12-1-2006

An Interactive Hebbian Account of Lexically Guided Tuning of Speech Perception

Daniel Mirman
University of Connecticut

James L. McClelland
Carnegie Mellon University

Lori L. Holt
Carnegie Mellon University, lholt@andrew.cmu.edu

An interactive Hebbian account of lexically guided tuning of speech perception

DANIEL MIRMAN
University of Connecticut, Storrs, Connecticut

and

JAMES L. McCLELLAND and LORI L. HOLT
Carnegie Mellon University, Pittsburgh, Pennsylvania

We describe an account of lexically guided tuning of speech perception based on interactive processing and Hebbian learning. Interactive feedback provides lexical information to prelexical levels, and Hebbian learning uses that information to retune the mapping from auditory input to prelexical representations of speech. Simulations of an extension of the TRACE model of speech perception are presented that demonstrate the efficacy of this mechanism. Further simulations show that acoustic similarity can account for the patterns of speaker generalization. This account addresses the role of lexical information in guiding both perception and learning with a single set of principles of information propagation.

Lexical knowledge can affect listeners' categorization of speech sounds. For example, an ambiguous /g/–/k/ sound—one that is classified about equally often as /g/ or /k/ when it occurs in a lexically neutral context—tends to be classified as /g/ when preceding *ift* but as /k/ when preceding *iss* (Ganong, 1980). Our focus here is on recent studies showing that lexical knowledge can also guide tuning of the mapping from auditory input representations to speech sound representations (Norris, McQueen, & Cutler, 2003; see also Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Eisner & McQueen, 2005; Kraljic & Samuel, 2005, 2006; Maye, Aslin, & Tanenhaus, 2003). In the basic paradigm (Norris et al., 2003), when listeners hear a perceptually ambiguous /s/–/f/ sound at the end of an utterance that would be a word if completed with /s/, they both identify the sound as /s/ and retune perception so that ambiguous sounds tend to be identified subsequently as /s/, even in lexically neutral contexts. Further studies employing this paradigm have revealed an interesting and complex pattern of generalization of this effect.

Our goal in this report is to demonstrate that interactive processing, initially proposed to account for lexical effects on perception (McClelland & Elman, 1986), provides the needed information to prelexical levels to support lexically guided tuning effects. The principle of interactive processing has been challenged by proponents of autono-mous models (Norris, McQueen, & Cutler, 2000), who attribute most lexical effects to postperceptual decision processes rather than to interactive processing. However, to account for lexically guided tuning of perception, these proponents allow feedback to guide tuning of perceptual mechanisms but not to guide the perceptual mechanisms themselves. We propose instead that feedback is indeed at work in perception and that this feedback has the right properties to successfully guide the retuning process.

There are now several findings supporting the view that lexical factors can affect prelexical processing, as predicted by the interactive approach and in contrast to the claims of autonomous models (for a full review, see McClelland, Mirman, & Holt, 2006). These effects of lexical factors on prelexical processing include lexically guided compensation for coarticulation (Elman & McClelland, 1988; Magnuson, McMurray, Tanenhaus, & Aslin, 2003; Samuel & Pitt, 2003) and lexically guided selective adaptation (Samuel, 1997, 2001). When taken together with these findings, lexically guided tuning of speech perception would simply be another instance of a prelexical consequence of lexical feedback.

More broadly, the principle of interactive processing predicts that context will affect processing at many levels, across many different domains and modalities, and that such effects should in turn contribute to the guidance of tuning. For example, recent studies showing that visual information can guide tuning of mappings from auditory to speech sound representations (Bertelson, Vroomen, & de Gelder, 2003) are completely consistent with the perspective presented here.

By introducing a simple learning algorithm into the existing structure of the interactive TRACE model and showing that the model can then address many features of the interesting pattern of lexically guided tuning effects, we demonstrate that the interactive processing mechanism

can provide the information needed to guide prelexical processing. The TRACE model was developed within the parallel distributed processing framework, in which connection-based learning plays a central role. Although the original formulation of TRACE did not include a mechanism for learning, McClelland and Elman (1986) noted that learning could be incorporated into TRACE as a means of tuning perception. The learning algorithm we propose takes advantage of the interactive mechanism wherein activation of lexical representations feeds back to the prelexical level and provides excitatory input to the lexically consistent units. Although we rely on TRACE as a representative interactive model in this report, other interactive models (e.g., that of Carpenter & Grossberg, 1991) are completely consistent with the ideas described here.

