

Short communication

## Lexical effects on compensation for coarticulation: the ghost of Christmash past

James S. Magnuson<sup>a,\*</sup>, Bob McMurray<sup>b</sup>,  
Michael K. Tanenhaus<sup>b</sup>, Richard N. Aslin<sup>b</sup>

<sup>a</sup>*Department of Psychology, Columbia University, 1190 Amsterdam Ave.,  
MC 5501, New York City, NY 10027, USA*

<sup>b</sup>*University of Rochester, New York, NY, USA*

Received 9 September 2002; received in revised form 16 December 2002; accepted 25 December 2002

---

### Abstract

The question of when and how bottom-up input is integrated with top-down knowledge has been debated extensively within cognition and perception, and particularly within language processing. A long running debate about the architecture of the spoken-word recognition system has centered on the locus of lexical effects on phonemic processing: does lexical knowledge influence phoneme perception through feedback, or post-perceptually in a purely feedforward system? Elman and McClelland (1988) reported that lexically restored ambiguous phonemes influenced the perception of the following phoneme, supporting models with feedback from lexical to phonemic representations. Subsequently, several authors have argued that these results can be fully accounted for by diphone transitional probabilities in a feedforward system (Cairns et al., 1995; Pitt & McQueen, 1998). We report results strongly favoring the original lexical feedback explanation: lexical effects were present even when transitional probability biases were opposite to those of lexical biases.

© 2003 Cognitive Science Society, Inc. All rights reserved.

*Keywords:* Psychology; Language understanding; Neural networks

---

... Scrooge, having his key in the lock of the door, saw in the knocker, without its undergoing any intermediate process of change: not a knocker, but Marley's face.

–*A Christmas Carol*, Charles Dickens (1843).

A central question in cognitive science is when and how information sources are integrated. Fodor (1983) argued that modularity *between* perceptual systems would allow gains in

---

\* Corresponding author. Tel.: +1-212-854-5667; fax: +1-212-854-3609.

*E-mail address:* [magnuson@psych.columbia.edu](mailto:magnuson@psych.columbia.edu) (J.S. Magnuson).

processing efficiency and maximize veridical perception. Similar arguments have been made for purely modular or feedforward stages *within* systems, perhaps most notably in language processing (e.g., Frazier & Clifton, 1996; Norris, McQueen, & Cutler, 2000). On this view, protection of the bottom-up signal from top-down knowledge is necessary to prevent our days from being filled with hallucinations and ghostly apparitions like Marley's face in the knocker (the reality of Scrooge's perception notwithstanding).

An alternative view is that because signals occur in noise, immediate use of top-down knowledge makes processing more reliable by allowing knowledge and the processing context to constrain interpretation of the signal (e.g., McClelland, 1987, 1996; McClelland & Elman, 1986). The present report focuses on a particular question within this broader debate; does lexical knowledge affect sublexical processing?

In many spoken language tasks, the lexical status of a carrier sequence (i.e., whether it is a word or not) influences phonemic judgments. For example, lexical status affects response times in phoneme monitoring and judgments about whether a phoneme is present in a carrier sequence containing noise (phoneme restoration, e.g., Pitt & Samuel, 1995; Samuel, 1981, 1996, 1997, 2001; Warren, 1970; Warren & Warren, 1970). Lexical status also affects the category boundary in an identification task for consonants that vary along a continuum: when only one endpoint forms a word (e.g., a *dash–tash* or *dask–task* continuum), the category boundary shifts significantly toward the lexical end of the continuum (the “Ganong” effect; Fox, 1984; Ganong, 1980; Pitt, 1995).

Although lexical effects on phoneme identification are well documented, the theoretical explanation has been widely debated, primarily because it bears on long-standing debates about models of hierarchically organized cognitive architectures. According to one class of model, lexical knowledge affects phonemic processing via feedback. Initially, an ambiguous phoneme will equally activate both candidate phonemes, which will in turn activate relevant lexical representations. When one potential match to an ambiguous phoneme would make the stimulus conform to a word and another would not (e.g., *dash* vs. *tash* given an ambiguous alveolar stop consonant), activation of the word feeds back and boosts activation of its corresponding phoneme (*dash* will send feedback to /d/, boosting its activation relative to /t/). This explanation was implemented in the influential TRACE model (McClelland & Elman, 1986). Proponents of lexical feedback hold that it would also compensate for noise inherent in speech by providing constraints on the interpretation of a bottom-up signal: lexical feedback serves as an implicit encoding of the probability that a phoneme will occur in a given context.

An alternative class of model eschews feedback in favor of a purely feedforward system. For example, in the Race model (Cutler & Norris, 1979), phonemic decisions can be based on the output of either lexical or purely phonemic processing routes; the one to reach a threshold first, or the one attended to based on task constraints, provides the basis for the decision. In the Merge model (Norris et al., 2000), phoneme decisions are based on post-perceptual phoneme decision units that receive input from perceptual phoneme units and lexical units, avoiding the need for lexical feedback. Thus, the crucial difference in feedforward and feedback accounts is the locus of lexical effects on phonemes. In feedback accounts, lexical knowledge influences phonemic perception. In feedforward accounts, prior knowledge influences phonemic perception indirectly via one of two mechanisms. In one class of feedforward models, lexical knowledge affects post-perceptual decisions (Norris et al., 2000). In another, apparent

lexical effects result from precompiled sublexical knowledge, such as diphone transitional probabilities (Cairns, Shillcock, Chater, & Levy, 1995; Pitt & McQueen, 1998).

