UvA-DARE (Digital Academic Repository)

Plosive consonant identification in ambiguous sentences

Pols, L.C.W.; Schouten, M.E.

DOI

[10.1121/1.393099](https://doi.org/10.1121/1.393099)

Publication date

1985

Published in

The Journal of the Acoustical Society of America

[Link to publication](#)

Citation for published version (APA):

Pols, L. C. W., & Schouten, M. E. (1985). Plosive consonant identification in ambiguous sentences. *The Journal of the Acoustical Society of America*, 78(1), 33-39.
<https://doi.org/10.1121/1.393099>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Plosive consonant identification in ambiguous sentences

L. C. W. Pols^{a)} and M. E. H. Schouten^{b)}

Institute for Perception TNO, Soesterberg, The Netherlands

(Received 6 September 1983; accepted for publication 12 March 1985)

The hypothesis tested in this paper is that the acoustic-phonetic information at the sentence level (irrespective of the meaning of the sentence) makes it easier to interpret those cues to plosive consonant identity which are contained in vocalic transitions. Listeners were given a two-way choice in meaningful Dutch sentences which allowed two different intervocalic plosive consonants in the same position, and from which various parts of the VCV segments (plosive burst, vocalic transitions) had been deleted. The results were compared with results from a condition in which the VCV segments were presented without the sentence context. The hypothesis was confirmed for CV transitions, except for /k/, but disconfirmed for nearly all VC transitions. This could mean that in experiments involving the perception of isolated VCV or CVC utterances, whether spoken in isolation or excised from longer utterances, the contribution of the CV transition to plosive consonant perception may be underestimated.

PACS numbers: 43.71.Es

INTRODUCTION

Real-time speech perception is a process of continuous adjustment of decision criteria on the basis of the available information. This information is of many kinds: acoustic, phonetic, lexical, syntactic, and pragmatic. In this paper it is our aim to find out how strong the effect of the *acoustic/phonetic* context at the sentence level is on phoneme identification. Restricting ourselves to plosive consonants, we want to create an experimental situation in which the listener is confronted with a choice between two response alternatives in a sentence position in which both alternatives are possible. We then want to compare this situation with one in which the *direct cues* to the plosive consonants in question (burst and surrounding vocalic transitions) are excised partly or completely from the sentences and presented in isolation. Ideally, both in the sentence context and in the isolated plosive consonant presentation, both response alternatives should be equally likely, so that the meaning of the sentence context provides no cues to the identity of the consonant to be identified. Any difference in scores between the two situations could then be attributed to the general acoustic/phonetic information: the *indirect cues* contained in the sentence context, such as relative duration, dynamic spectral characteristics, speaker-dependent characteristics, and so on. In practice, however, this ideal is not quite attainable, for the following reasons:

(1) It is very difficult to construct a sufficiently large set of appropriate sentences which can take on more than one meaning, depending on which plosive consonant occurs in one position in a sentence. It is impossible, however, to make all response alternatives syntactically, lexically, and pragmatically equally probable. As a result, we have to determine and, if possible, correct our listeners' response preferences

before subjecting them to the spoken sentences.

(2) Since we are talking about the identification of phonemes inside and outside their appropriate contexts, the decision about where to place the phoneme boundaries around our plosive consonants is of great importance. Obviously, we want to aim for boundaries which are placed in such strategic positions that all the direct phoneme cues are to be found between them, so that the rest of the sentence can only contain the more general acoustic and phonetic information (assuming for the moment that lexical response preferences play no part). We opted for the hypothesis that all perceivable cues to plosive consonant identity are contained within the plosive burst and the vocalic transitions to and from the surrounding vowel(s). To be on the safe side, we placed our boundaries in the middle of the "steady state" portion(s) of the surrounding vowel(s); even this, however, is no cast-iron guarantee that all direct cues to plosive consonant identity have been included; to check on this, we also included an experimental condition in which listeners had to identify a plosive consonant on the basis of "the rest of the sentence," i.e., with both burst and transitions removed; if all direct cues have been removed, this should lead to responses at chance level.