The learning rule we consider is a variant of Hebbian learning. By the term *Hebbian* we mean that the learning algorithm relies on Hebb's postulate, which we paraphrase as follows: When a sending unit s participates in firing a receiving unit r, the strength of the connection determining the influence of s on r will be increased (Hebb, 1949). Models incorporating this type of learning have been used to account for a broad range of data in the domains of speech and visual perception (e.g., Carpenter & Grossberg, 1991; Grossberg, 1976), including initial acquisition of speech sound representations through statistical learning (e.g., Guenther & Gjaja, 1996) and successes and failures of adults learning nonnative speech sound contrasts (e.g., McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002).

The Hebb-TRACE Model of Speech Perception[1]

The Hebb-TRACE model of speech perception, introduced here, integrates a form of Hebbian learning into the version of the TRACE model presented by McClelland and Elman (1986). The TRACE model consists of three layers: an acoustic/articulatory feature layer, where input is represented in seven banks of units corresponding to values along each feature dimension (e.g., voicing); a phoneme layer, where each unit corresponds to a particular phoneme; and a lexical layer, where each unit corresponds to a particular word. When TRACE is presented with an ambiguous input, feedback from the lexical layer gives lexically consistent phoneme interpretations an advantage over lexically inconsistent ones, and this advantage is increased through competitive lateral inhibition among phoneme units. At this point in processing, a Hebbian learning algorithm can associate the ambiguous feature input with the lexically consistent phoneme. Following this learning, the ambiguous input will come to activate the lexically consistent phoneme, even in the absence of lexical feedback (i.e., in lexically neutral contexts).

A specific learning rule consistent with Hebb's principle and often used for unsupervised category learning is the *competitive learning* rule (Grossberg, 1976; Oja, 1982; Rumelhart & Zipser, 1985; von der Malsburg, 1973):

$$\Delta W_{s \to r} = \lambda a_r \left(a_s - W_{s \to r} \right).$$

Here, $W_{s \to r}$ is the weight from the sending unit s to the receiving unit r, a_r and a_s are the activations of units s and r, and λ is the learning rate. In the weight change equation, the product of a_r and a_s embodies the Hebbian principle that when a sending unit s participates in firing a receiving unit r, the strength of connection determining the influence of s on r will be increased. The subtraction of the existing value of the weight from a_s causes the algorithm to align the weights to the receiving units with the pattern of incoming activation (the weights stop changing when they match the pattern of incoming activation). Following Rumelhart and Zipser, weights and feature activations were normalized within each feature dimension:

$$\Delta W_{s \to r} = \lambda a_r \left[\left(a_s / S_s \right) - \left(W_{s \to r} / S_w \right) \right],$$

where S_s is the sum of all feature unit activations within a given feature bank and S_w is the sum of all weights from a given feature bank to the receiving phoneme unit (r). The normalization keeps the sum of the weights to a receiver from all the units in a feature bank stable, even as changing inputs adjust the distribution of the weights.

In TRACE, activation is propagated in a cascading fashion; as a result, the temporal patterns of activation of feature and phoneme units are overlapping, but they tend to be offset from each other in time. That is, feature unit activations build up and begin to activate phoneme units, but by the time phoneme unit competition has been resolved and activation has built up to near-peak levels, feature unit activations are already decaying toward their rest values. This temporal asynchrony poses a problem for the learning algorithm, because the algorithm requires the sending and receiving units to be active simultaneously. The normalization of feature activations counteracts this temporal asynchrony to some extent, but it is ineffective when the activations of all feature units have decayed to 0. One solution to this problem would be to introduce a temporal offset into the learning rule (see, e.g., Bi & Poo, 2001). However, in the interest of simplicity, this was not implemented in the present simulations. Instead, the learning was turned off when there was no activity in the feature layer (i.e., when $S_s \le 0$). We see this approach as a simplification of the biological learning algorithm that maintains the basic principles of Hebbian learning, thus allowing a comparatively simple investigation of perceptual tuning by lexical feedback. Also, consistent with the principles of competitive learning and interactive processing, learning was applied only to those phoneme units that were active above their interactive threshold (0).

When input is presented to the TRACE model, it is processed in a series of time steps. On each time step, net inputs (excitatory inputs from units at adjacent levels and inhibitory inputs from units at the same level) are computed for each level, and activations are updated. In Hebb-TRACE, an additional weight update step (the learning rule described above) is also performed on every time step on the basis of the current activations and weights. The learning rule is applied to all phoneme units in all time slices, and it affects feature-to-phoneme as well as

phoneme-to-feature weights (which are symmetric with respect to each other).

