# The Detection of Mispronunciations and the Influence of Context

I.B. Ottevanger
*Utrecht, the Netherlands*

## 1. Introduction

This report reviews a series of experiments investigating word recognition and the influence of context. The series was set up to test the cohort theory of word recognition (Marslen-Wilson and Welsh, 1978), a model which makes precise predictions about the way in which recognition takes place. In short it claims that in the perception of speech 'the word' is the level at which data-driven and knowledge-driven processing strategies are optimally co-operative. The model assumes that on the basis of acoustic information a word-initial 'cohort' is activated, which contains all words in a language that begin with the same two or three phonemes as the input word. Next, word candidates are removed from the cohort as soon as their acoustic characteristics are no longer compatible with the acoustics of the flow of new input; the same happens when word candidates are in conflict with contextual specifications. When one word candidate is left, word recognition has been achieved. Going from left to right in the word, the phoneme that distinguishes that word from all others in the cohort is called the recognition point.

## 2. Method

### *2.1. Stimuli*

Recognition points of twelve Dutch polysyllabic words were determined with the aid of a standard Dutch dictionary (Kruyskamp, 1976). Each word was mispronounced by changing one phoneme into another at four or five successive points, the 3rd point being the phoneme that functioned as the recognition point. Care was taken that the initial two phonemes of the stimulus words and the final one were not mispronounced, so that word boundaries were kept intact. Other requirements were that all mispronunciations were phonotactically legal and that the initial part of the words up to and including the misplaced phoneme was not identical with the beginning of any other Dutch word.

The stimulus words were spoken in isolation and in a final position in short auditory[1] context sentences. These sentences were alternative versions of the phrase 'The next word is ....' Out of the auditory context sentences the words

were spliced onto structural context sentences, again sentence-finally. The latter set consisted of five sentences which were ambiguously constraining, i.e., syntactically and semantically they led both to the target word and at the same time to another word that shared its first two phonemes with the stimulus word. The extent to which targets and alternatives turned out to be predictable on the basis of preceding context combined with the acoustics of the first phonemes was $\overline{X}$ 52% and 55%, respectively. The remaining seven sentences were uniquely constraining: syntactically and semantically they led to the target words only. On the basis of prior context and acoustic characteristics of the first phonemes their mean predictability was 92%.

### *2.2. Procedure*

For each of the three conditions, isolation, auditory context and structural context (ambiguous and unique), five groups of eight to ten subjects were instructed to listen for mispronunciations (cf. Cole, 1973) and to press a response key as soon as an error was heard. The four or five different mispronunciations of a target word were presented to the different groups of subjects. Reaction times (RTs) were measured from the onset of the mispronounced phoneme.

### *2.3. Predictions*

On the assumption that word recognition is prior to error detection, the cohort model predicts long RTs to mispronounced phonemes preceding the recognition point or coinciding with it, and short RTs to errors following the recognition point.

Because in the isolated condition recognition is based on the interaction of acoustic input and lexical knowledge only, the cohort theory predicts long RTs to mispronounced 1st, 2nd and 3rd points (the 3rd point being the recognition point), and short RTs to 4th and 5th points. The same prediction applies to the auditory context condition: since auditory context has no power to remove word candidates from the cohort, words are predicted to be recognized at the same point as when presented in isolation.

For the ambiguous structural context condition the model predicts long RTs to mispronounced 1st and 2nd points, since on the basis of acoustic input, lexical knowledge and contextual constraints two word candidates (the stimulus word and the alternative word) are left in the cohort, and therefore, word recognition has not yet taken place. Short RTs are predicted to 3rd, 4th and 5th points, because one of the two candidates has now been removed on account of its incompatibility with the acoustic input.

[1] The term is taken from Pollack and Pickett (1964); auditory context implies that the sentence-final word is not constrained by prior context in a syntactic and/or semantic manner, in the case of structural context prior context does have such constraints on the sentence-final word.

In the case of the seven uniquely constraining structural context sentences the cohort theory claims that word recognition has occurred on the basis of context well in advance of the earliest mispronunciation point and all RTs should be short.

## 3. Results

The results of the detection experiments are displayed in Table I and graphically represented in figure 1.

A one-way analysis of variance showed that for the isolated condition RTs to 4th and 5th points were significantly shorter than to 1st, 2nd and 3rd points as predicted (F(3,370)=2.90, $p < .05$); for the auditory and the ambiguous structural context condition there was no significance. For the unique structural context condition there was a highly significant difference between RTs to 1st and 2nd points on the one hand and 3rd, 4th and 5th points on the other (F(4,351)=4.69, $p < .01$); this was not in accordance with the prediction that RTs to the successive points would be equally short.

## 4. Discussion and conclusions

The recognition of words spoken and presented in isolation is adequately accounted for by the cohort model. The same is not true when words are presented in auditory and structural context. In auditory context the pattern which reflects the crucial role of the recognition point in word recognition, namely the large difference between the 3rd and the 4th mispronunciation point, has disappeared. The ambiguous structural context sentences have not achieved that recognition occurs at an earlier point in the word. The unique structural context results show that recognition has taken place at an earlier point, but not so early as the interaction of acoustic analysis, lexical knowledge and syntactic/semantic constraints permits.

