Identification of Intervocalic Plosive Consonants: the Importance of Plosive Bursts vs. Vocalic Transitions

M.E.H. Schouten and L.C.W. Pols
Utrecht and Amsterdam, the Netherlands

1. Introduction

When plosive consonants occur between vowels, as they usually do in conversational speech, their information-bearing elements are silence, vocal murmur, the plosive burst, and the two vocalic transitions (VC and CV). In the literature there has been some controversy about the relative perceptual importance of burst, CV-transition, and VC-transition; for a short survey see Schouten and Pols (1983). It is clear that the plosive burst is the most invariant one of the three but it is also true that the vocalic transition is probably more resistant to masking, and may therefore be more important perceptually.

We have run a series of experiments in which subjects had to identify plosive consonants from signals in which various parts of those consonants had been deleted (replaced by silence). The aim was to find out what the relative contributions of the plosive burst and of the two kinds of vocalic transition are to the identification of plosive consonants. In this presentation we restrict ourselves to a subset of our data: we shall only discuss those conditions which all experiments had in common; a more comprehensive report will be published at a later date. There were four experiments:

1. Pairs of ambiguous sentences: each pair consisted of two nearly identical sentences, which differed only in one plosive consonant. Segments of this consonant were deleted in various ways, and the consonant had to be identified; subjects chose between the two possible alternatives.
2. In order to find out whether the sentence context, which, strictly speaking, was completely redundant, had any influence on the scores, we lifted the VCV sequences containing the relevant plosives from the sentences and presented them in isolation. Again, the two possible alternatives were available for subjects to choose from.
3. As a link to experiment 4, the stimuli of experiment 2 were presented to the same subjects, who this time were allowed to choose among the five plosive consonants Dutch has: /p, t, k, b, d/. This experiment was actually run before experiment 2.
4. In a re-run of the experiment described in Pols and Schouten (1982) and Schouten and Pols (1983), we asked subjects to identify the three plosive consonants in CVCVC nonsense sequences. There was, however, one difference: in the earlier experiment every transition was defined operationally as consisting of three pitch periods, but this time the transitions were defined as beginning (VC) and ending (CV) at the midpoints of the vowels. This made the segmentation procedure the same as the one employed in experiments 1, 2, and 3. As in experiment 3, subjects had five alternatives to choose from. In what follows, we shall only deal with the second of the three plosive consonants in the CVCVC utterance, since this was the only intervocalic one.

2. Method

In all four experiments the segment defined as the VC-transition preceding the burst started near the midpoint of the vowel and ended where the plosive burst (including vocal murmur or silent interval) began; everything was reversed for the CV-transition following the burst. The deleted segments were replaced by silence, and a 5 ms smoothing window was applied at the segmentation points. Four conditions are considered here: nothing deleted, only the burst left in, only VC left in, and only CV left in.

Experiment 1: ambiguous sentences

We constructed 50 pairs of sentences, each pair consisting of two sentences which were identical except for one plosive consonant (example: 'hij telde twee keer' = 'he counted twice' vs. 'hij belde twee keer' = 'he rang twice'). Each of the five Dutch plosive consonants occurred five times opposite each of the other four Dutch plosives in varying intervocalic contexts; since there were ten possible pairings, this resulted in 50 sentence pairs. The 100 sentences were read by one of the authors, who took great care to pronounce both members of each pair identically. The recordings were digitised (20 kHz, 12 bits) and regenerated in a random order of sentences and of segmentation conditions.

Subjects were 19 university students: half of them did this experiment before experiment 4 on the same day, and half of them did experiment 4 first. For each of the sentences they had to encircle one of two letters on a response sheet which gave the whole sentence. On the whole they enjoyed this task, which was experienced as quite a natural one.

Experiment 2: VCV segments from the sentences, 2 alternatives

The stimuli were the VCV segments containing the 'ambiguous' plosive consonants from experiment 1, from the middle of the preceding vowel to the middle of the following one. Either the whole segment was presented, or the burst, or one of the transitions.

Subjects were 24 university students: they took part in this experiment after experiment 3. They had to encircle the appropriate one of two letters on

a response sheet; this sheet did not contain any information on the sentences. The task was strongly disliked.

Experiment 3: VCV segments from the sentences, 5 alternatives

This experiment, which always took place before experiment 2, and in which the same 24 subjects participated, was identical to experiment 2, except that subjects now had the five Dutch plosive consonants to choose from.

Experiment 4: CVCVC sequences

45 Meaningless CVCVC sequences were read out by the other author, digitised (12 kHz, 10 bits) and regenerated in a random order of sequences per condition. Whatever was deleted from one of the consonants was deleted from all three, so that subjects were presented with three bursts, or two CV-transitions, or two VC-transitions, or the whole utterance (as in all the experiments reported here, there were more conditions, but those do not concern us now). However, in the CV-condition the third consonant in the CVCVC-utterance was represented by its burst; the same applied to the initial consonant in the VC-condition. This was done in order to maintain the impression that each stimulus consisted of three consonants. Stimuli were embedded in carrier phrases.

Experiment 4 was alternated with experiment 1, and the same 19 subjects participated. Since this was the only one of the four experiments to be run on-line (tapes were used in the other three), in this experiment responses were given by typing in the appropriate letters. Subjects found it difficult to remember the correct order of the three plosives, but they expressed no strong dislikes.