SIMULATION 1
## Lexically Guided Tuning of Speech Perception

Materials and Method

Simulation 1 was designed to test whether Hebb-TRACE could account for the basic findings of Norris et al. (2003). To this end, the simulation procedure mimicked their experimental procedure. A pretest simulation was conducted to assess baseline perception of an ambiguous fricative (/s/ or /ʃ/) in a lexically neutral context. The key simulations consisted of an exposure phase, in which the model was presented with a sequence of fricative-final words, followed by an identification phase identical to the pretest. There were two types of exposure phases: In one, the /s/ in all the /s/-containing words was replaced with the ambiguous fricative ("[?s]+[ʃ] words"); in the other, the /ʃ/ in all the /ʃ/-containing words was replaced with the ambiguous fricative ("[s]+[?ʃ] words"). The word contexts for these simulations are in the Appendix. For these simulations, the learning rate (λ) was set to 0.001 during the exposure phases and 0.0 during the test phases (i.e., learning was turned off during the test phases). Otherwise, the stan-

dard TRACE parameter values were used[2] (McClelland & Elman, 1986; Mirman, McClelland, & Holt, 2005).

Results and Discussion

Luce choice probabilities for the ambiguous fricative (top row) and the two unambiguous fricatives (middle and bottom rows) are shown in Figure 1. At pretest (top left panel), the ambiguous fricative was equally likely to be perceived as /s/ or /ʃ/; following /s/-biased exposure (middle), the sound was perceived as /s/; following /ʃ/-biased exposure (right), the sound was perceived as /ʃ/. The middle and bottom rows demonstrate that tuning had no effect on perception of unambiguous sounds.

Tuning of the voicing boundary between /d/ and /t/ generalizes to the voicing boundary between /b/ and /p/ (Kraljic & Samuel, 2006). The localist phoneme representation of TRACE precludes this type of generalization, because the similarity between the /d/–/t/ distinction and the /b/–/p/ distinction is not represented at the phoneme layer. An interactive Hebbian learning model in which phonemes are represented in terms of sets of contrastive or distinctive features, however, might capture this similarity and consequently produce this type of generalization. In future work, detailed patterns of generalization could be used to constrain and inform the details of representations.
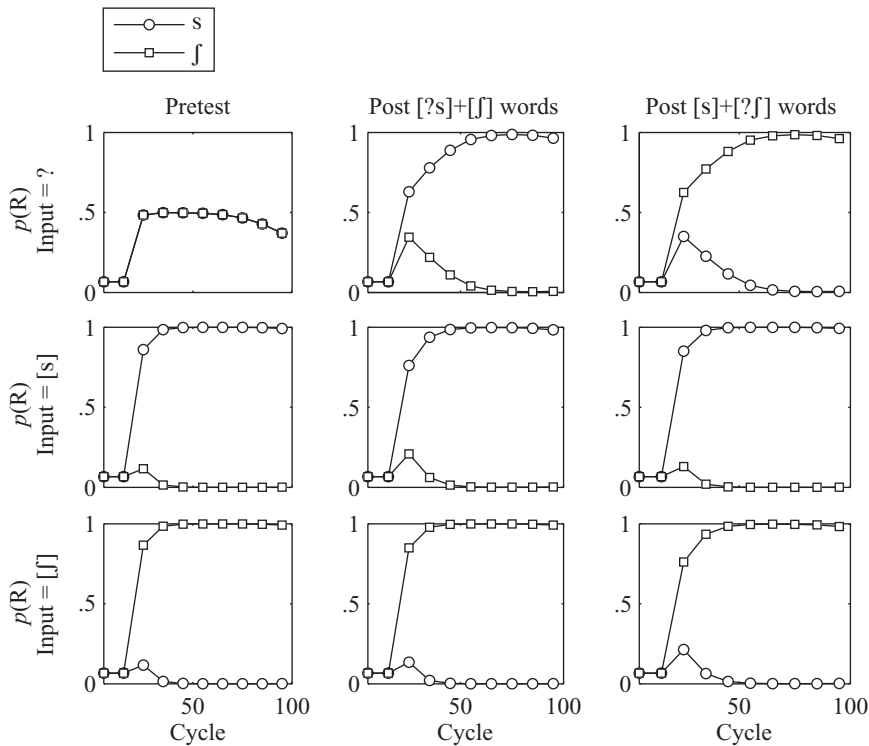
Figure 1. Top row: Response likelihood for the two interpretations of an ambiguous frica-tive at pretest, after /s/-biased exposure, and after /ʃ/-biased exposure. At pretest (left), the sound was perfectly ambiguous between the two interpretations; following /s/-biased expo-sure (middle), the sound was perceived as /s/; following /ʃ/-biased exposure (right), the sound was perceived as /ʃ/. Middle and bottom rows: These demonstrate that tuning had no effect on perception of unambiguous sounds.

