ISCA ITRW on
Spoken Word Access Processes (SWAP)
Nijmegen, The Netherlands
May 29-31, 2000

# ACTIVATION FLOW IN MODELS OF SPOKEN WORD RECOGNITION

**U. H. Frauenfelder**
Laboratoire de psycholinguistique expérimentale
Université de Genève, Switzerland
and **A. Content**
Laboratoire de psychologie expérimentale
Université libre de Bruxelles, Belgium

## ABSTRACT

Models of spoken word recognition in the localist connectionist tradition appeal to different types of activation flow to characterize the lexical selection process during which a competitor set is activated and the target word is identified in this set. This study investigates the nature of activation flow by means of experiments and simulations with the objective of (1) distinguishing between the effects of bottom-up and lateral inhibition and (2) te :ing the effects of top-down word to phoneme activation flow upon word recognition. Our results support the concept of bottom-up inhibition and raise doubts about the utility of top-down feedback.

## 1. INTRODUCTION

Current models generally agree that spoken word recognition involves the activation of a set of lexical competitors and the selection of the target word from this activated set. By all accounts the amount of bottom-up activation received by any given word unit —target or competitor— represents this unit's fit with the sensory input. This fit depends upon the quality and the quantity of the match between the sensory input and the lexical form representation. Thus, all the models assume that competitors which match and are aligned with the input are activated.

In contrast, current models differ in their view of the exact mechanisms underlying the selection process. Different selection mechanisms have been proposed. According to COHORT [1], lexical selection is achieved by means of bottom-up activation combined with bottom-up inhibition. When mismatching sensory information is received, it deactivates the inappropriate lexical units. According to TRACE [2], competitor set reduction is achieved through lateral inhibition. This inhibition between lexical competitors allows the stronger candidates, and in particular, the target, to dominate and deactivate the weaker ones. Moreover, this model also includes top-down word to phoneme feedback, which is assumed to help it recover from a minimal phonological mismatch in the input. Finally, the SHORTLIST model [3] includes both bottom-up inhibition and lateral inhibition, but no top-down feedback.

It is important to identify the cases in which these selection mechanisms lead to divergent empirical predictions. First, the tolerance of word initial mismatch in the sensory input by the models varies. It can range from extreme intolerance and total lexical deactivation by any mismatch in the sensory input via bottom-up inhibition (like in COHORT) to relative tolerance and partial activation despite some mismatch. In this latter case, there is presumably some function that relates the phonological distance between the input and a given competitor to the level of activation attained by that competitor. The closer and more complete the fit, the greater the activation received by the word.

A second difference is related to the effects of a phonological mismatch that appears later in the word after it has received a certain amount of activation. In models with bottom-up inhibition, the lexical candidate is strongly deactivated or even eliminated from the lexical competition. In contrast, in models that are solely based upon lateral inhibition, once a word has reached a certain level of activation and has no more lexical competitors, a phonological mismatch in the input will no longer influence its activation level.

In what follows, we first present two phoneme monitoring studies that examine how lexical candidates are activated and deactivated by partial phonological mismatches at onset and internal positions (see [4] for a similar study). Then we assess the effects of top-down feedback on the recognition of correct and distorted inputs in TRACE.

## 2. ON BOTTOM-UP INHIBITION

### 2.1. Experiment 1

French native speakers detected phonemes in words and in matched nonwords which were derived from these words by changing 1) the initial phoneme by a single distinctive feature (close nonword or CNW) or 2) multiple initial phonemes (baseline nonword or BNW).

| Activation | Baseline nonword | Word/nonword |
|---|---|---|
| Full | satobu**L**ary | vocabu**L**ary |
| Full | satobula**R**y | vocabula**R**y |
| Partial | satobu**L**ary | focabu**L**ary |
| Partial | satobula**R**y | focabula**R**y |

**Table 1**. Examples of the four comparisons used in Experiment 1 (target phonemes in boldface uppercase and mismatching information underlined)

Table 1 illustrates the four comparisons used. The difference in detection latencies to the phoneme in words and baseline nonwords represents full lexical activation, whereas the difference between the close nonword and the baseline nonword reflects the partial activation of the word by the close nonword. The position of the target

