *Vision,* W.H. Freeman

9 Fodor, J.A. (1983) *Modularity of Mind,* MIT Press
10 Crick, F.H.C. and Asanuma, C. (1986) Certain aspects of the anatomy and physiology of the cerebral cortex. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Vol. 2)* (McClelland, J.L. and Rumelhart, D.E., eds), pp. 333–371, MIT Press
11 Hupe, J. *et al.* (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394, 784–787
12 Lee, T.S. and Nguyen, M. (2001) Dynamics of subjective contour formation in early visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 98, 1907–1977
13 Massaro, D.W. (1989) Testing between the TRACE model and the fuzzy logical model of speech perception. *Cogn. Psychol.* 21, 398–421
14 Massaro, D.W. and Cohen, M.M. (1991) Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cogn. Psychol.* 23, 558–614
15 McClelland, J.L. (1991) Stochastic interactive processes and the effect of context on perception. *Cogn. Psychol.* 23, 1–44
16 Movellan, J.R. and McClelland, J.L. (2001) The Morton–Massaro law of information integration: Implications for models of perception. *Psychol. Rev.* 108, 113–148
17 Mirman, D. *et al.* (2005) Computational and behavioral investigations of lexically induced delays in phoneme recognition. *J. Mem. Lang.* 52, 424–443
18 Elman, J.L. and McClelland, J.L. (1988) Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *J. Mem. Lang.* 27, 143–165
19 Holt, L.L. (2005) Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychol. Sci.* 16, 305–312
20 Stephens, J.D.W. and Holt, L.L. (2003) Preceding phonetic context affects perception of nonspeech. *J. Acoust. Soc. Am.* 114, 3036–3039
21 Mann, V.A. and Repp, B.H. (1981) Influence of preceding fricative on stop consonant perception. *J. Acoust. Soc. Am.* 69, 546–558
22 Pitt, M.A. and McQueen, J.M. (1998) Is compensation for coarticulation mediated by the lexicon? *J. Mem. Lang.* 39, 347–370
23 McQueen, J.M. (2003) The ghost of Christmas future: didn't Scrooge learn to be good? Commentary on Magnuson, McMurray, Tanenhaus, and Aslin. *Cogn. Sci.* 27, 795–799
24 Magnuson, J.S. *et al.* (2003) Lexical effects on compensation for coarticulation: The ghost of Christmash past. *Cogn. Sci.* 27, 285–298
25 Samuel, A.G. and Pitt, M.A. (2003) Lexical activation (and other factors) can mediate compensation for coarticulation. *J. Mem. Lang.* 48, 416–434
26 Samuel, A.G. (1986) Red herring detectors and speech perception: In defense of selective adaptation. *Cogn. Psychol.* 18, 452–499
27 Samuel, A.G. and Kat, D. (1996) Early levels of analysis of speech. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 676–694
28 Samuel, A.G. (2001) Knowing a word affects the fundamental perception of the sounds within it. *Psychol. Sci.* 12, 348–351
29 Samuel, A.G. (1997) Lexical activation produces potent phonemic percepts. *Cogn. Psychol.* 32, 97–127
30 Norris, D. *et al.* (2003) Perceptual learning in speech. *Cogn. Psychol.* 47, 204–238
31 Davis, M.H. *et al.* (2005) Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *J. Exp. Psychol. Gen.* 134, 222–241
32 Eisner, F. and McQueen, J.M. (2005) The specificity of perceptual learning in speech processing. *Percept. Psychophys.* 67, 224–238
33 Kraljic, T. and Samuel, A.G. (2005) Perceptual learning for speech: Is there a return to normal? *Cogn. Psychol.* 51, 141–178
34 Kraljic, T. and Samuel, A.G. Generalization in perceptual learning for speech. *Psychon. Bull. Rev.* (in press)
35 McQueen, J.M. *et al.* Phonological abstraction in the mental lexicon. *Cogn. Sci.* (in press)