Elman and McClelland (1988) provided an apparently crucial test of these accounts by demonstrating lexical effects on *compensation for coarticulation*. Compensation for coarticulation (Mann & Repp, 1981; Repp & Mann, 1981, 1982) occurs when category boundaries in a phoneme identification task are shifted by the preceding coarticulatory context. Mann and Repp found that following a segment with an alveolar place of articulation (e.g., /s/), categorization of non-endpoint steps on an immediately following alveolar–velar continuum (/t/–/k/) was biased towards the velar place of articulation. The opposite shift was observed following a context with a velar place of articulation (e.g., /ʃ/). Thus, subjects are more likely to respond /k/ if the target immediately follows /s/, and /t/ if it follows /ʃ/.

Mann and Repp's interpretation of these effects was that, in natural production, when a velar or palatal segment must be produced immediately following an alveolar segment, the articulators are unlikely to reach the ideal target for the second segment. The result is a realization of the velar or palatal segment that is acoustically more similar than normal to its alveolar counterpart. They proposed that the speech perception system is tuned to dynamically shift category boundaries depending on context through perceptual learning, compensating for effects of coarticulation. Given an ambiguous segment in the compensation for coarticulation experimental paradigm, the perceptual system attributes the non-ideal realization of midpoint steps along the continuum to coarticulation due to the preceding segment.

Elman and McClelland (1988) combined compensation for coarticulation with the Ganong (1980) effect. If the Ganong effect results from lexical feedback to a perceptual phonemic level, then an ambiguous segment disambiguated and restored based on lexical status ought to drive compensation for coarticulation.

Elman and McClelland presented subjects with auditory contexts such as *fooliX* and *christmaX*, where *X* was a segment that, in isolation, was perceptually halfway between /s/ and /ʃ/. This ambiguous fricative was immediately followed by a word from a *tapes* to *capes* continuum. The expectation was that *X* would be perceived as /ʃ/ given *fooliX* and as /s/ given *christmaX* because of the Ganong effect. Then, if the restoration were due to true lexical influence on phonemic perception, the lexically restored /s/ or /ʃ/ percept should modulate the perception of the following /t/–/k/ continuum—that is, the lexically restored fricative percept should drive perceptual compensation for coarticulation.

Elman and McClelland's results supported this lexical-feedback prediction. Responses on the *tapes/capes* continuum were shifted towards *capes* following lexical contexts biased towards /s/ (e.g., *christmaX*), and towards *tapes* following contexts biased towards /ʃ/ (e.g., *fooliX*). These results have been challenged by claims that a purely feedforward model based on transitional probabilities among phonemes can account for the results without lexical representations. Cairns et al. (1995) analyzed the London–Lund corpus (Svartvik & Quirk, 1980) and reported that Elman and McClelland's items confounded diphone transitional probability (TP) with lexical status. That is, across a corpus of British English, they found the sequence / $\Delta$ s/ to be more likely than / $\Delta$ ʃ/, and /tʃ/ more likely than /ts/ (we discuss the specific statistic they used in detail later).

This allows an explanation of the lexical effects on compensation for coarticulation that does not invoke lexical representations: sensitivity to diphone TPs could explain the lexical

bias in the ambiguous fricative. Moreover, [Shillcock, Lindsey, Levy, and Chater \(1992\)](#); see also [Norris, 1993](#)) trained a recurrent network to output the previous, current and predicted next phoneme given phoneme-by-phoneme transcriptions of conversations from the same corpus and found that the network exhibited lexical effects on compensation for coarticulation.

These results from British English prompted [Pitt and McQueen \(1998\)](#) to devise an empirical test of the TP hypothesis in American English. They used two lexical contexts (*juice* and *bush*) in which the vowels were equally predictive of /s/ and /ʃ/. They contrasted these lexical contexts with nonword contexts with TP biases: *nai-*, biased towards /ʃ/, and *der-*, biased towards /s/. Pitt and McQueen predicted that if TP is the true basis for the lexical effects reported by [Elman and McClelland \(1988\)](#), then compensation for coarticulation should be observed in the nonword contexts (where TPs differed) but not in the lexical contexts (where TPs were equated). This is precisely what they found.

There are several reasons why we felt it was important to revisit whether there are lexical effects on compensation for coarticulation. First, [Pitt and McQueen's \(1998\)](#) result with equi-biased lexical contexts is a null effect, and must be interpreted with caution. Second, the tested lexical contexts and TPs were based on only two items, and may not generalize to other contexts. Third, as we later discovered, lexical status and TP were not confounded for all of the [Elman and McClelland \(1988\)](#) items in corpora of American English.