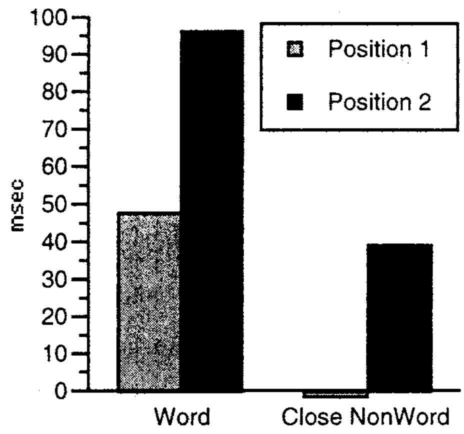


**Figure 1.** RT differences, relative to the Baseline Nonwords, for Word and Close Nonword comparisons for the two target positions.

was also manipulated, being either just after the uniqueness point or at word offset. This manipulation was included to determine whether lexical activation increases as the target arrives later and more bottom-up support has been received.

It is widely believed that TRACE tolerates minor initial mismatches so that "shigarette" will be recognized as "cigarette". However, the simulations presented in the next section (see also [5]) suggest that this is not totally correct since TRACE does not systematically recover from such mismatches. In contrast, SHORTLIST, thanks to its reset mechanism, appears to be more tolerant of such mismatches, allowing the recognition of words by mismatching stimuli when sufficient matching input has been received.

Figure 1 shows RT differences that correspond to the amount of lexical activation for the word (RT BNW-W) and close nonword (RT BNW-CNW) conditions for the two positions. Lexical activation was found for the word condition in both positions, with greater activation for the later position. In contrast, there was no lexical activation by the close nonword for position 1, but some activation emerged for position 2.

This experiment provides evidence for late lexical activation in the case of single-feature mismatches at word onset. Apparently, considerable bottom-up support for the word is required to override the inhibitory effect of the initial mismatching phoneme. This result appears to be more consistent with SHORTLIST than TRACE.

## 2.2. Experiment 2

The second phoneme monitoring experiment examines the deactivating influence of mismatching information arriving after the lexical candidates have already received some activation. In the first pair of comparisons (see Table 2), the target occurs before the mismatching information (Position 1), whereas in the second comparisons, the target occurs after the mismatching information (Position 2). The item internal mismatching phoneme differed from the original phoneme by several distinctive features.

| Activation | Baseline nonword | Word/nonword |
|---|---|---|
| Full | Satobu**L**ary | Vocabu**L**ary |
| Partial | Satobu**L**asy | Vocabu**L**asy |
| | | |
| Full | Satobula**R**y | Vocabula**R**y |
| Partial | Satobuna**R**y | Vocabuna**R**y |

**Table 2.** Examples of the four comparisons used in Experiment 2 (target phonemes in boldface uppercase and mismatching information underlined)

TRACE and SHORTLIST make divergent predictions concerning these comparisons. For SHORTLIST, the presence of mismatching information in the signal produces bottom-up inhibition that strongly deactivates the lexical candidate. In contrast, TRACE includes lateral but not bottom-up inhibition, and since in the present case the target word no longer has any lexical competitors at the moment that the mismatching information is received, no difference is predicted between activation curves for word and nonword conditions. The predictions of the two models are shown in Figure 2.

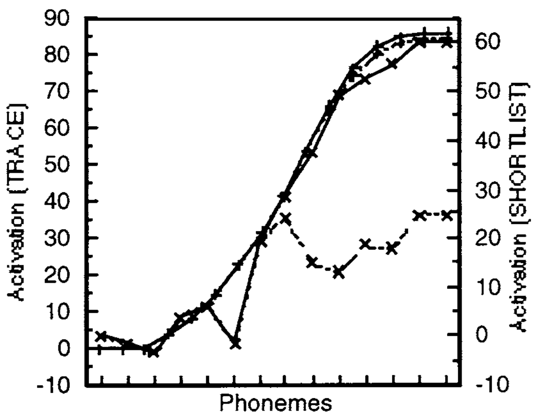


**Figure 2.** The activation curves for words (solid lines) and nonwords with medial mismatches (dashed lines) for TRACE (+) and SHORTLIST (x).

It is interesting to note that unlike the quantitative differences in models' predictions for experiment 1, the predictions here are qualitatively different and result from architectural choices on activation flow.

