Cognitive Science 27 (2003) 795–799
COGNITIVE SCIENCE
http://www.elsevier.com/locate/cogsci

Letter to the Editor

# The ghost of Christmas future: didn't Scrooge learn to be good? Commentary on Magnuson, McMurray, Tanenhaus, and Aslin (2003)

James M. McQueen*

*Max Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands*

Received 3 February 2003; received in revised form 14 April 2003; accepted 15 April 2003

**Abstract**

Magnuson, McMurray, Tanenhaus, and Aslin [Cogn. Sci. 27 (2003) 285] suggest that they have evidence of lexical feedback in speech perception, and that this evidence thus challenges the purely feedforward Merge model [Behav. Brain Sci. 23 (2000) 299]. This evidence is open to an alternative explanation, however, one which preserves the assumption in Merge that there is no lexical–prelexical feedback during on-line speech processing. This explanation invokes the distinction between perceptual processing that occurs in the short term, as an utterance is heard, and processing that occurs over the longer term, for perceptual learning.


*Keywords:* Speech perception; Lexical feedback; Neural networks

Norris, McQueen, and Cutler (2003) show that perceptual learning in the speech recognition system can be modulated by lexical knowledge. Dutch listeners heard words and nonwords including 20 words ending in each of [f] and [s] (these fricatives occurred nowhere else). For some listeners, an ambiguous sound [?] (midway between [f] and [s]) replaced the [f] sounds, and the [s] sounds were natural. For other listeners, the [f]'s were natural and the [s]'s became [?]. The first group thus heard words such as *witlo?* (*witlof* means chicory; *witlos* is a nonword), and hence could use the lexicon to infer that [?] was [f]. The second group heard words such as *naaldbo?* (*naaldbos* means pine forest; *naaldbof* is a nonword) and thus could infer that [?] was

* Tel.: +31-24-3521373; fax: +31-24-3521213.
*E-mail address:* james.mcqueen@mpi.nl (J.M. McQueen).

0364-0213/$ – see front matter 
doi:10.1016/S0364-0213(03)00069-7

[s]. During this exposure, the listeners made lexical decisions; they judged most [?]-final items to be words. After exposure, listeners categorized an ambiguous [f]-[s] continuum. The first group labeled significantly more of these fricatives as [f] than the second group. The listeners had adjusted their phonetic categories, based on lexical information.

This result indicates that lexical information can be used in perceptual learning. Current research is attempting to clarify the nature of this kind of learning (Eisner & McQueen, 2003). But it is certainly plausible to assume that it has a prelexical locus (i.e., occurs at a level of processing preceding lexical access). Adjustments to prelexical representations of fricatives would facilitate lexical access during continued exposure to the speaker producing those unusual sounds (who might be, e.g., a speaker of an unfamiliar dialect). The findings of Norris et al. (2003) thus show that there is lexical feedback for perceptual learning. Norris et al. (2000, 2003) suggested that on-line feedback, where lexical knowledge influences prelexical processing of an utterance as that utterance is heard, serves no useful function. They also point out that on-line feedback can distort perception to produce hallucinations (like Scrooge's vision of Marley). In contrast, the effect of feedback for learning is likely to be entirely beneficial to the listener.

Magnuson et al. (2003; MMTA hereafter) show that listeners are more likely to label an ambiguous [t]-[k] sound as [k] when that sound follows *bliss* than when it follows *brush*, confirming that perception of stops can be biased by preceding fricatives (Mann & Repp, 1981). They also show that this bias can occur after an ambiguous fricative (between [s] and [ʃ]), where fricative identity is provided by lexical knowledge (i.e., after *bli?* and *bru?*). MMTA interpret this result (as did Elman & McClelland, 1988) as evidence that there is on-line feedback of fricative information from the lexicon to the prelexical level (the assumed locus of the compensation process which adjusts stop perception according to the fricative context).

But consideration of the Norris et al. (2003) findings leads to an alternative explanation. The MMTA result could be due to perceptual learning rather than on-line feedback. The listeners heard unambiguous fricatives only in lexically consistent contexts (i.e., [s] in *bliss* and [ʃ] in *brush*). They could thus have learned, as the experiment proceeded, that [s] was more likely after [blɪ] and that [ʃ] was more likely after [brʌ]. This contextual bias on interpretation of the ambiguous fricative could have induced the compensation process. Note, however, that this kind of learning would not necessarily depend on the top-down use of lexical knowledge (as does the learning in Norris et al.'s study), since the contextual dependencies were present in the stimuli themselves. In short, the possibility that perceptual learning may have taken place means that the MMTA experiment does not provide unambiguous evidence that the lexicon influences prelexical processing of fricatives as that fricative information is heard, that is, via on-line feedback.

In the original Elman and McClelland (1988) study, and a recent study by Samuel and Pitt (2003), experiment-internal contextual biases were controlled. Lexically mediated shifts in stop identification after ambiguous fricatives were nevertheless observed. This apparent evidence for feedback is also questionable, however. It needs to be established whether the effects in these two studies are due to lexical knowledge per se, or to sensitivity to the likelihood of [s] and [ʃ] in particular phonological contexts. Experiment 1 in Elman and McClelland may or may not have vowel–consonant diphone transitional probability (TP) biases, depending on which TP counts one considers. Diphone TPs were controlled in Elman and McClelland's

Experiment 3, and by Samuel and Pitt, but in none of these experiments were higher-order TPs controlled (i.e., TPs spanning units larger than diphones). Such TP biases, rather than lexical biases, could thus be responsible for the effects found in those experiments. (Note, however, that this criticism does not apply to the MMTA experiment, where both diphone and higher-order TPs were controlled.)