For a more elaborate discussion of these results and for a presentation of the complete stimulus set of which these stimuli were a subset, the reader is referred to Ottevanger (1982; 1984).

*Table I.* Mean RT and standard error in ms per mispronunciation point for the three conditions

| | Isolated | Auditory context | Structural context | |
|---|---|---|---|---|
| | | | ambiguous | unique |
| 1st point | 833 (25) | 769 (31) | | 753 (75) |
| 2nd | 831 (18) | 717 (24) | 659 (50) | 613 (28) |
| 3rd | 790 (23) | 642 (26) | 644 (58) | 455 (28) |
| 4th | 601 (17) | 556 (21) | 660 (97) | 454 (24) |
| 5th | 576 (18) | 495 (18) | 567 (59) | 458 (19) |

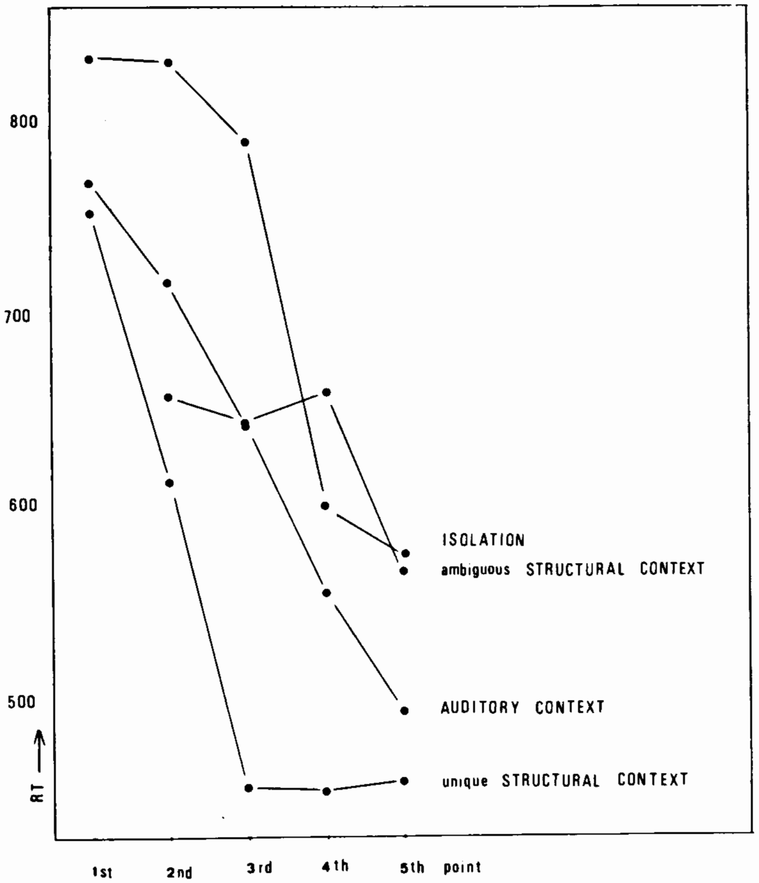

*Figure 1.* Mean RTs to the successive mispronunciation points in the target words as found in the isolated condition, in auditory context and in structural context (ambiguous and unique).

It can be concluded that, although reactions to mispronunciations in words presented in context are *faster*, context has no accelerating effect on word recognition in the sense that words are recognized *earlier*.

The finding that RTs are shorter to mispronunciations in words presented in auditory context compared to isolated words, fits in well with the results of Pollack and Pickett's (1964) experiment, in which they found that additional context contributed to the intelligibility of excerpts even though the contents were known to their subjects beforehand.

Finally, it is striking to see that, however long RTs to mispronunciations in isolation are, standard errors are small, indicating much conformity between subjects. The same degree of conformity is not found in the other conditions; the extremely high standard errors for ambiguous structural context indicate that subjects were very much hampered by the presence of alternative word candidates.

## Acknowledgement

This research was made possible by a grant of the researchpool of Utrecht University.

## References

Cole, R.A. (1973). Listening for mispronunciations: a measure of what we hear during speech. *Perception and Psychophysics* **11**: 153-156.

Kruyskamp, C. (1976). *Van Dale's groot woordenboek der Nederlandse taal*, 10th edition, Martinus Nijhoff, 's-Gravenhage.

Marslen-Wilson, W.D. and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* **10**: 29-63.

Ottevanger, I.B. (1982). Word recognition in non-constraining context in comparison with recognition of isolated words. *Progress Report of the Institute of Phonetics* (PRIPU) **7**,(2): 41-56.

Ottevanger, I.B. (1984). Word recognition in syntactically and semantically constraining context. To appear in *Progress Report of the Institute of Phonetics* (PRIPU) **9**,(1).

Pollack, I. and Pickett, J.M. (1964). Intelligibility of excerpts from fluent speech: auditory vs. structural context. *Journal of Verbal Learning and Verbal Behavior* **3**: 79-84.