The results shown in Figure 3 reveal lexical activation for both word positions that, unlike in the first experiment, is approximately of the same amount. In the critical comparison in which the target followed the mismatch, no significant lexical activation is found. This suggests that the mismatching information completely deactivates the lexical candidate. In contrast, when the target preceded the mismatching information, significant but reduced lexical activation was found. This finding - that detection latencies are unaffected by the immediately following mismatch information - suggests that the phoneme monitoring task provides an immediate measure of lexical activation. With respect to the

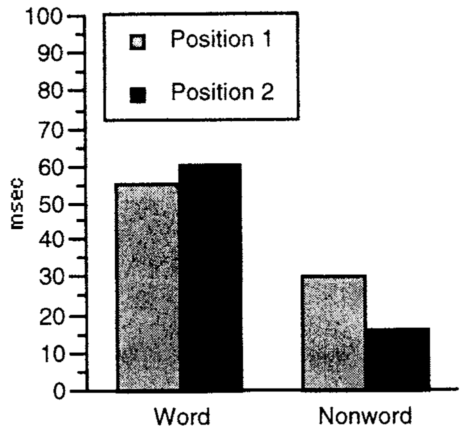


**Figure 3**. RT differences, relative to the Baseline Nonwords, for Word and Close Nonword comparisons for the two target positions.

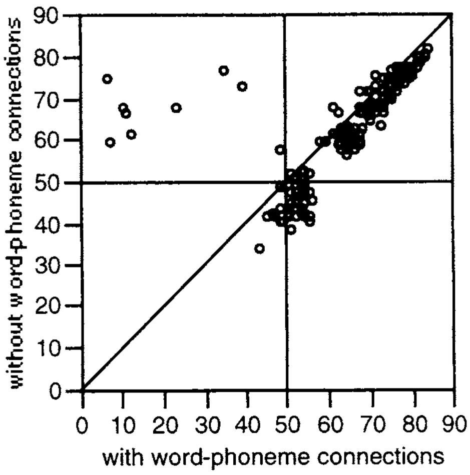


**Figure 4.** Maximum activation level attained by words with and without top-down feedback

prediction of the models described above, the present findings are inconsistent with the predictions of TRACE, but are in line with those of SHORTLIST.

## 3. ON TOP-DOWN FEEDBACK IN TRACE

A final important difference between the architectures of the models presented here is whether or not they include top-down feedback. To characterize the activation flow in lexical processing completely, it is necessary to determine whether top-down feedback exists, and if so, what effect it has upon word recognition. Much of the debate has centered on the existence of top-down feedback upon phoneme recognition [6, 7]. Unfortunately, little attention has been paid to the how such feedback might actually affect processing at the lexical level. It has been simply been assumed that any facilitation at the phoneme level would automatically translate into faster word recognition. It is important, therefore, to test this assumption more carefully with simulations.

### 3.1. Top-down feedback and words

In a series of TRACE simulations, we evaluated the effect of the presence or absence of top-down feedback on the recognition of words. In a first simulation we selected 400 words from the BIGLEX lexicon (with more than 1000 lexical entries) and submitted the words to a TRACE simulation. First, the maximum activation reached and the recognition points for every input word were calculated with the normal top-down parameter setting. Then in a second simulation, the same words were tested with the top-down parameter turned off (set to 0). Finally, the maximum activation level reached for the words in each condition was compared. The results of this comparison are plotted in Figure 4.

The figure shows that most words reach approximately the same level of activation with (94.5 %) and without (87.5 %) top-down feedback, with a slight advantage for the former condition. However, some words were considerably more activated in the absence of any top-down feedback. Thus, the results of this simulation do not confirm the predicted strong facilitation of word recognition by top-down lexical to phoneme feedback. According to TRACE, all the words in the simulation should have reached a higher activation level with top-down feedback than without.

### 3.2. Top-down feedback and input mismatch

Proponents of interactive models [2] claim that top-down feedback is especially important in the recognition of an ambiguous or distorted input. Accordingly, when the input is not decisive, the top-down information is assumed to compensate. However, this claim about facilitatory feedback effects with distorted inputs has never been evaluated fully at the lexical level.

