

University of Montana

ScholarWorks at University of Montana

Anthropology Faculty Publications

Anthropology

6-2000

Merging Information Versus Speech Recognition

Irene Appelbaum

University of Montana - Missoula, irene.appelbaum@umontana.edu

Follow this and additional works at: https://scholarworks.umt.edu/anthro_pubs



Part of the [Anthropology Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Appelbaum, Irene, "Merging Information Versus Speech Recognition" (2000). *Anthropology Faculty Publications*. 4.

https://scholarworks.umt.edu/anthro_pubs/4

This Response or Comment is brought to you for free and open access by the Anthropology at ScholarWorks at University of Montana. It has been accepted for inclusion in Anthropology Faculty Publications by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.

ACKNOWLEDGMENTS

We thank Steve Levinson for suggesting the name of the Merge model to us. We also thank Peter Jusczyk, Antje Meyer, and Joanne Miller for very helpful comments on the original version of this paper, and Dom Massaro, Arthur Samuel, Mark Pitt and two anonymous reviewers for later comments. Please address correspondence to: Dennis Norris, MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 2EF, UK; E-mail: dennis.norris@mrc-cbu.cam.ac.uk.

NOTE

1. In fact, in TRACE, this is not strictly true. Feedback in TRACE can actually alter the pattern of sensitivity (Massaro 1989a), but this is a consequence of the simplifying assumptions made in implementing it. All processing is performed without noise, and response probabilities are derived by applying the Luce choice rule to the resulting activations. The noise simulated by this rule is independent of all other processing. Top-down feedback can alter the signal (arbitrarily) during processing without altering noise. That is, feedback can alter sensitivity. When noise is added to these systems at input, or during processing, and the Luce rule is replaced by the Best One Wins decision rule, feedback will alter both signal and noise together, leaving sensitivity unchanged (McClelland 1991). So, the tendency of TRACE to show effects of feedback on sensitivity is really the result of the simplified, but unrealistic, noise-free processing assumption. This problem can be remedied by replacing the IAM with a Stochastic Interactive Activation Model (McClelland 1991).

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Merging information versus speech recognition

Irene Appelbaum¹

Department of Philosophy, University of Montana, Missoula, MT 59812.
appel@selway.umt.edu

Abstract: Norris, McQueen & Cutler claim that all known speech recognition data can be accounted for with their autonomous model, “Merge.” But this claim is doubly misleading. (1) Although speech recognition is autonomous in their view, the Merge model is not. (2) The body of data which the Merge model accounts for, is not, in their view, speech recognition data.

Norris, McQueen & Cutler claim that all known speech recognition data can be accounted for with an autonomous model (“Merge”) that merges lexical and prelexical information. This claim elicits an immediate objection. Autonomous models, by definition, are non-interactive. Merging, on the face of it, is a form of interaction. How can a model that features interaction be defended as one that does not? In response, Norris et al. waver between denying that merging is interaction and denying that it is problematic, if it is. In particular, they imply that the part of Merge that does what they call “phonemic decision-making” need not be autonomous, because phonemic decision-making is not part of ordinary speech recognition. But this response gives rise to a second difficulty. If phonemic decision-making is distinct from speech

recognition, then the data Merge accounts for – phoneme decision data – is not speech recognition data, making their central claim misleading.

Four features distinguish Merge from TRACE. (1) Merge consists of two phonemic stages: a first “processing” stage and a second “decision-making” stage. (2) Because the first (processing) stage lacks intra-node inhibition, its output preserves phonemic ambiguity. (3) Connections between lexical and phonemic information are flexible – built “on the fly” in response to task demands and attentional shifts. And crucially: (4) Lexical information does not affect the first (processing) stage; instead, lexical information and the output of the first phonemic (processing) stage are merged in the second (decision) stage, which outputs a determinate phonemic identification.