36 Rumelhart, D.E. and Zipser, D. (1985) Feature discovery by competitive learning. *Cogn. Sci.* 9, 75–112
37 Mirman, D. *et al.* An interactive Hebbian account of lexically guided tuning of speech perception. *Psychon. Bull. Rev.* (in press)
38 Smolensky, P. (1986) Neural and conceptual interpretation of PDP models. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Vol. 2)* (McClelland, J.L. and Rumelhart, D.E., eds), pp. 390–431, MIT Press
39 Elman, J.L. (1990) Finding structure in time. *Cogn. Sci.* 14, 179–211
40 Gaskell, M.G. and Marslen-Wilson, W.D. (1997) Integrating form and meaning: A distributed model of speech perception. *Lang. Cogn. Process.* 12, 613–656
41 Bybee, J. and McClelland, J.L. (2005) Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *Linguist. Rev.* 22, 381–410
42 Lotto, A.J. and Holt, L.L. (2000) The illusion of the phoneme. In *Chicago Linguistic Society (Vol. 35): The Panels* (Billings, S.J. *et al.*, eds), pp. 191–204, Chicago Linguistic Society
43 Smyth, P. (1997) Belief networks, hidden Markov models, and Markov random fields: A unifying view. *Pattern Recog. Lett.* 18, 1261–1268
44 Geisler, W.S. and Diehl, R.L. (2003) A Bayesian approach to the evolution of perceptual and cognitive systems. *Cogn. Sci.* 118, 1–24
45 Strauss, T.J. *et al.* jTRACE: A reimplementation and extension of the TRACE model of speech perception and spoken word recognition. *Behav. Res. Methods Instrum. Comput.* (in press)
46 Frauenfelder, U.H. *et al.* (1990) Lexical effects in phonemic processing: Facilitatory or inhibitory? *J. Exp. Psychol. Hum. Percept. Perform.* 16, 77–91
47 Wurm, L.H. and Samuel, A.G. (1997) Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring. *J. Mem. Lang.* 36, 165–187
48 Marslen-Wilson, W. and Warren, P. (1994) Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychol. Rev.* 101, 653–675
49 McQueen, J.M. *et al.* (1999) Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 1363–1389
50 Dahan, D. *et al.* (2001) Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Lang. Cogn. Process.* 16, 507–534
51 Pitt, M.A. *et al.* (2006) Global model analysis by parameter space partitioning. *Psychol. Rev.* 113, 57–83
52 Eimas, P.D. *et al.* (1990) Attention and the role of dual codes in phoneme monitoring. *J. Mem. Lang.* 29, 160–180
53 Cutler, A. *et al.* (1987) Phoneme identification and the lexicon. *Cogn. Psychol.* 19, 141–177
54 Mirman, D. *et al.* Attentional modulation of lexical effects on speech perception: Computational and behavioral experiments. *Proc. 28th Annu. Conf. Cogn. Sci. Soc.*, Erlbaum (in press)
55 Hartsuiker, R.J. *et al.* (2005) The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Belated reply to Baars *et al.* (1975). *J. Mem. Lang.* 52, 58–70
56 Monsell, S. *et al.* (1992) Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 452–467
57 Bonte, M. *et al.* (2006) Time course of top-down and bottom-up influences on syllable processing in the auditory cortex. *Cereb. Cortex* 16, 115–123
58 Desimone, R. and Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222
59 O'Craven, K.M. *et al.* (1997) Voluntary attention modulates fMRI activity in human MT-MST. *Neuron* 18, 591–598
60 Magnuson, J.S. *et al.* (2003) Lexical effects on compensation for coarticulation: A tale of two systems? *Cogn. Sci.* 27, 801–805
61 Laing, C.R. and Chow, C.C. (2002) A spiking neuron model of binocular rivalry. *J. Comput. Neurosci.* 12, 39–53