(3) Any context is likely to be beneficial to the perception of isolated consonants: it alerts listeners and often makes their task more enjoyable. To check on this, it may be desirable to include conditions in which the isolated consonant is embedded in a standard carrier phrase, which is produced separately from the consonant, and to compare this with the situation in which the consonants are presented entirely by themselves. We could think of no good way, however, of embedding the isolated plosive bursts and vocalic transitions we used (see below) in a carrier phrase. Moreover, as we will indicate in the next paragraph, we were interested not so much in the overall lowering of the scores as a result of removing the acoustic-phonetic sentence context, but more in the differential effect this removal might have on the contribution of plosive burst and vocalic transitions to plosive identification.

^{a)} Present address: Institute of Phonetic Sciences, University of Amsterdam, The Netherlands.

^{b)} Present address: Institutes of English and Phonetics, University of Utrecht, The Netherlands.

It has been known since the fifties (see, e.g., Delattre *et al.*, 1955) that in synthetic speech vocalic transitions can cue the perception of the place of articulation of plosive consonants. This does not mean, of course, that in natural speech plosive consonants are cued primarily by vocalic transitions; it only means that the information is present and can be used if necessary. This was confirmed for "natural" speech (in so far as words produced in isolation may be regarded as natural) by, among others, Fischer-Jørgensen (1972), La Riviere *et al.* (1975), Fujimura *et al.* (1978), Streeter and Nigro (1979), Pols and Schouten (1978), Pols (1979), Sharf and Ohde (1981), Ohde and Sharf (1981), and Pols and Schouten (1981). What all these studies failed to indicate, however, was whether the information contained in the transition actually takes precedence over other cues, or whether the plosive burst is the more important cue-bearing element, as is suggested by Winitz *et al.* (1972), Cole and Scott (1974), Wolf (1978), and Blumstein and Stevens (1980). Two studies addressing this question with respect to nonsynthetic speech were conducted by Dorman *et al.* (1977) and Schouten and Pols (1983). The main difference between these two studies was the very low identification score for the bursts obtained by Dorman *et al.*, whereas Schouten and Pols found very high scores, in common with the studies on plosive bursts cited above. In both studies bursts or transitions were deleted from the same words; however, Dorman *et al.*, used voiced American English plosives (three alternatives), whereas Schouten and Pols used voiced and unvoiced (but unaspirated) Dutch plosives (five alternatives).

Using unvoiced (aspirated) American English plosives (three alternatives), Amerman and Parnell (1981) obtained a 91% correct score with isolated plosive bursts; this score increased to close to 100% when the carrier phrase preceding the burst was left intact. The main cause of the difference was probably the presence in the carrier phrase of a vocalic transition from the final vowel of the carrier to the plosive burst which, in our terminology, is a direct cue. Consequently, this experiment does not answer our questions with regard to the contribution to plosive perception of the indirect cues provided by the wider acoustic/phonetic context of the sentence: As we have already indicated, we regard the vocalic transition as an integral part of the plosive consonant. The same considerations probably apply to an experiment by Lynn and Brotman (1981), who also investigated the effects of presence or absence of a carrier phrase on plosive consonant identification. What we are after, however, is the possible effect of "the rest of the sentence," which, ideally, should contain neither syntactic, lexical, and pragmatic cues, nor *direct* phonetic or acoustic cues to the identity of the relevant plosive consonant. We do suppose, however, that the sentence context contains indirect acoustic/phonetic cues such as relative duration, dynamic spectral characteristics, speaker-specific characteristics, and so on. We suspect that absence of this neutral sentence context will have an especially strong effect whenever consonant recognition is cued mainly by transitional information, since vocalic transitions are much more variable than plosive bursts: they differ substantially from vowel to vowel. It stands to reason, therefore, that the perception of vocalic transitions may benefit from a

wider acoustic-phonetic context, which may provide additional stability in the form of, for instance, relative timing cues. In fact, the contribution of transitions to plosive consonant recognition in nonsynthetic speech may have been underestimated as a result of being presented outside of an appropriate acoustic-phonetic context.

We decided to conduct an experiment in which subjects had to identify intervocalic plosive consonants in naturally spoken sentences, with either the burst or one or both of the surrounding vocalic transitions removed. In all cases, subjects were given a choice of two response categories, both of which yielded a meaningful sentence. In order to determine how much information is contributed by the acoustic-phonetic sentence context, subjects were also asked to make the same binary decisions on the same plosive consonants in meaningless VCV sequences excised from the sentences.