## 1. Experiment

We designed materials using the same /ʌ/ and /ɪ/ vowels that Elman and McClelland used, but we embedded them in contexts with opposite lexical biases. They found lexical effects on compensation for coarticulation with *Christmas*—we used *brush* to embed the same vowel in a context with the opposite (lexically-based) fricative bias. Elman and McClelland found lexical effects with *foolish*. We used *bliss* to embed the same vowel in a context with the opposite fricative bias. Using opposite lexical biases creates a strong test of whether there are lexical effects beyond effects of diphone TP.

We hypothesize that lexical status might have more powerful effects than diphone TPs because of the increased redundancy afforded by lexical information. We agree that diphone transitions could guide perception of ambiguous segments by combining the bottom-up signal with acquired knowledge of the most likely segments to follow. We argue that words, however, are more predictive because they span multiple segments and provide a compact, implicit representation of context-specific statistics.

### 1.1. Method

#### 1.1.1. Materials

We created an /s/-/ʃ/ continuum by recording natural, isolated utterances of the two fricatives. Samples were excised from the center of each fricative to make their durations 233 ms. These served as the endpoints of the continuum. We then created intermediate steps using a waveform averaging technique similar to that used by [Pitt and McQueen \(1998\)](#) for their *tapes/capes* continuum (see also [McQueen, 1991](#); [Repp, 1981](#)). We created weighted averages

of matrix representations of the /s/ and /ʃ/ waveforms in 2.5% steps. Thus, the /s/ endpoint was 100% /s/, 0% /ʃ/. The next step was 97.5% /s/ and 2.5% /ʃ/. We created 39 steps between the /s/ and /ʃ/ endpoints. Consistent with previous reports, our pilot identification tests with the /s-/ʃ/ continuum revealed substantial individual differences in the maximally ambiguous token. Rather than finding the maximally ambiguous token for each participant in our study (the procedure Pitt and McQueen used), and potentially alerting participants to our interest in the fricative component of the stimuli, we used two intermediate fricatives: the 50 and 60% /s/ tokens. Both were ambiguous for six pilot participants (mean for the 50% stimulus in isolation was 51% “s” responses; mean for the 60% stimulus was 58%, though the ranges overlapped).

To avoid coarticulatory cues in the vowels, the lexical contexts were naturally produced tokens of the first three phonemes of the words *bliss* and *brush*, i.e., /bli/ (318 ms) and /brʌ/ (314 ms).<sup>1</sup> Acoustic analyses comparing these with complete productions of *bliss* and *brush* did not reveal inadvertent coarticulatory cues. The two word-initial CCV contexts were combined with three fricative segments by appending them to the CCV: the appropriate endpoint (*bli-* + /s/, or *bru-* + /ʃ/), and the two ambiguous fricatives, 50 and 60% /s/.

A *tapes/capes* continuum was constructed in a similar fashion, except that the /t-/k/ endpoints were recorded in their full lexical contexts. The initial portion of each endpoint through the fourth pitch period in the vowel /eɪ/ was cut from the full lexical context. The resulting stimuli were 89 (/keɪ/) and 83 ms long (/teɪ/). To make the endpoints the same length, 6 ms were excised from the interior of the noise burst in /keɪ/. Pilot tests on the *tapes/capes* continuum indicated that the following seven steps on the continuum would yield non-ceiling/floor levels of “t” responses and a graded change from mainly “t” to mainly “k” responses: 45, 47.5, 50, 52.5, 55, 57.5, and 60% /t/. The auditory stimuli were recorded and presented with 16-bit resolution and a 22.05 kHz sampling rate.

### 1.1.2. Procedure

The experiment was conducted using PsyScope 1.2.5 (Cohen, MacWhinney, Flatt, & Provost, 1993). On each trial, participants heard one of six fricative-final stimuli (*bliss*, *brush*, *bli-50*, *bli-60*, *bru-50*, or *bru-60*, where 50 and 60 indicate the percentage /s/ for the ambiguous tokens) immediately followed by an item from the *tapes/capes* continuum. The task was similar to that used by Pitt and McQueen (1998): participants pressed one of four buttons, labeled “s t,” “s k,” “sh t,” and “sh k,” indicating the sequence of segments they heard at the word boundary.

On each trial, the four orthographic labels appeared on the screen, aligned with correspondingly labeled keys on the keyboard. After an 800 ms delay, one of the *bliss/brush* fricative stimuli was presented (*bliss*, *brush*, *bli-50*, *bli-60*, *bru-50*, or *bru-60*) followed immediately by one of the nine tokens from the *tapes/capes* continuum.

The experiment began with 36 practice trials to familiarize participants with the task. The practice trials consisted of two repetitions of each pairing of the endpoint *bliss/brush* stimuli with each of the nine continuum steps (the *tapes* and *capes* endpoints along with the seven intermediary steps) in random order. These practice trials were not included in the analyses.

Following the practice trials, there were 324 experimental trials, consisting of six repetitions of each of the 2 (lexical context) × 3 (fricative stimuli) × 9 (*tapes/capes* steps) combinations of the stimulus elements. These were presented in random order in six blocks of 54 trials.