Letters

Are there really interactive processes in speech perception?

James M. McQueen[1], Dennis Norris[2] and Anne Cutler[1]

[1] Max Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands
[2] MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK

On both empirical and theoretical grounds, we argue that the affirmative answer of McClelland *et al.* [1] is premature. Contrary to the predictions of the TRACE model, which postulates interactive processing in speech perception, there is no lexically mediated compensation for coarticulation when there is a lexical bias in interpretation of the preceding fricative [2]. This cannot be dismissed with criticism of the fricative-final lexical contexts and their distance from the stops (Box 3 in Ref. [1]). The contexts in Ref. [2] were effective – there was a lexical bias on the fricatives; and the timing was appropriate – there was a compensation effect with the same fricative-stop gap in the nonword-context conditions. Furthermore, we suggest that the evidence on compensation that purports to show interaction is unconvincing. Some apparent lexical effects [3,4] are possibly higher-order transitional-probability effects [2]. Others perhaps reflect learning about experiment-internal biases; indeed, the effect in Ref. [5] did not seem to increase over time, but this null result is not definitive because learning could have occurred in the practice block.

Perceptual retuning can explain the selective-adaptation data (e.g. Ref. [6]). Retuning of phonetic perception can arise after only ten critical trials [7]; selective adaptation effects depend on many more exposures (e.g. 768 in Ref. [6]). Thus, selective adaptation can follow from perceptual retuning. After listeners learn that the ambiguous phoneme is /s/, for example, it acts as an adaptor, reducing the number of /s/ responses to the test stimuli. Further analysis of the data in Ref. [6] reveals exactly this: early blocks of trials show retuning, whereas later blocks show adaptation (Figure 3 in Ref. [8]).

The evidence on whether there is on-line lexical influence on prelexical processes is thus inconclusive. However, consensus has been reached on the existence of lexical feedback for learning [7,9] and, as just shown, this can also explain apparent evidence of on-line interaction. Why might there be feedback that affects learning but not processing? Feedback for learning is helpful because it enables the listener to adjust to speaker-related variability [9], but feedback in on-line processing is not beneficial and could even be harmful [10]. These views are based on rational analysis [11]. Analyzing the nature of the perceptual task generated a hypothesis (that was confirmed in Ref. [9]) about how speech perception should operate.

According to the rational-analysis perspective, and McClelland *et al.* [1], the goal of the speech-recognition device is optimal interpretation. Bayesian methods provide the optimal way to combine independent sources of information for perceptual decisions. However, if interaction were permitted between information sources, those sources would no longer be independent and the decision would be suboptimal [10]. Interaction, therefore, makes optimal interpretation harder. If an interactive algorithm could be made to compute the correct Bayesian decision function, the interaction would be a property only of that algorithm, not of the underlying computation. What would be computed is exactly what a non-interactive system would compute. Hence, commitment to the computational principle of optimality requires no commitment to the algorithmic principle of interactive processing.

No data require direct influences of the lexicon on prelexical mechanisms, and evidence and computational principles argue against interactive processing. Further evidence of lexical mediation of prelexical processes might yet appear. Indeed, if lexical retuning were implemented in the TRACE model using Hebbian learning [9,12], there could well be on-line processing effects. Such demonstrations would be further evidence of feedback in learning and not evidence of interactive processing, which, other than potentially as part of a learning mechanism, serves no useful function.

References

1 McClelland, J.L. *et al.* (2006) Are there interactive processes in speech perception? *Trends Cogn. Sci.* 10, 363–369
2 Pitt, M.A. and McQueen, J.M. (1998) Is compensation for coarticulation mediated by the lexicon? *J. Mem. Lang.* 39, 347–370
3 Elman, J.L. and McClelland, J.L. (1988) Cognitive penetration of the mechanisms of perception: compensation for coarticulation of lexically restored phonemes. *J. Mem. Lang.* 27, 143–165
4 Samuel, A.G. and Pitt, M.A. (2003) Lexical activation (and other factors) can mediate compensation for coarticulation. *J. Mem. Lang.* 48, 416–434
5 Magnuson, J.S. *et al.* (2003) Lexical effects on compensation for coarticulation: the ghost of Christmash past. *Cogn. Sci.* 27, 285–298
6 Samuel, A.G. (2001) Knowing a word affects the fundamental perception of the sounds within it. *Psychol. Sci.* 12, 348–351
7 Kraljic, T. and Samuel, A.G. Perceptual adjustments to multiple speakers. *J. Mem. Lang.* (in press)
8 Vroomen, J. *et al.* Visual recalibration and selective adaptation in auditory-visual speech perception: contrasting build-up courses. *Neuropsychologia* (in press)
9 Norris, D. *et al.* (2003) Perceptual learning in speech. *Cogn. Psychol.* 47, 204–238
10 Norris, D. *et al.* (2000) Merging information in speech recognition: feedback is never necessary. *Behav. Brain Sci.* 23, 299–370
11 Anderson, J.R. (1990) *The Adaptive Character of Thought.* Erlbaum
12 Mirman, D. *et al.* An interactive Hebbian account of lexically guided retuning of speech perception. *Psychon. Bull. Rev.* (in press)

Corresponding author: McQueen, J.M. (james.mcqueen@mpi.nl)
Available online 25 October 2006.