To test generalization to novel lexical contexts, simulations of minimal pairs (which were not presented during the exposure phase) were carried out. For these simulations, an ambiguous fricative replaced the distinguishing phoneme in two /s/–/ʃ/ minimal-pair words (*parcel*/ *partial*, *police*/*polish*).[3] Lexical unit activations[4] for the words are shown in Figure 2 and illustrate that learning from the exposure phase generalizes to novel lexical contexts. Following [?s]+[ʃ] exposure (left panel), the model interpreted the presented word as containing an /s/ (i.e., *parcel* or *police*), and following [s]+[?ʃ] exposure (right panel), the model interpreted the same presented word as containing a /ʃ/ (i.e., *partial* or *polish*). This type of generalization arises because phonemes are represented as abstract units independent of lexical or acoustic content; thus, the changes to feature-to-phoneme weights generalize across lexical contexts. These results are consistent with recent behavioral experiments that have shown generalization in perception of minimal-pair words that were not part of the training set (McQueen, Cutler, & Norris, 2006).

SIMULATION 2
Speaker Generalization

Studies building on the initial findings of lexically guided tuning have examined whether the tuning is speaker specific or generalizes to novel speakers. Eisner and McQueen (2005) found that lexically guided tuning for fricatives (/s/–/f/) did not generalize to a novel speaker. Kraljic and Samuel (2005) also tested fricatives (/s/–/ʃ/) and found that tuning did not generalize across speakers and that speaker-specific exposure was required to undo the effects of lexically guided tuning. When a male voice was heard during the exposure phase, at posttest the tuning effect was reliable for male-produced tokens but not for female-produced tokens, and in this context posttuning exposure to male-produced unambiguous tokens reduced the tuning effect, but posttuning exposure to female-produced

unambiguous tokens did not. That is, untuning displayed the same pattern of speaker specificity as tuning. In contrast, Kraljic and Samuel (2006) tested stop consonants (/d/–/t/) and found that tuning *did* generalize to a novel speaker.

Kraljic and Samuel (2005) suggested that the patterns of generalization may be due to acoustic similarity among the different exposure and test tokens. They argued that the acoustic cues that distinguish the test fricatives are more variable between speakers than the voicing cue that distinguishes the test stops. Thus, tuning of auditory-to-speech-sound mappings may generalize to acoustically similar sounds but not to acoustically dissimilar sounds. In a more detailed analysis, Kraljic and Samuel (2005) then showed that tuning generalized for fricatives when the spectral mean (one cue to fricative identity) of the exposure fricative fell within the range of the test (i.e., the new speaker's) fricatives (their female→male condition), but not when the spectral mean of the exposure fricative fell outside the range of the test fricatives (their male→female condition). Eisner and McQueen's (2005) data are also consistent with this perspective: The tuning effect on fricative identification was strongest when the test and exposure vowel–fricative stimuli were produced by the same speaker, weaker when the stimulus consisted of the exposure fricative spliced with a vowel produced by a different speaker, and weakest when the vowel and fricative at test were both produced by a different speaker. In the following simulations, we show that the Hebb-TRACE model can capture these hypothesized effects of acoustic similarity.