### 1.1.3. Participants

Seventeen volunteers with normal hearing were paid for their participation. One participant's data was excluded because he always perceived the ambiguous fricative tokens as /s/.

## 1.2. Results and discussion

To determine whether the lexical contexts influenced responses to the ambiguous fricatives, we examined the proportion of “s” responses to the endpoint and ambiguous fricative stimuli at each *tapes/capes* step. The endpoints (*bliss* and *brush*) were responded to at ceiling and floor levels of /s/ response (96% [ranging from 93 to 98% across steps], 4% [ranging from 2 to 7%], respectively). Lexical context strongly affected responses to the ambiguous fricatives. Across all *tapes/capes* stimuli, the 50% /s/ token was labeled “s” 84% of the time (range: 79–88%) given the *bli-* context and labeled “sh” 93% of the time (range: 89–97%) given the *bru-* context; the 60% /s/ token was labeled “s” 86% of the time (range: 83–91%) in the *bli-* context, and labeled “sh” 92% of the time (range: 86–97%) given the *bru-* context. We conducted two ANOVAs (one each for 50 and 60% /s/ fricatives), with  $2 \times 9$  levels (lexical context  $\times$  *tapes/capes* steps) on the “s”-response proportions to verify that the pattern held across participants and *tapes/capes* level. The effect of lexical context was significant for both the 50% /s/ stimulus ( $F(1, 15) = 110.9, p < .001, \omega^2 = .77$ ) and the 60% /s/ stimulus ( $F(1, 15) = 121.2, p < .001, \omega^2 = .79$ ). The effect of step was not reliable for either ambiguous fricative, nor was the interaction of context and step. Thus, the lexical contexts were effective at shifting responses to the ambiguous fricatives, and the stimuli exhibit the prerequisite lexical effect on the fricative (the [Ganong, 1980](#) effect) for examining whether lexical bias influences compensation for coarticulation.

We next asked whether fricative perception affected compensation for coarticulation in the “t/k” responses. [Fig. 1](#) shows the effect of lexical context on the “t/k” responses. Proportions of “k” responses are plotted at each *tapes/capes* continuum step, with separate curves for *bli-* and *bru-* contexts and separate plots for each fricative type. We conducted ANOVAs on the proportion of “k” responses as a function of lexical context and ambiguous *tapes/capes* steps (steps 2–8; we excluded steps 1 and 9 because there is no reason to expect effects on unambiguous consonants). Thus, we conducted two (50 and 60% /s/)  $2 \times 7$  ANOVAs (lexical context  $\times$  *tapes/capes* continuum step).

For the endpoint stimuli (top panel), there were significant effects of context (*bli-* = 69% “k,” *bru-* = 49%;  $F(1, 15) = 41.4, p < .001, \omega^2 = .56$ ), step (ranging from 21 to 93%;  $F(6, 90) = 103.9, p < .001, \omega^2 = .85$ ), and a significant interaction of context and step ( $F(6, 90) = 2.5, p < .05, \omega^2 = .04$ ). This weak interaction depended on the relatively small effect at step 8; with step 8 removed, the interaction was not reliable ( $F(5, 75) = 1.4, p = .25$ ). In the case of the 50% /s/ fricative stimuli, there were significant effects of context (49% “k” responses given *bliss*, 43% given *brush*;  $F(1, 15) = 5.5, p < .05, \omega^2 = .12$ ) and step (ranging from 7 to 90%;  $F(6, 90) = 97.2, p < .001, \omega^2 = .84$ ), but the interaction was not significant ( $F < 1$ ). The pattern was the same for the 60% /s/ stimuli: there were significant effects of context (*bli-* = 52% “k” responses, *bru-* = 46%;  $F(1, 15) = 4.6, p < .05, \omega^2 = .19$ ) and step (ranging from 10 to 90%;  $F(6, 90) = 92.3, p < .001, \omega^2 = .83$ ) but the interaction was not reliable ( $F < 1$ ).