1364-6613/$ – see front matter
doi:10.1016/j.tics.2006.10.004

Letters Response

Response to McQueen *et al.*: Theoretical and empirical arguments support interactive processing

Daniel Mirman[1], James L. McClelland[2] and Lori L. Holt[3]

[1] Department of Psychology, University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, CT 06269-1020, USA
[2] Department of Psychology, Stanford University and Center for Mind, Brain and Computation, Jordan Hall, Building 420, Stanford, CA 94305, USA
[3] Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

McQueen *et al.* [1] continue to argue against interactive processes in speech perception, but we suggest that their arguments are unconvincing. Theoretical and empirical arguments support the interactive account. Concerning their theoretical points, a rational analysis is consistent with interactive models because they can produce optimal information integration [2]. We argue that interactive, rather than feedforward, processing is the algorithm that the brain uses to accomplish optimal information integration. Interactive processing provides a more parsimonious algorithm than the feedforward approach of McQueen *et al.*, which requires an additional decision level and a specialized feedback mechanism that affects learning but not processing.

We suggest that the empirical arguments offered by McQueen *et al.* are also unconvincing. The failure to find lexically mediated compensation for coarticulation in Ref. [3] is not problematic; the lexically mediated effect will necessarily be smaller than the effect that is produced by an unambiguous phoneme (Figure 2 in Ref. [4]) and might be too small to be detected reliably. Furthermore, one failure to replicate cannot outweigh three independent successful replications that were based on 16 different lexical contexts (reviewed in Ref. [4]). Regarding the 'higher-order transitional probability' argument of McQueen *et al.*, there is no definition of 'higher-order transitional probability' that can account for the full set of data [5].

Perceptual learning cannot explain lexically induced selective adaptation as neatly as McQueen *et al.* claim. They cite audiovisual recalibration data from the ambiguous condition in Ref. [6] that showed learning followed by adaptation-driven unlearning. However, the lexically mediated selective-adaptation data (Figure 3 in Ref. [6]) correspond more closely to the unambiguous condition (Figure 1 in Ref. [6]), showing selective adaptation relative to baseline. This correspondence suggests that lexically mediated selective adaptation operates in the same way as perceptually mediated selective adaptation (the unambiguous condition in Ref. [6]), as predicted by interactive processing.

In sum, McQueen *et al.* [1] have provided neither a theoretical basis nor a sufficient argument to bring into doubt the evidence that supports interactive processes in speech perception. Lexically guided learning is not a special case for which feedback must be introduced; it is just one of many benefits of interactive processing.

References

1 McQueen, J.M. *et al.* (2006) Are there really interactive speech processes in speech perception? *Trends Cogn. Sci.* 10, 533
2 Movellan, J.R. and McClelland, J.L. (2001) The Morton–Massaro law of information integration: implications for models of perception. *Psychol. Rev.* 108, 113–148
3 Pitt, M.A. and McQueen, J.M. (1998) Is compensation for coarticulation mediated by the lexicon? *J. Mem. Lang.* 39, 347–370
4 McClelland, J.L. *et al.* (2006) Are there interactive processes in speech perception? *Trends Cogn. Sci.* 10, 363–369
5 Magnuson, J.S. *et al.* (2003) Lexical effects on compensation for coarticulation: a tale of two systems? *Cogn. Sci.* 27, 801–805
6 Vroomen, J. *et al.* Visual recalibration and selective adaptation in auditory–visual speech perception: contrasting build-up courses. *Neuropsychologia* (in press)

1364-6613/$ – see front matter
doi:10.1016/j.tics.2006.10.003

Corresponding author: Mirman, D. (daniel.mirman@uconn.edu)
Available online 30 October 2006.