Materials and Method

The lexical contexts and simulation parameters were the same as those in Simulation 1. To model speaker similarity, we created "male" and "female" versions of ambiguous stop consonant (/d/ or /t/) and ambiguous fricative (/s/ or /ʃ/) input patterns. The two versions of the ambiguous stop differed not with respect to the features that distinguish /d/ and /t/, but instead on a feature that was irrelevant to stop
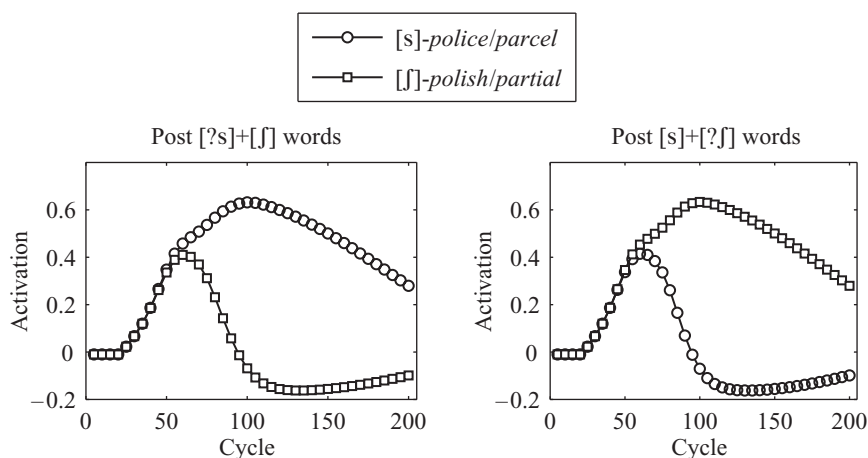
Figure 2. Lexical activations of the two possible interpretations of /s/–/ʃ/ minimal-pair words after the fricative has been replaced by an ambiguous fricative. The left panel shows the pattern of activation following /s/-biased exposure, the right panel the pattern of activation following /ʃ/-biased exposure.

identity. The two versions of the ambiguous fricative had different values on the features that distinguish /s/ and /ʃ/ (see the Appendix for more details). Note that although the "male" and "female" versions of the ambiguous inputs were different at the feature level, they were the same at the phoneme level. This was because their feature values were balanced such that their phonetic interpretations (before tuning) would be identical. It is important to stress that our implementation was not intended as a veridical representation of the acoustic differences between male and female versions of stops and fricatives. Rather, our goal was to investigate the effect of acoustic similarity between speakers on generalization of lexically guided tuning as a result of Hebbian learning in an interactive model of speech perception.

To test the effect of speaker similarity on the generalization of learning, the model was trained using one version of the ambiguous phoneme, and the tuning effect was evaluated for the two versions of that ambiguous phoneme. The tuning effect was calculated as the difference between (1) /d/ (or /s/) response likelihood following /d/-biased (or /s/-biased) exposure and (2) the response likelihood for the same phoneme following /t/-biased (or /ʃ/-biased) exposure [i.e., $p(/d/|/d/\text{-biased exposure}) - p(/d/|/t/\text{-biased exposure})$].[5] Stops were used for the similar condition; fricatives were used for the dissimilar condition. If the acoustic similarity account is correct, tuning would generalize for the similar sounds (i.e., the tuning effect would be of approximately equal size for both versions), but not for the dissimilar sounds (i.e., there would be a large difference in tuning effect size between the two versions). To test the speaker specificity of untuning following tuning to one version of the ambiguous fricative, the model was exposed to unambiguous inputs based on either the same version of the ambiguous phoneme used in the tuning phase or the other version. If the acoustic similarity account is correct, the tuning effect should be reduced following exposure to unambiguous inputs based on the same version but not following exposure to unambiguous sounds based on the different version.

Results and Discussion

Figure 3 shows the results of the tuning and untuning simulations. When the two versions of an ambiguous phoneme were similar (left panel), tuning generalized from the exposure version ("Same") to the generalization version ("Diff"). In contrast, when the two versions of the ambiguous phoneme were dissimilar (middle panel), tuning showed weak generalization across versions. The untuning results (right panel) also matched the behavioral data (Kraljic & Samuel, 2005): Presentation of unambiguous sounds in the same "voice" as in the exposure phase caused a reduction in the lexically guided tuning effect (although it did not eliminate it), and presentation of unambiguous sounds in a different "voice" did not dampen the tuning effect.