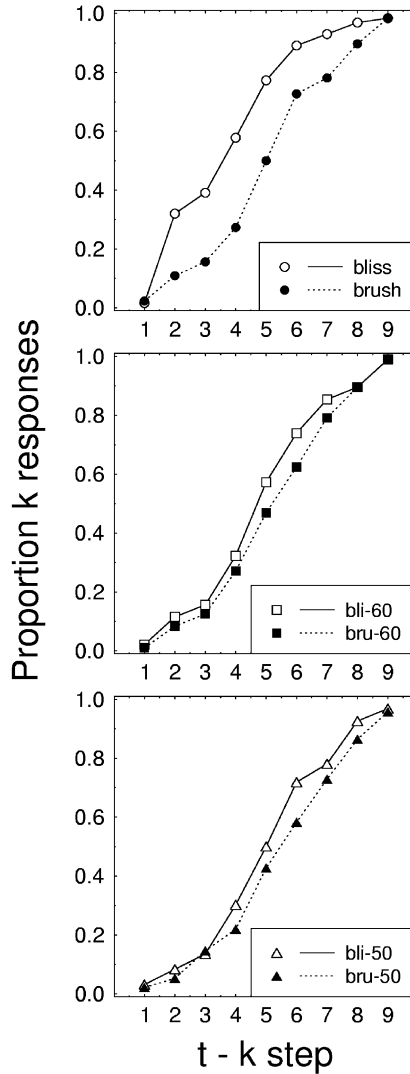


Fig. 1. Proportion of *tapes* responses at each step along the *tapes/capes* continuum as a function of the preceding fricative context. Top panel: endpoints (*bliss*, 100% /s/, and *brush*, 0% /s/). Middle panel: lexical contexts with ambiguous final fricative (60% /s/). Bottom panel: lexical contexts with ambiguous final fricative (50% /s/).

The most important result in each case is the main effect of lexical context, indicating that compensation for coarticulation was modulated by lexical status.<sup>2</sup> This is consistent with a recent study by Samuel and Pitt (2003). They tested a number of items with the same vocalic context but opposite lexical biases and found lexical effects on compensatory coarticulation for most of their items. Since the ambiguous fricative was always preceded by the same vocalic context, Samuel and Pitts' results cannot be accounted for by diphone TPs.

How can we explain the divergence between Pitt and McQueen (1998), who did not find lexical compensation with TP-neutral contexts, and Elman and McClelland (1988), Samuel

and Pitt (2003), and our study (all of which found lexically mediated compensation, even when TPs were at odds with lexical bias)? One possibility is that differences in stimulus preparation techniques are responsible. For example, our ambiguous fricatives were highly pliable (cf. the large changes in “s” responses depending on lexical context). Another possibility is that there is something unusual about the two lexical items Pitt and McQueen used, and another factor overrode the lexical bias. Indeed, Samuel and Pitt found that the compensation effect is strongly influenced by perceptual grouping phenomena, and, in a separate test of perceptual grouping, that fricatives cohered more strongly with Pitt and McQueen’s lexical than nonword contexts, making them less susceptible to compensation.

A third possibility is that the corpus analyses used to determine diphone TPs were sufficiently different in the various studies that the resultant selection of stimulus items was unbalanced. According to the corpus analyses reported by Cairns et al. (1995), the lexical biases used in our stimulus materials were opposite to the TP biases for our chosen vowel-fricative contexts. However, there may be higher-order TPs correlated with lexical context. In order to test whether a more elaborate phoneme-based TP explanation might account for our results, we conducted a series of corpus analyses.

## 2. Corpus analyses

We analyzed two pronunciation dictionaries, Moby and MIT,<sup>3</sup> weighted by the frequency counts in Francis and Kucera (1982), and one phonemically transcribed corpus of spoken, conversational American English (the CALLHOME corpus; Kingsbury, Strassel, McLemore, & McIntyre, 1997a).<sup>4</sup> Forward and backward TPs based on the MIT and CALLHOME corpora are shown in Table 1 (results with the MIT and Moby lexicons are nearly identical). The only contexts strongly biased towards /f/ were /ʊ/ and /et/. Even if we compute the statistic used by Cairns et al. (1995)—backward TP (which we hold is the incorrect statistic<sup>5</sup>)—we find a similar pattern. Across all corpora, five contexts are biased towards /f/, none of which corresponds to a context lexically biased towards /f/ used by Elman and McClelland or us. Thus, the claim that TP was confounded with lexical status in the crucial contexts in Elman and McClelland’s materials does not hold (at least not for American English). We also tested whether larger *n*-phone TPs might account for the results of Experiment 1. Table 2 presents forward and backward TPs for all possible preceding TP contexts (V, CV, CCV) and targets (fricative [F], VF, CVF, CCVF) based on the MIT lexicon (similar patterns hold for Moby and CALLHOME). None of the context/target combinations account for both the *bliss* and *brush* effects. Thus, neither simple diphone TPs nor more complex TPs account for our results.<sup>6</sup>

Perhaps the explanation does not lie with lexical knowledge *per se*, but with higher-order statistics than simple TPs. While we explored a large set of TPs, it is possible that another analysis exists which would account for the results. However, as one invokes higher- and higher-order statistics, several problems will emerge. For example, what is the proper order? If diphones do not suffice, a fixed *n*-phone TP model (where  $n > 2$ ) will not be able to account for lexical effects with two-segment words.