Given the possibility that the MMTA results may be due to contextual biases in their experiment, and the possibility that higher-order TP biases could underlie all other apparent lexical effects in compensation for coarticulation, it is premature to conclude that there is on-line feedback during speech recognition. Indeed, there is evidence from the compensation paradigm which suggests that there is no such feedback. Pitt and McQueen (1998) asked listeners to identify fricatives and stops (e.g., in the sequence *jui? ?apes*, listeners judged whether the ambiguous word-final sound was [s] or [ʃ] and whether the ambiguous word-initial sound was [t] or [k]). A dissociation was found between identification of the fricatives and the stops: Listeners showed a lexical bias on fricative identification (e.g., more [s] responses to *jui?*) on the same trials for which there was no lexical bias on stop identification. This pattern of results is consistent with the feedforward Merge model (Norris et al., 2000), but challenges the on-line feedback account: If feedback were adjusting the prelexical interpretation of the ambiguous fricative [?], as the fricative judgments might suggest, the prelexical compensation process ought to have adjusted stop identification. Any account that postulates on-line feedback to explain contextual biases on fricative-stop compensation must also be able to explain this critical dissociation. MMTA do not offer such an explanation.

The debate on feedback that centers on compensation for coarticulation is therefore not yet resolved. The ghost of Christmas future warns us, however, that further research in this area will not be easy. As Samuel and Pitt (2003) show, it is hard to engineer the specific conditions required for the emergence of contextual biases in fricative-stop compensation (lexical biases were only observed in four of the eight stimulus sets they tested). In addition, given the role of TPs in the effect, more accurate measures of the frequency of occurrence of particular phonetic sequences are required. For example, it seems inappropriate in such frequency counts to consider the vowel in *brush* ([ʌ]) to be the same as the last vowel in *Christmas* ([ə]), as MMTA do. In my dialect of English at least, irrespective of speaking rate and speaking style, these vowels differ in both quality and stress ([ʌ] is full; [ə] is reduced).

The ghost of Christmas future also tells us that further research on compensation will involve higher-order TPs. It will be necessary to establish whether higher-order TP effects on phoneme identification exist, and what the window size is over which sequential dependencies are computed. MMTA, in their discussion of statistical learning in simple recurrent networks, make the important point that the window size in such models is likely to be of variable length—for the sequence of phonemes in any given word, the effective context is likely to be the length of that word. It is an empirical question whether this is true for human listeners (while inter-phoneme dependencies are likely to be determined largely by word-internal sequences, common between-word sequences could also have some role). Let us assume, however, for the sake of argument, that listeners are sensitive to higher-order TPs, and that these phoneme likelihoods are computed over word-sized units. MMTA argue that lexical-prelexical feedback would offer an efficient way of instantiating this statistical knowledge (i.e., that higher-order TP effects would be lexical effects). But it is also possible that this knowledge could be coded entirely at the prelexical level, as it is in the recurrent network simulations of Elman and McClelland's (1988) data (Cairns, Shillcock, Chater, & Levy, 1995; Norris, 1993). The networks in these simulations contained no explicit lexical representations and were simply trained to predict the phonemes in the input. The simulations showed that a probabilistic bias in fricative identification at the prelexical level of processing could trigger compensatory adjustments in stop identification at that same level. It should also be emphasized that, contrary to what MMTA suggest, statistical learning does not require top-down feedback from the lexicon, since the sequential dependencies are present in the bottom-up speech signal itself (Norris, 1993). That is, the sequential regularities in the speech stream can be learned and stored prelexically, without the involvement of the lexicon. Higher-order TP effects in the compensation paradigm would therefore not necessarily be evidence of feedback. Experiments testing whether higher-order TP effects in compensation are distinct from lexical effects (i.e., whether they dissociate, like the diphone TP and lexical effects in Pitt & McQueen, 1998) will be necessary to distinguish between the feedback and bottom-up learning accounts of TP effects.

The most important lesson the ghost of Christmas future has for us, however, concerns the more general issue of perceptual learning in speech recognition. The distinction between learning and on-line processing needs to be considered in the debate on feedback in speech perception (as in other domains, see, e.g., Pylyshyn, 1999). Effects in the compensation paradigm (or other speech perception paradigms) that purport to be evidence of on-line feedback may instead be due to learning about contextual biases within an experiment. Furthermore, such effects might be due to longer-term statistical learning about the language's phonological structure (i.e., learning about TPs; Pitt & McQueen, 1998), or to an off-line lexical influence on perceptual learning (see Norris et al., 2003, for further discussion). Critically, none of these three types of learning requires the on-line modulation of prelexical processing through feedback from the lexicon.

## Acknowledgments

I thank Anne Cutler, Frank Eisner, Jim Magnuson, Dennis Norris and an anonymous reviewer for useful feedback.

## References

Cairns, P., Shillcock, R., Chater, N., & Levy, J. P. (1995). Bottom-up connectionist modeling of speech. In J. P. Levy, D. Bairaktaris, J. A. Bullinaria, & P. Cairns (Eds.), *Connectionist models of memory and language* (pp. 289–310). London: UCL Press.

Eisner, F., & McQueen, J. M. (2003, June). *Does perceptual learning in speech perception generalize across speakers?* Presented at the second Dutch Endo-Neuro-Psycho Meeting, Doorwerth, The Netherlands.

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.

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

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.

Norris, D. (1993). Bottom-up connectionist models of 'interaction'. 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 and Brain Sciences, 23,* 299–325.

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

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, *39*, 347–370.

Pylyshyn, Z. (1999). Is perception continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*, 341–423.

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.