Recently, Kraljic and Samuel (in press) provided yet another important piece of data on the speaker specificity of tuning effects. They found that when both voices were presented during exposure in opposite lexical contexts
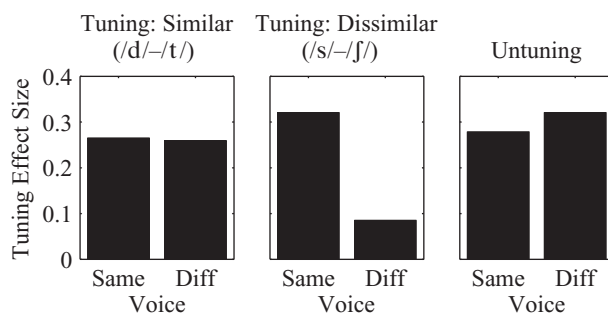
Figure 3. Speaker specificity of lexically guided tuning. Left panel: Tuning effect size for exposure version ("Same") and generalization version ("Diff") for similar sounds (/d/–/t/). Middle panel: Tuning effect size for exposure version and generalization version for dissimilar sounds (/s/–/ʃ/). The tuning effect generalizes for similar sounds but not for dissimilar sounds. Right panel: Tuning effect size following an untuning phase in which unambiguous fricatives were presented in the exposure "voice" or in a different "voice." The tuning effect size is reduced only when the stimuli in the untuning phase were in the exposure "voice."

(e.g., a male ambiguous stop in /d/ lexical contexts and a female ambiguous stop in /t/ lexical contexts, or a male ambiguous fricative in /s/ lexical contexts and a female ambiguous fricative in /ʃ/ lexical contexts), there was no net tuning for stops, but there was speaker-specific tuning for fricatives. That is, the data suggest that when the male and female versions are acoustically similar (ambiguous stops), tuning affects the same representations; thus, the male /d/-bias is undone by the female /t/-bias. In contrast, when the male and female versions are acoustically different (ambiguous fricatives), tuning is speaker specific (e.g., in the example above, the male ambiguous fricative was perceived as /s/ and the female ambiguous fricative as /ʃ/). Figure 4 shows that Hebb-TRACE produced exactly the same pattern. When the two versions of the ambiguous fricative were presented in different lexical contexts, the tuning effect was restricted to the same version only (filled symbols). As demonstrated above, tuning for similar phonemes (i.e., stops) generalized across ambiguous phoneme versions; as a result, Kraljic and Samuel's (in press) multiple-speaker similar-phoneme exposure condition necessarily led to zero net tuning, since generalization from one speaker undid any tuning from the other speaker (open symbols; the /d/ lexical contexts were slightly stronger, and thus there is a small /d/ bias across all exposure and input conditions).

Our simulations show that the different patterns of speaker generalization of lexically guided tuning for stops and fricatives can be captured under the assumption that the acoustic realizations of stop voicing cues are more similar across speakers than are the acoustic realizations of fricative place cues. This assumption can also account for other differences in the tuning effect between stops and fricatives. In particular, the persistence of the fricative tuning effect relative to the stop tuning effect could be simply a product of the speaker specificity of fricative tuning and untuning: An effect that can be undone only by a specific speaker should last longer than an ef-
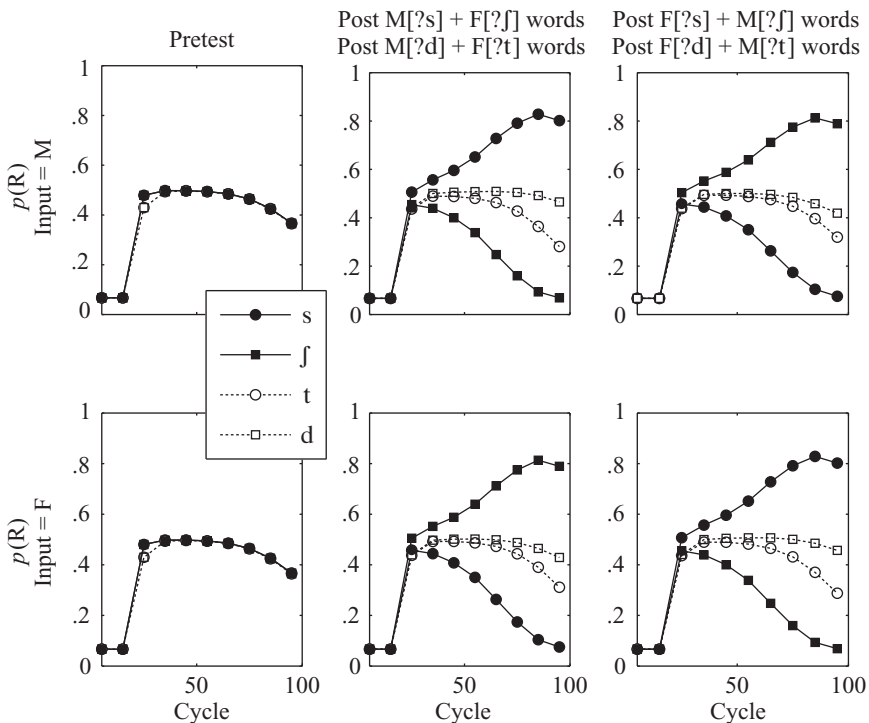
Figure 4. Simultaneous speaker-specific tuning following multiple-speaker exposure. Filled symbols: "Male" and "female" versions are different (i.e., fricatives). Open symbols: "Male" and "female" versions are similar (i.e., stops). Top row: Response likelihood for two interpretations of ambiguous "male" speech sound. Bottom row: Response likelihood for two interpretations of ambiguous "female" speech sound.