Critics of feedback in spoken word recognition have appealed to the fact that simple recurrent networks can simulate the Elman and McClelland (1988) effects without explicit lexical

Table 1

Forward and backward diphone transitional probabilities for the occurrence of /s/ and /ʃ/ given vowels based on a frequency-weighted written lexicon (MIT) and one spoken corpus (CALLHOME) of American English

|   | Forward TP   |              |      | Backward TP   |              |      |
|---|--------------|--------------|------|---------------|--------------|------|
|   | $p(f V)$     | $p(s V)$     | Bias | $p(V f)$      | $p(V s)$     | Bias |
| MIT (written frequency-weighted lexicon)        |              |              |      |               |              |      |
| æ   | .0068        | <b>.0288</b> | s    | .0359         | .0293        | eq   |
| ɛ   | .0162        | <b>.0759</b> | s    | .0510         | .0457        | eq   |
| ɪ   | .0164        | <b>.0804</b> | s    | .1487         | .1401        | eq   |
| ɑ <sup>w</sup>                                  | .0000        | <b>.1808</b> | s    | .0000         | <b>.0035</b> | s    |
| ʌ   | .0044        | <b>.0892</b> | s    | .0094         | <b>.0366</b> | s    |
| ɚ   | .0032        | <b>.0576</b> | s    | .0085         | <b>.0298</b> | s    |
|   | <b>.0048</b> | .0005        | ʃ    | <b>.0032</b>  | .0001        | ʃ    |
| ɔɪ  | .0000        | <b>.0309</b> | s    | .0000         | <b>.0044</b> | s    |
| aɪ  | .0000        | <b>.0297</b> | s    | .0000         | <b>.0089</b> | s    |
| ɑ   | .0002        | <b>.0411</b> | s    | .0003         | <b>.0128</b> | s    |
| ɔ   | .0024        | <b>.0312</b> | s    | .0036         | <b>.0090</b> | s    |
| eɪ  | <b>.1295</b> | .0695        | ʃ    | <b>.2316</b>  | .0238        | ʃ    |
| i   | .0013        | <b>.0209</b> | s    | .0049         | <b>.0151</b> | s    |
| o   | .0122        | <b>.0447</b> | s    | .0313         | .0220        | eq   |
| u   | .0081        | .0084        | eq   | <b>.0219</b>  | .0044        | ʃ    |
| ɚ   | .0014        | <b>.0211</b> | s    | .0136         | <b>.0390</b> | s    |
| CALLHOME (spoken telephone conversation corpus) |              |              |      |               |              |      |
| æ   | .0033        | <b>.0931</b> | s    | .02091        | <b>.0913</b> | s    |
| ɛ   | .0070        | <b>.0779</b> | s    | .02660        | <b>.0465</b> | s    |
| ɪ   | .0107        | <b>.0610</b> | s    | .09343        | .0837        | eq   |
| ɑ <sup>w</sup>                                  | .0057        | <b>.0625</b> | s    | .00045        | <b>.0007</b> | s    |
| ʌ   | .0048        | <b>.0382</b> | s    | .03001        | .0372        | eq   |
| ɚ   | .0103        | <b>.0985</b> | s    | .02137        | .0319        | eq   |
|   | <b>.0041</b> | .0000        | ʃ    | <b>.00341</b> | .0000        | ʃ    |
| ɔɪ  | .0072        | .0107        | eq   | <b>.00705</b> | .0016        | ʃ    |
| aɪ  | .0038        | <b>.0508</b> | s    | .01910        | <b>.0399</b> | s    |
| ɑ   | .0089        | <b>.0149</b> | s    | <b>.01637</b> | .0043        | ʃ    |
| ɔ   | .0064        | <b>.0305</b> | s    | .00796        | .0059        | eq   |
| eɪ  | .0229        | <b>.0557</b> | s    | <b>.08707</b> | .0332        | ʃ    |
| i   | .0083        | <b>.0484</b> | s    | .05024        | .0457        | eq   |
| o   | .0139        | <b>.0305</b> | s    | <b>.07229</b> | .0248        | ʃ    |
| u   | .0071        | <b>.0579</b> | s    | .02842        | .0363        | eq   |
| ɚ   | .0026        | <b>.253</b>  | s    | .00796        | <b>.0122</b> | s    |

“Bias” for /s/ or /ʃ/ was operationalized as one TP being 1.5 times greater than the other; biases are indicated by bold type.

knowledge (Cairns et al., 1995; Norris, 1993)—when TP and lexical biases are correlated in the training corpus, which Cairns et al. (1995) reported to be true of English.<sup>7</sup> However, such models do learn to represent word-specific statistics. Recurrent networks have the potential to make use of recent history (e.g., by making the states of some units at time *t* part of the input at time *t* + 1), representing context-specific statistics over potentially large temporal windows

Table 2

All possible forward and backward fricative transitional probabilities based on the MIT corpus