I. METHODS

A. Stimuli

Fifty pairs of Dutch sentences were constructed, such that the two members of each pair were identical, except for one intervocalic plosive consonant (example: "hij telde twee keer" = "he counted twice" versus "hij belde twee keer" = "he rang twice"). Most of the plosive consonants occurred at the beginning of a stressed syllable and were preceded by an unstressed vowel, usually of a different word. Dutch has five plosives (/p,t,k,b,d/), so there were ten possible pairings; each pairing occurred five times in different vocalic contexts; because of the much larger choice of vowels it was practically impossible to balance the vocalic contexts over the plosive pairings. The resulting 100 sentences were read by one of the authors, who took great care to pronounce both members of each pair identically, apart from the two intervocalic plosives. The sentences were read in pairs, which enabled the speaker to concentrate on avoiding any differences between the members of a pair. Each pair was read out repeatedly, until the authors were satisfied that no apparent differences remained, except between the two plosive consonants.

The recording were digitized (20 kHz, 12 bits). Around each of the disambiguating plosive consonants four segmentation points were defined:

(1) the middle of the preceding vowel; since we had found in Schouten and Pols (1983) that presence or absence of the stationary part of a vowel made very little difference to the identification of a neighboring plosive consonant, this fairly roughly defined segmentation point is treated in the present paper as the beginning of the vocalic transition towards the plosive;

(2) the end of the vocalic part of the vowel preceding the plosive—and the beginning of the closure (silence or vocal murmur);

(3) the end of the plosive burst—and the beginning of the vocalic transition towards the vowel following the plosive (Dutch has virtually no plosive aspiration);

(4) the middle of the following vowel, which in this paper is treated as the end of the vocalic transition.

Figure 1 shows waveforms and spectrograms for the relevant

FIG. 1. Waveforms and spectrograms of two members of one pair of sentences. Only the immediate environment of the disambiguating consonant is shown. Vertical dashed lines indicate segmentation points.

parts of the two members of a sentence pair; the four segmentation points are indicated by means of vertical dashed lines.

Segmentation was done by visual inspection of the relevant parts of the waveform; any chosen segment could also be made audible. In order to minimize transients, all segmentation points were chosen as close as possible to the zero line and smoothed by means of a 4-ms cosine window. One author carried out the segmentations, the other author checked them independently.

The four segmentation points were used to produce six different segmentation conditions, each symbolized by plus or minus signs at three positions, the first one representing presence or absence of the transition toward the burst (VC transition), the second one that of the plosive burst, and the third one that of the transition after the burst (CV transition):

- (1) no deletion (+ + +);
- (2) deletion of both vocalic transitions (− + −);
- (3) deletion of the plosive burst (+ − +);
- (4) deletion of burst and VC transition (− − +);
- (5) deletion of burst and CV transition (+ − −);
- (6) deletion of burst and both transitions (− − −).

The deleted segments were replaced by silence.

Although we did as much as we could to make the two members of a sentence pair acoustically and phonetically neutral with respect to the response alternatives, and although both response alternatives were possible and the sentences had been chosen carefully, this did not of course guarantee that both alternatives were equally likely. To obtain a measure of response bias, we asked our subjects before the listening experiments to indicate their preference by encircling one of the two alternatives presented on a typewritten list of the 50 sentences: they were told not to think about it but to go by instinct and to finish the list within 3 min. The result is shown in Table I, in terms of response percentages; the 23 subjects gave a response for each of the five different sentences belonging to a particular plosive consonant opposition. It is clear from Table I that our stimuli are not as balanced as we would have liked them to be: several pairs are rather far from a 50-50 split. We will see that quite different preferences emerge with an acoustic presentation.

For the *sentence condition*, the 600 stimuli (6 conditions $\times$ 100 sentences) were recorded on three tapes; each tape consisted of a complete randomization of the six condi-

TABLE I. Percentage of preference for the first consonant in the sentence pairs.

Consonant pair	Sentences					Average preference
	1	2	3	4	5	
p-t	48	35	30	52	52	44
p-k	48	43	70	74	87	65
p-b	91	35	22	56	61	54
p-d	83	48	43	78	43	60
t-k	30	17	48	74	9	37
t-b	56	52	87	30	91	65
t-d	52	91	9	83	61	60
k-b	74	52	48	17	65	52
k-d	4	22	35	74	39	36
b-d	13	48	17	83	56	44

tions for subsets of 34 or 32 sentences. The division into three subsets had to do with the limited storage capacity of our computer disks. Each tape took about 20 min, and between the tapes there was a short break.