We tested this claim with a further simulation similar to the preceding one. We first transformed the 400 word stimuli into nonwords by changing the initial phoneme by one distinctive feature. Then we conducted the same simulations as before with and without top-down feedback. The results are plotted in Figure 5.

Unlike what is commonly assumed, our simulation results showed that TRACE's recognition performance was globally quite poor for stimuli with small initial phonological mismatches, and this with or without top-down feedback. Indeed, if the activation level of 50 is taken arbitrarily to represent the recognition threshold, many words are not recognized in either condition.

Separate analyses showed that more mispronounced words reached their recognition threshold without top-down (42%) than with top-down (22%) feedback. In other words, when the parameter controlling the top-down feedback from word to phoneme levels was turned off, the recognition rate for the mismatching stimuli improved considerably. These results show a detrimental rather than beneficial effect of top-down feedback in word recognition.

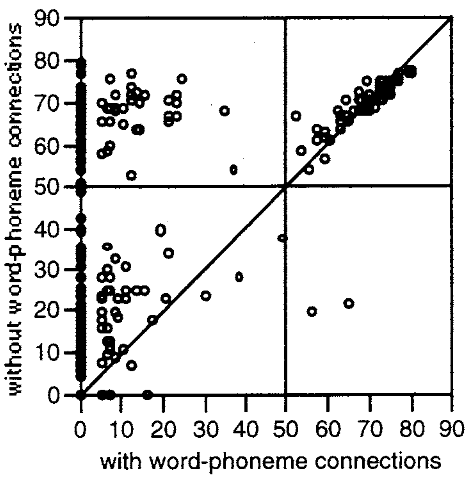


**Figure 5**. Maximum activation level attained by close nonwords with and without top-down feedback

## 4. CONCLUSION

One of the basic goals in the study of spoken word recognition is characterizing the nature of the information flow which allows lexical selection and recognition. We have contrasted two mechanisms of lexical selection, one based on bottom-up inhibition and the other on lateral inhibition. Our experimental results have shown the need to include bottom-up inhibition in any model: The phoneme monitoring results for medial mismatches show a strong effect of lexical deactivation that can only be accounted for in terms of a bottom-up inhibition mechanism. In sum, these results in favor of bottom-up inhibition, taken together with other findings in the literature that support lateral inhibition [8], suggest that the selection process is based on a combination of both - as in SHORTLIST.

Our experimental findings with initial mismatching stimuli showed that listeners could recover the intended word given sufficient bottom-up support for a word. In contrast, the corresponding TRACE simulations showed relatively poor recognition performance (about 22% correct recognition). Finally, the results of our evaluation of the effect of top-down feedback upon the recognition of such distorted words were problematic for interactive models since they showed a detrimental rather than beneficial effect of this feedback. Hence, these simulations raise serious doubts about the utility of such top-down feedback at the lexical level.

## 5. ACKNOWLEDGEMENTS


This research was supported by a grant from the Fonds National de la Recherche Suisse (FNRS 1113-049698.96). Thanks go to Mark Scholten and Jean-Philippe Goldman for their contribution to the experimental and simulation parts respectively.


Correspondence should be sent to Uli Frauenfelder, Laboratoire de Psycholinguistique Expérimentale, Université de Genève, e-mail: frauenfe@uni2a.unige.ch.


## 6. REFERENCES


[1] Marslen-Wilson, W.D. (1984). Function and process in spoken word recognition. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and Performance X: Control of Language Processes*, Hillsdale, NJ: Lawrence Erlbaum Associates.

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

[3] Norris, D. (1994). SHORTLIST: A connectionist model of continuous speech recognition. *Cognition*, **52**, 189-234.

[4] Connine, C.M., Blasko, D.M. & Titone, D.A. (1993). Do the beginning of words have a special status in auditory word recognition? *Journal of Memory and Language*, **32**, 193–210.

[5] Frauenfelder, U.H. & Peeters, G. (1998). Simulating the time-course of word recognition: An analysis of lexical competition in TRACE. In J. Grainger & A. Jacobs (Eds.). *Symbolic/Connectionist approaches to Human Cognition.* Cambridge: LEA.

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

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

[8] Norris, D., Mcqueen, J.M. & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **21**, 1209-1228.