|     | Forward transitional probability |                  |      | Backward transitional probability |                  |      |
|-----|----------------------------------|------------------|------|-----------------------------------|------------------|------|
|     | $p(f context)$                   | $p(s context)$   | Bias | $p(context f)$                    | $p(context s)$   | Bias |
| i   | .0164                            | <b>.0804</b>     | s    | .1487                             | .1401            | eq   |
| Λ   | .0044                            | <b>.0892</b>     | s    | .0094                             | <b>.0366</b>     | s    |
| li  | .0405                            | <b>.0956</b>     | s    | <b>.0174</b>                      | .0079            | f    |
| lΛ  | .0220                            | <b>.1199</b>     | s    | .0012                             | .0012            | eq   |
| bli | .1043                            | .0735            | eq   | <b>.0030</b>                      | .0004            | f    |
| brΛ | .0131                            | .0131            | eq   | .0001                             | <.0001           | eq   |
|     | $p(lf context)$                  | $p(lis context)$ | Bias | $p(context lf)$                   | $p(context lis)$ | Bias |
| l   | .0050                            | <b>.0118</b>     | s    | <b>.1172</b>                      | .0562            | f    |
| bl  | .0269                            | .0190            | eq   | <b>.0201</b>                      | .0029            | f    |
|     | $p(Λf context)$                  | $p(Λs context)$  | Bias | $p(context Λf)$                   | $p(context Λs)$  | Bias |
| l   | .0003                            | <b>.0018</b>     | s    | <b>.1250</b>                      | .0336            | f    |
| br  | .0009                            | .0009            | eq   | <b>.0104</b>                      | .0005            | f    |
|     | $p(lif context)$                 | $p(lis context)$ | Bias | $p(context lif)$                  | $p(context lis)$ | Bias |
| b   | .0012                            | .0009            | eq   | <b>.1718</b>                      | .0514            | f    |
|     | $p(rΛf context)$                 | $p(rΛs context)$ | Bias | $p(context rΛf)$                  | $p(context rΛs)$ | Bias |
|     | <.0001                           | <.0001           | eq   | <b>.0833</b>                      | .0153            | f    |

None accounts for both the *bliss* and *brush* compensation for coarticulation effects.

of *context-dependent size*. In this case, the learned dependencies are regularities controlled by words (e.g., lexical items control inter-phoneme statistical dependencies, with the weakest TPs at word boundaries; e.g., [Harris, 1955](#)). Thus, these networks encode a dynamic statistical representation, such that for a given word, the crucial “*n*-phone” resolves to word length. A fixed *n*-phone model cannot work, since the effective size of *n* is context dependent; indeed, the relevant context is best described as lexical. In other words, lexical knowledge subsumes the relevant statistics (cf. [McClelland & Elman, 1986](#)).

Feedback between explicit lexical and phonemic representations provides an efficient way of instantiating this knowledge for processing. Recurrent networks may represent such knowledge and feedback via context or history units. However, to date, models without explicit lexical representations have not simulated lexical biases on compensation for coarticulation when TP and lexical biases are at odds (i.e., lexical bias and TP bias have been correlated in the training corpora, which has been the basis for arguing that lexical knowledge is unnecessary for explaining the phenomenon). Our corpus analysis demonstrates that both we and [Elman and McClelland \(1988\)](#) have found this result. The challenge now is for models without lexical representations to simulate the dominance of lexical bias over TP bias. We expect this is possible with models like those used by [Cairns et al. \(1995\)](#) and [Norris et al. \(2000\)](#). However, we predict that this will not be possible with a truly bottom-up model—one whose interpretation

of the current bottom-up signal is not constrained by recent context beyond the level of diphones.

### 3. Conclusions

Our results demonstrate a clear influence of lexical knowledge on spoken word recognition that is distinct from lower-order TPs. Using the same vocalic contexts as Elman and McClelland (1988) but opposite lexical biases, we found reliable, lexically-mediated compensation for coarticulation. Corpus analyses showed that lexical bias and diphone TP were not confounded in all of Elman and McClelland's original stimuli, and ruled out possible explanations based on diphone TPs as well as other plausible higher-order phonemic TPs. Our results are supported by recent converging evidence reported by Samuel and Pitt (2003) that we discussed earlier. Thus, the preponderance of evidence clearly indicates that the perceptual phenomenon of compensation for coarticulation is mediated by lexical information. While TPs also have an influence (Pitt & McQueen, 1998), our study shows that the influence of lexical status is stronger, as it dominates when lexical and TP biases are in opposition. This pattern of results strongly favors spoken word recognition models incorporating lexical feedback, and, by extension, the importance of feedback in perceptual processing.

### Notes

1. In pilot studies using other materials, we found that coarticulatory cues in the vowels of the stimuli to which the ambiguous fricatives were attached had strong effects on compensation; given lexically-consistent coarticulation, we found compensation effects nearly as large as those found with unambiguous endpoint fricatives; given lexically-inconsistent coarticulation, we found almost no compensation.
2. While the effect sizes for the ambiguous fricatives were substantially smaller than for the endpoint stimuli, they were medium ( $.06 < \omega^2 < .15$ ) and large ( $\omega^2 > .15$ ) effects (Cohen, 1977).
3. Moby was developed and placed in the public domain by Grady Ward. See <http://www.speech.cs.cmu.edu/comp.speech/Section1/Lexical/moby.html>. The MIT lexicon was developed from the 1965 *Webster's Pocket Dictionary* for the analyses reported by Shipman and Zue (1982).
4. This phonemic corpus was created by replacing each orthographic form in the transcripts of the CALLHOME corpus with the phonemic form in the CALLHOME lexicon, PRON-LEX (Kingsbury et al., 1997b).
5. The causal relationship between phoneme restoration and backward transitional probability—that is, the probability that the preceding segment was a particular vowel given the identity of the current fricative—is not clear, especially given that *ambiguous* fricatives were used in the stimulus materials.
6. Given relationships between TP and neighborhood density (e.g., Vitevitch & Luce, 1999), frequency and neighborhood density could influence compensation. *Bliss* occurs