So: the first (processing) stage gets no feedback from either the lexicon or the second (decision-making) stage; the lexicon gets input only from the first (processing) stage; and the second (decision-making) stage gets information from both the first (processing) stage and the lexicon. Given this architecture, the first (processing) stage does seem to meet normal standards of autonomy. The question is what to say about the second (decision-making) stage. Is it also autonomous? It would seem not to be since “phoneme decision units are . . . directly susceptible to facilitatory influences from the lexicon” (sect. 5.1, para. 4). Yet if the second stage is not autonomous, then neither is the Merge model overall, and Norris et al.’s claim to account for all speech recognition data with an autonomous model fails.

Faced with this objection, one strategy is to argue that the second, decision-making stage is autonomous after all. At times Norris et al. seem tempted by this strategy. Though not stated explicitly, this view is implied by their contrasting “merging,” “combining,” and “integrating” (sect. 5) lexical and prelexical information with interactive processing. For example: “It is not interaction that offers the benefit, but the process of combining two different kinds of information – the lexical and the prelexical” (sect. 8, para. 4).

How can merging or integration be considered a form of autonomy, rather than interaction? Their idea seems to be that the decisive criterion for distinguishing autonomous from interactive processes is *feedback*. That is, Norris et al. seem to consider a system directly influenced by higher-level information to still be autonomous as long as the intra-modular information flow involves no feedback loops. But this is a problematic move. Reinterpreting the interactive/autonomous distinction in terms of a feedback/non-feedback distinction artificially eclipses the theoretical space for other (i.e., non-feedback) kinds of interactive processing. Moreover, to automatically classify all non-feedback top-down processes as autonomous is contrary to normal usage. However, although Norris et al. gesture in this direction, I do not think it is (or should be) their considered view.

Instead, Norris et al. seem to endorse a second strategy: acknowledge that decision-making is not autonomous, but deny that this threatens the autonomy of speech recognition. Norris et al.’s crucial move in defense of this view is to deny that the second, decision-making stage is part of ordinary speech recognition.

One might worry that Norris et al. are simply stipulating that phonemic decision-making is outside speech recognition in order to justify their autonomy claim. It would certainly appear this way if phonemic decision-making were a necessary final stage of the overall phonemic task. But an extremely interesting and consequential claim of their target article is that it is not. In their view, the ambiguity-preserving output of the first phonemic stage is as complete as phonemic processing gets in *ordinary speech recognition*. The explicit phoneme decision tasks that subjects perform in the lab, they claim, are distinct from the kind of phonemic processing that goes on in ordinary speech recognition. Real-world speech recognition requires only the first stage; only explicit phoneme decision tasks (of the sort performed in laboratories) require both.

And sure enough, if second-stage phonemic decision-making is

not part of speech recognition, then it does not matter whether its merging processes are classified as autonomous or interactive. Either way, speech recognition remains autonomous, because first-stage phonemic processing – the only stage it requires – unproblematically is.

Nevertheless, this line of response creates its own complications. For one thing, even if speech recognition is autonomous, the Merge model is not, because whether or not phonemic decision-making is part of speech recognition, it is still part of Merge. The system that accounts for the data (i.e., Merge) is not autonomous; and the system which is autonomous (i.e., phonemic processing) does not account for the data. This leads to an even more consequential difficulty. It is no longer clear how the body of data that Merge is advertised as accounting for bears on the processes of speech recognition. For this body of data is derived from exactly the sorts of experimental phoneme decision tasks that Norris et al. claim are different from ordinary speech recognition.

Norris et al. are aware of this difficulty. They respond by claiming that from the data on explicit phoneme decision tasks we can infer what ordinary speech recognition is like, since the model that accounts for the explicit data (theirs) contains an autonomous speech recognition component. This is an interesting claim. Even if it is their considered view, however, their overarching claim that Merge is an autonomous model that accounts for all the data from speech recognition would still need to be qualified. Merge is not autonomous, even if one of its components is; and the data that Merge accounts for is not (in Norris et al.'s view) speech recognition data, even if it may shed indirect light on the process of speech recognition.

ACKNOWLEDGMENTS

I would like to thank Brian C. Smith for discussion and comments, and the Center for the Study of Language and Information, Stanford University, for supporting this research.

NOTE

1. Author is also affiliated with the Center for the Study of Language and Information, Stanford University, Stanford, CA 94305, irenea@csl.stanford.edu.