In the *isolated VCV condition*, the stimuli were the same as those in the sentence context, except that the rest of the sentences was now completely left out. Only bursts and vocalic transitions were presented, but it should be remembered that the transitions started or ended in the middle of a vowel, so that they included part of the stationary vowel as well. All segmentation edges were smoothed by means of a 4-ms cosine window.

B. Subjects

Subjects were 43 university students who had no reported hearing loss. They were paid for their services. Twenty-three subjects were assigned to the sentence condition, and 20 to the isolated VCV condition.

C. Procedure

Up to four subjects at a time listened to the stimuli over headphones. In the sentence condition they responded by putting a circle around the preferred letter on a sheet showing the complete sentence. In each sentence the two possible response alternatives were given: one in the line space above and one in the line space below the rest of the sentence. These positions were randomized in such a way that half the time the upper alternative, and half the time the lower alternative was the correct one; moreover, in each sentence pair, each one of the two plosives was the upper alternative in half of the conditions, and the lower one in the other half. Since it was such a simple task, subjects were not given any practice. They were told in advance that bits of particular consonants had been tampered with, but except in condition (6) (− − −), where nothing of the consonant remained, they noticed hardly any gaps.

In the isolated VCV condition, a short practice session was included. A stimulus was generated every 2 s and recorded on tape. Four subjects at a time listened to the tape over headphones and recorded their responses on a sheet. This time, no printed context was given; instead there were ten pairs of letters on each line of the response sheet. After every ten stimuli a brief tone was recorded, in order to pre-

vent subjects losing track. An important difference from the sentences condition was that conditions (4) (- - +) and (5) (+ - -) were not mixed with the other conditions, but came in separate random orders at the end of the sessions. This was done because in these conditions only a single vocalic transition was present, and subjects had to be told to try to identify a plosive consonant at the beginning [condition (4)] or at the end [condition (5)] of the stimulus. An additional difficulty in condition (5) (+ - -) was that 40% of the time a voiced plosive had to be identified at the end of a transition, although in Dutch voiced plosives do not occur in word-final position.

II. RESULTS

The results are presented in Table II. The left-most column shows the two consonants forming a particular binary opposition, followed by two columns for each of the conditions (1)–(5), showing the correct response percentages for the sentences (S) and the isolated plosive consonants (I). Unequal percentages for the two members of a pair in one condition are due to response preferences on the part of the sub-

TABLE II. Percent correct scores per consonant pair in the sentence context (S) and in isolation (I).

cond.:	(1) + + +	(2) - + -	(3) + - +	(4) - - +	(5) + - -
	S I	S I	S I	S I	S I
p	99	100	98	100	100
t	100	100	100	94	77
				36	51
				26	31
				98	31
p	100	100	98	100	96
k	100	100	99	99	82
					48
					56
					85
					98
					76
					79
p	100	95	97	92	87
b	99	98	54	85	89
					82
					94
					75
					86
					64
p	100	100	99	99	100
d	98	96	77	86	98
					84
					96
					84
					98
					80
t	99	93	99	100	77
k	100	99	96	98	97
					55
					63
					40
					58
					39
t	100	100	100	100	82
b	100	100	97	99	99
					41
					73
					45
					47
					41
t	99	91	95	99	62
d	98	100	78	88	96
					16
					68
					37
					58
					58
k	100	100	98	99	88
b	100	100	98	100	99
					66
					74
					53
					74
					95
k	100	100	99	100	93
d	98	100	99	100	96
					79
					84
					52
					88
					90
b	98	99	93	97	100
d	100	97	98	99	97
					89
					99
					83
					90
					76
					65

jects. These response preferences may have many causes: they may be due to, for example, a lexical preference (cf. Table I), or to new, unintended, cues introduced by the excision of part of the signal. The term “new cues” implies that we are dealing with what may be genuine perceptual effects, and not merely a response bias. The effect is that of a response bias, however, and it may be useful to transform the data in such a way that the effects of the “new cues” are neutralized. As we shall see below, the “bias” is one which favors some responses (bilabials) at the expense of other responses; a successful neutralization procedure should therefore reduce the number of bilabial responses, but keep the other internal relations within the data intact. Such a procedure is the symmetrization procedure devised by Wagenaar (1968), intended to turn asymmetrical 2×2 confusion matrices into symmetrical ones [Wagenaar (1968) used Luce’s choice axiom, as Clarke (1957) had done in order to turn $n \times n$ matrices into 3×3 matrices]. The symmetrization procedure assumes that each of the response percentages consists of a “real” response term and a “bias” response term in the following way:

$$\begin{pmatrix} A & B \\ B & A \end{pmatrix} \times \begin{pmatrix} 1 + \lambda & -\lambda \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} A + A\lambda & B - A\lambda \\ B + B\lambda & A - B\lambda \end{pmatrix},$$

in which A and B are the “real,” underlying response percentages, and λ is the “bias” factor.

Another representation of the data is given in Fig. 2. The response percentages in Fig. 2 were obtained by averaging the percentages of correct responses for each of the four times a particular consonant occurred in a binary opposition in a particular condition. As a result, Fig. 2 shows the correct percentages per consonant both before (top) and after (bottom) symmetrization of the 2×2 confusion matrices. Within each panel of Fig. 2 solid lines connect the conditions for isolated VCV segments; the scores for the sentences are con-

FIG. 2. Correct identification percentages for each of the five plosive consonants, based on the original data (upper half) and on symmetrized confusion matrices (lower half). The five segmentation conditions are indicated by number in the middle of the figure, and by plus and minus signs at the bottom. The solid lines in the panels connect data points in the VCV context; the dashed lines connect the sentence context. The significance of the differences between the two contexts is indicated in the upper panels; ns means nonsignificant, * means significant at the 5% level, and ** means significant at the 1% level (Newman-Keuls *post-hoc* analysis).

nected by means of dashed lines. It is clear that the symmetrization has done what was expected of it: Virtually all relationships within the data have been preserved, except the differences between the consonants which are due to cues favoring bilabial responses introduced by the excision procedure.

The data underlying the top half of Fig. 2 (untransformed percentages of correct responses) were submitted to an analysis of variance, with CONTEXTS (sentences or segments) as a between-subjects factor and CONDITIONS (5) and CONSONANTS (5) as within-subject factors. Since our analysis program does not allow unequal numbers of subjects, three randomly chosen subjects of the 23 in the sentence context condition were dropped from the analysis. No analysis of variance could be done on the "symmetrical" data shown in the lower half of Fig. 2, since this procedure does not yield individual data; consequently, all significant effects are given with the untransformed data.

TABLE III. Correct response percentages in condition (6) (— — —) of the sentence context, presented separately for each of the sentence pairs. Horizontally and vertically, the order of the sentences is the same as that in Table I. In the second consonant pair data for the fifth sentence pair are missing because something had gone wrong in the recording.

Consonant pair	1	2	Sentence pairs		5	Average
			3	4		
p	100	91	83	96	91	92
t	0	13	91	22	13	28
p	70	96	100	100		91
k	35	4	39	100		45
p	100	96	91	87	74	90
b	4	26	22	26	35	23
p	96	100	96	91	100	97
d	17	4	26	9	13	14
t	39	48	13	39	9	34
k	70	52	96	78	91	87
t	22	4	43	83	17	29
b	100	96	61	35	96	87
t	78	65	70	61	56	66
d	43	52	74	78	43	58
k	30	26	61	48	61	45
b	96	100	87	74	91	90
k	100	78	65	96	70	71
d	70	52	65	48	83	60
b	83	61	83	100	96	84
d	61	52	48	30	13	44

All three main effects {CONTEXT [$F(1,38) = 103$], CONDITIONS [$F(4,152) = 532$], and CONSONANTS [$F(4,152) = 99$]} were highly significant ($p < 0.001$). A *post-hoc* Newman-Keuls test was used to determine the significance of the differences within the pairs of data points belonging to the same condition, but to different contexts. The significance level is indicated in Fig. 2 below these pairs; in this way the effect of context (sentence or VCV) can be read off at a glance. In addition to the Newman-Keuls test, we also performed a double check on the differences by means of separate analyses of variance for each pair of data points. The same picture emerged from both procedures, with only a few very marginal differences.