four times per million words (Francis & Kucera, 1982), while *brush* occurs 36 times as a noun and 38 as a verb. *Bliss* has six neighbors (words differing by one phoneme) and a neighborhood density of 4.9 (summed log neighbor frequencies), while *brush* has 11 neighbors and a neighborhood density of 9.4. *Bliss* and *brush* had similar frequency-weighted neighborhood probabilities (Luce & Pisoni, 1998): *bliss* = .12, *brush* = .17. Since the items have similar neighborhood densities and probabilities, our results cannot be attributed to differences in these dimensions.

7. A second element of the argument is that such networks are purely bottom-up. We do not have space to address this point in detail here, but we disagree, because such networks are trained via an error signal fed back through the network, and, after training, performance depends on knowledge encoded in connection weights. Part of the input to an SRN, for example, comes from context units representing the state of the hidden units at the previous time step. The knowledge in connection strengths is top-down in that the bottom-up input is interpreted with respect to states of units fed back after they have been processing the preceding context. Indeed, when Norris et al. (2000) allow the possibility making Merge sensitive to TPs, they implicitly abandon their position that a word recognition system could not do better than to directly map from the bottom-up signal to representations; a TP-sensitive system would constrain the bottom-up mapping with context-specific knowledge stored in memory.

## Acknowledgments

This work was supported by NSF SBR-9729095 and NIDCD, DC-005071 to MKT and RNA, an NSF Graduate Research Fellowship to JSM, and a Grant-in-Aid of Research from the National Academy of Sciences, through Sigma Xi, to JSM. We thank Howard Nusbaum for helpful theoretical discussions, Gary Dell, Jeff Elman and an anonymous reviewer for comments that strengthened this paper, Dana Subik for assistance in running the experiment, and Gina Cardillo for help in preparing the manuscript.

## References

- Cairns, P., Shillcock, R., Chater, N., & Levy, J. P. (1995). Bottom-up connectionist modeling of speech. In J. P. Levy & D. Bairaktaris (Eds.), *Connectionist models of memory and language* (pp. 289–310). London, England: UCL Press Limited.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (revised edition). New York: Academic Press.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, 25, 257–271.
- Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113–134). Hillsdale, NJ: Erlbaum.
- Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, 27, 143–165.

- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 526–540.
- Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton-Mifflin.
- Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.
- Ganong, W. F. (1980). Phonetic categorization in auditory perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110–125.
- Harris, Z. S. (1955). From phoneme to morpheme. *Language*, 31, 190–222.
- Kingsbury, P., Strassel, S., McLemore, C., & McIntyre, R. (1997a). *CALLHOME American English transcripts, LDC97T14*. Philadelphia: Linguistic Data Consortium.
- Kingsbury, P., Strassel, S., McLemore, C., & McIntyre, R. (1997b). *CALLHOME American English lexicon (PRON-LEX), LDC97L20*. Philadelphia: Linguistic Data Consortium.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36.
- McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 3–36). Hillsdale, NJ: Erlbaum.
- McClelland, J. L. (1996). Integration of information: Reflections on the theme of attention and performance XVI. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI. Information integration in perception and communication* (pp. 633–656). Cambridge, MA: MIT Press.
- Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, 69, 548–558.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception & Performance*, 17, 433–443.
- Norris, D. (1993). Bottom-up connectionist models of ‘interact.’ In G. T. M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 211–234). Hillsdale, NJ: Erlbaum.
- Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral & Brain Sciences*, 23, 299–370.
- Pitt, M. A. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 1037–1052.
- Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, 39, 347–370.
- Pitt, M. A., & Samuel, A. G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology*, 29, 149–188.
- Repp, B. H. (1981). Perceptual equivalence of two kinds of ambiguous speech stimuli. *Bulletin of the Psychonomic Society*, 18, 12–14.
- Repp, B. H., & Mann, V. A. (1981). Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America*, 69, 1154–1163.
- Repp, B. H., & Mann, V. A. (1982). Fricative-stop coarticulation: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 71, 1562–1567.
- Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, 110, 474–494.
- Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, 125, 28–51.
- 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 and Language*, 48, 416–434.

- Shillcock, R. C., Lindsey, G., Levy, J., & Chater, N. (1992). A phonologically motivated input representation for the modeling of auditory word perception in continuous speech. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 408–413). Mahwah, NJ: Erlbaum.
- Shipman, D. W., & Zue, V. W. (1982). Properties of large lexicons: Implications for advanced isolated word recognition systems. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 546–549). Piscataway, NJ: IEEE.
- Svartvik, J., & Quirk, R. (1980). *A corpus of English conversation*. Lund: Gleerup.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40, 374–408.
- Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, 167, 392–393.
- Warren, R. M., & Warren, R. P. (1970). Auditory illusions and confusions. *Scientific American*, 223, 30–36.