3. Results and Discussion

All results that concern us here are summarised in Fig. 1, which shows the percentages of correctly identified plosives in the four experiments in the form of histograms. Unvoiced and voiced plosives are shown separately. The hatched bar in the middle of each histogram represents the condition in which only the plosive burst was heard; the bars to its left and to its right represent the conditions with only VC-transitions and only CV-transitions, respectively. The top horizontal line reflects the identification score for the uncut utterance.

We should like to highlight three features from Fig. 1.:

1. Overall, the burst elicits many more correct identifications than do either of the two transitions. This suggests that the vocalic transitions contribute relatively little to the perception of plosive consonants. Transitions help, presumably, but in many cases the plosive burst seems to contain enough information by itself. The results from experiment 4 contradict this

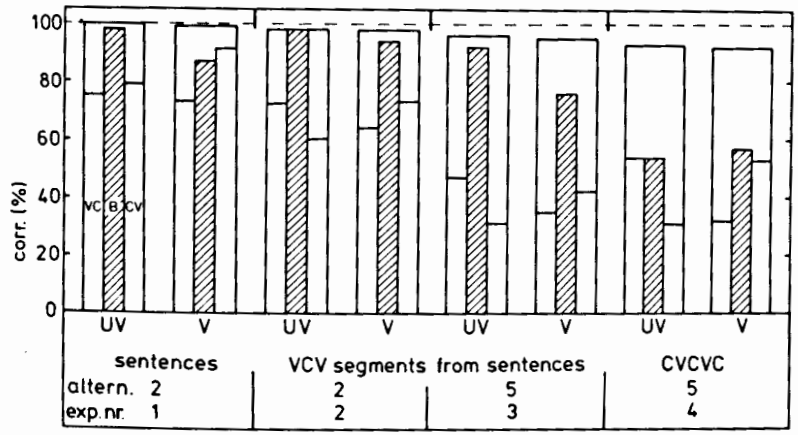

Figure 1: The results in terms of percentage of correct plosive identification responses in all four experiments. The hatched bar represents the score for the bursts without the vocalic transitions; the bars to its left and right represent the scores for the VC- and the CV- transition respectively. The top line in each histogram indicates the score for the whole (uncut) utterance. UV means unvoiced; V means voiced. Chance levels are 50% in experiments 1 and 2, and 20% in experiments 3 and 4.

conclusion to some extent. Since in experiment 4 the speaker was different from the one used in the other experiments, it could be that our main speaker just happens to be a good articulator of bursts. We would tend to support a different explanation, however: in experiment 4, subjects had to identify three plosives, which came in rapid succession. Hearing them as three vocalic transitions, one perceived a certain amount of continuity and tended to remember a (nonsense) word, whereas with three usually very brief plosive bursts the effect was one of three completely disjointed sounds suggesting a short burst of machine-gun fire. In such a situation low scores are almost inevitable, and the conclusion could be that vocalic transitions are indispensable under (normal) circumstances in which many consonants have to be identified in a short time: if transitions were absent, continuity could fall by the wayside.

2. Occasionally, the CV-transition rivals the burst in the percentage of correct identifications. This happens only in the case of voiced plosives, and then only in experiments 1 and 4. The same applies to the VC-transitions of unvoiced plosives in experiment 4. In experiment 4, these are probably the result of the low scores for the bursts, but in experiment 1 no such explanation is available. In fact, this is just one of the differences between the results from experiments 1 and 2. It should be remembered that stimuli and procedure were the same in experiments 1 and 2, except that in experiment 1 a supposedly completely redundant context surrounded the stimuli. Either the context was not as redundant as we believed it to be, or the 'naturalness' of the stimuli does have a strong influence on identification behaviour, turning CV-transitions into important cue-bearers in voiced plosives.

3. There appears to be a fairly systematic difference between VC- and CV-transitions: the latter are more important in voiced plosives, whereas the former are more important in unvoiced plosives. Contrary to English, there is virtually no voice-onset-time between the burst and the vocalic parts of Dutch plosives, so our finding here may be a language-specific one. We predict, however, that voice-onset-time cuts off a part of the CV-transition and thus reduces the overall score for CV-transitions, but should not affect the relative scores. For the time being, therefore, we claim that our finding should also hold for a language such as English.

We realise that forcing people to listen to utterances from which bits have been deleted does not tell us everything about the role of the remaining bits in normal connected speech. By introducing silence into an utterance one does not simply eliminate cues - one also adds new ones. We have attempted to avoid that trap by conducting an experiment with sentences in which one plosive was mutilated, and which the subjects regarded as quite natural: they hardly noticed anything unusual about the sentences. However, we have not managed to avoid the trap: in all experiments, including the sentence experiment, the gaps in the stimuli provoked a strong tendency towards /p/- and /b/- responses, a tendency which seemed to overrule subjects' written preference for one sentence rather than the other of a pair. These things will, however, have to wait for future evaluation.

References

L.C.W. Pols and M.E.H. Schouten (1982). Perceptual relevance of coarticulation. In: R. Carlson and B. Granström (eds.). *The representation of speech in the peripheral auditory system*. Amsterdam, Elsevier Biomedical Press, 203-208.

M.E.H. Schouten and L.C.W. Pols (1983). Perception of plosive consonants. The relative contributions of bursts and vocalic transitions. In: M.P.R. van den Broecke, V.J.J.P. van Heuven and W. Zonneveld (eds.), *Sound Structures*, Dordrecht, Foris.