fect that can be undone by any speaker. Similarly, since tuning of the boundary between one pair of stops (/d/ vs. /t/) generalizes to other stop pairs with the same voicing contrast (e.g., /b/ vs. /p/; Kraljic & Samuel, 2006), Kraljic and Samuel's (2006) inclusion of unambiguous /g/ and /k/ stimuli during exposure intended to tune the /d/–/t/ boundary would be expected to attenuate the magnitude of the tuning effect. This would account for the observation that fricatives (Kraljic & Samuel, 2005) produce much larger effects than stops do (Kraljic & Samuel, 2006), though this effect size difference could also be due to a number of possible stimulus factors.

GENERAL DISCUSSION

The present simulations demonstrate that an interactive model with Hebbian learning provides a simple and integrated account of a complex pattern of data. In this model, lexical feedback enhances activation of units corresponding to prelexical representations that are consistent with lexical information. These activations in turn allow Hebbian learning to tune mappings from units participating in auditory input representations to prelexical speech sound representations. It is a straightforward consequence of the architecture of the model that this tuning generalizes to words that were not presented during the exposure phase, since the prelexical units mediate between the input and

lexical levels. In the model, as in experimental data, the acoustic similarity of cues across speakers determines when tuning will generalize from one speaker to another and determines whether posttuning exposure to a different speaker will counteract the tuning effect. We suggest that the combined principles of interactive activation and Hebbian learning may have wide applicability, offering accounts for learning due to audio–visual interactions (Bertelson et al., 2003) and evidence of interactive effects in other modalities (e.g., figure–ground perception; Lee & Nguyen, 2001).

The development of cognitive theories depends on an understanding of the principles of cognitive processing. One controversial principle is interactive processing: the bidirectional information flow that allows multiple levels to work in tandem to develop and constrain a percept. Proponents of the autonomous view (Norris et al., 2000) propose that feedback exists for learning but not for perception (Norris et al., 2003). We see several drawbacks to this approach. First, the autonomous proposal requires what seems to us an arbitrary distinction between the information propagation principles governing perception and learning. Second, it requires either the dismissal or separate treatment of other findings pointing to the idea that lexical context can affect prelexical speech representations. These other findings include lexically mediated compensation for coarticulation (Elman & McClelland,

1988; Magnuson et al., 2003; Samuel & Pitt, 2003) and lexically mediated selective adaptation of speech perception (Samuel, 1997, 2001). Third, we note that the separate propagation of feedback for the sake of learning introduces complexities that are not needed if such propagation is a natural part of processing itself. Algorithms such as back propagation that employ such separate signals are often viewed as biologically implausible (Grossberg, 1987). We share with Grossberg (1987), O'Reilly (1996), and others the view that the necessary signals for learning in multilayer perceptual systems arise through the process of interactive activation. The present simulations of interactive Hebbian tuning of speech perception demonstrate the power and parsimony of interactive processing in accounting for lexical effects on both perceptual processing and learning.

REFERENCES

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science*, **14**, 592-597.

Bi, G.-Q., & Poo, M.-M. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annual Review of Neuroscience*, **24**, 139-166.

Carpenter, G. A., & Grossberg, S. (1991). *Pattern recognition by self-organizing neural networks*. Cambridge, MA: MIT Press.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, **134**, 222-241.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, **67**, 224-238.

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory & Language*, **27**, 143-165.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 110-125.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: Part I. Parallel development and coding of neural feature detectors. *Biological Cybernetics*, **23**, 121-134.

Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, **11**, 23-63.

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America*, **100**, 1111-1121.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, **51**, 141-178.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, **13**, 262-268.

Kraljic, T., & Samuel, A. G. (in press). Perceptual adjustments to multiple speakers. *Journal of Memory & Language*.

Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences*, **98**, 1907-1911.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: The ghost of Christmash past. *Cognitive Science*, **27**, 285-298.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2003, March). *In search of the Weckud Wetch: Online adaptation to speaker accent*. Poster presented at the CUNY Conference on Sentence Processing, Cambridge, MA.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]–[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience*, **2**, 89-108.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, **10**, 363-369.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, **30**, 1113-1126.

Mirman, D., McClelland, J. L., & Holt, L. L. (2005). Computational and behavioral investigations of lexically induced delays in phoneme recognition. *Journal of Memory & Language*, **52**, 424-443.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral & Brain Sciences*, **23**, 299-370.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, **47**, 204-238.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, **15**, 267-273.

O'Reilly, R. C. (1996). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation*, **8**, 895-938.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science*, **9**, 75-112.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology*, **32**, 97-127.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, **12**, 348-351.

Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory & Language*, **48**, 416-434.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, **14**, 85-100.

NOTES

1. The model code, including parameter and lexicon files and an example script file, is available at magnuson.psy.uconn.edu/mirman/research.htm or by contacting the first author.

2. The default phoneme-to-feature feedback value is 0.0. Changes to this value did not affect model performance as long as feature unit decay and phoneme–phoneme inhibition parameters were also changed to maintain the dynamics of unit activation and decay.

3. Although *police* and *polish* are not a minimal pair in English, their representations in TRACE were distinguished solely by the final fricative.

4. Note that lexical unit activations peak later than phoneme unit activations. Generalization to novel lexical contexts is presented in terms of lexical activation rather than phoneme response probability in order to provide a more direct account of behavioral priming data (McQueen, Cutler, & Norris, 2006).

5. This data format was chosen for simplicity of presentation and to match the Kraljic and Samuel (2005) data representation.

APPENDIX

Word Contexts Used for Simulations

Fricative

/s/-bias: *decrease*, *produce*, *carcass*, *glorious*
/ʃ/-bias: *abolish*, *brackish*, *publish*, *galosh*

Stop

/t/-bias: *abrupt*, *carpet*, *secret*, *biscuit*, *product*
/d/-bias: *crooked*, *regard*, *placid*, *solid*, *garbled*

Specification of Male and Female Versions of Ambiguous Speech Sounds

The TRACE model feature input consists of seven banks of units that represent the value along each of seven acoustic/articulatory features. The phonemes /s/ and /ʃ/ are defined as having identical preferred values on five of the dimensions and different preferred values on the other two ("diffuse" and "acute"). The standard ambiguous fricative (used in Simulation 1 and in previous studies using the TRACE model; e.g., McClelland & Elman, 1986; Mirman et al., 2005) was created by setting five of the feature values to match both /s/ and /ʃ/ and the feature values for "diffuse" and "acute" to intermediate values that would be equally consistent with /s/ and /ʃ/. To create featurally different versions of an ambiguous fricative, an ambiguous value of "diffuse" and no value of "acute" were specified for the "male" version, and an ambiguous value of "acute" and no value of "diffuse" were specified for the "female" version. Thus, each ambiguous fricative was equally consistent with /s/ and /ʃ/, but their featural specifications were different with respect to the cues that distinguish /s/ and /ʃ/. This approach also allowed for the creation of unambiguous speaker-specific phonemes for the untuning simulations. As stated in the main text, this approach should not be interpreted as our view of the acoustic differences between male and female fricatives. Rather, this approach is a simple implementation of acoustically distinct ambiguous phonemes to allow straightforward testing of the acoustic similarity hypothesis.

The phonemes /d/ and /t/ differ in their values on the "voiced" and "burst" features, and the standard ambiguous coronal stop is defined by intermediate values on those features. To create featurally similar, but not identical, ambiguous coronal stops, the value along a different, nondistinctive feature ("power") was slightly changed: The "male" version had a slightly higher value, and the "female" version had a slightly lower value. The key aspect of this implementation was that the two versions of the ambiguous coronal stop were not identical but also did not differ with respect to the distinguishing features. As a result, lexically guided tuning that affected the distinction between /d/ and /t/ would generalize across the acoustic difference between the two versions. As with the fricatives, this implementation was designed to test the consequences of assuming that male and female versions of /d/ and /t/ do not differ with respect to the voicing feature, and it should not be taken as a statement about the acoustic difference between male and female stop consonants.