Assuming that the symmetrization procedure described above does remove all or most of the response preferences, we can now see in Fig. 2 which of the significant differences between the sentence and the VCV context are due to these preferences, and which are not. It is clearly visible that the difference between the two contexts is hardly affected by response preferences in the case of the voiced plosives /b/ and /d/: the "nonsymmetrical" and "symmetrical" data points for these consonants reveal very similar patterns. With the unvoiced plosives, however, especially condition (5) (+ — —) is unreliable for drawing comparisons between sentences and VCV segments.

The results for condition (6) (— — —) (which only occurred in the sentences context) are shown in Table III, which has largely the same organization as Table I, except that in Table III two percentages are given per cell, since here the two do not add up to 100%. Tables I and III allow one to compare the genuine response bias (preference for a particular meaning of the sentence) with the response tendencies present when all consonant information has supposedly been removed from the auditory signal. It was expected that there would be no great differences between responses to printed stimuli and to printed-cum-auditory stimuli, if neither contained any information about the response to be given, unless the preference for one phoneme or another in a particular sentence conflicted strongly with the auditory results of cutting away parts of the speech signal. As can be seen in Tables I and III, the differences are considerable.

III. DISCUSSION

As we had expected on the basis of Schouten and Pols (1983), and counter to Dorman *et al.* (1977), deleting the vocalic transitions while leaving in the plosive burst [condition (2), — + —] leads to only a very small deterioration of the scores [see Fig. 2, condition (2)]. A *post-hoc* Newman-Keuls analysis showed that this deterioration is only significant [$p < 0.01$, difference between conditions (1) and (2)] for /b/ and /d/ in the sentences. The explanation is probably that the relatively weak plosive burst of voiced consonants is partly masked by the preceding part of the sentence, whereas no such masking is present with voiced plosive bursts presented in isolation. The sentence context does not make much difference to the score: apart from a marginal difference ($p < 0.05$) involving /t/, the only significant difference between sentence contexts and VCV segments in condition

(2) (- + -) is the one involving /b/, where the reverse is true. Thus, plosive bursts did not benefit from appearing in the sentence context; context actually seemed to have a slight masking effect on the burst.

With respect to the vocalic transitions, our hypothesis was that they would be perceived better in an appropriate sentence context than in isolated VCV segments. Indeed, transitional information did lead to higher scores in the sentences than in the VCV stimuli. Although this agrees with our hypothesis, we have to be careful about accepting it at face value, since at least two objections could be raised: (1) The sentence context may have retained some cues about the identity of the deleted plosive consonant, despite segmentation in the middle of the vowel. (2) Cutting away parts of utterances (bursts or transitions) not only destroys cues, but actually introduces new ones: It is quite possible that the formant configuration at the cutting point is appropriate for a different (bilabial) consonant. However, this seems not to affect our arguments based on differences in correct scores between sentences and VCV stimuli.

Table III appears to lend support to the idea that the sentence context did retain cues to the identity of the deleted plosive consonant: condition (6) consisted only of the supposedly neutral sentence context, but the correct response percentages are above the chance level of 50%: the average over all sentence pairs is 61%. There was a strong response preference for /p/, but correcting for this imbalance does not materially affect the overall correct score of a little over 60%: it only increases from 61% to 63%. Since we do not believe that subjects were able to make use of any unintended differences between the two members of a sentence pair, we can only conclude that even cutting in the middle of a vowel, as we have done, does not remove all the coarticulatory information about the preceding or the following plosive consonant (see the spectrograms in Fig. 1). The question clearly is: to what extent has coarticulatory information at the sentence level contributed to the differences in the scores between the VCV and sentence contexts in conditions (3), (4), and (5)?

The answer is to be found in Table IV. The idea is that, given certain assumptions, it should be possible to predict the sentence scores from the VCV scores (first column) and from the nonrandom scores for condition (6) (second column).

The following steps were taken to obtain the predicted sentence scores:

The original correct score percentages were first corrected for response preference by symmetrizing the 2×2 confusion matrices, in order to remove the large differences in preference strength between sentences and VCV segments.

The "symmetrized" scores were then corrected for chance by means of the formula $Q = 2Q' - 100$ in which Q' is the experimental score, and Q is the corrected score. This stretched the original 50%–100% range to a necessary 0%–100% range.

From these corrected scores the error scores were derived by subtraction from 100%.

If we assume independent sources of information, we

TABLE IV. "Sentence gain." The score percentages for the separate VCV segment and the corresponding scores for condition (6) (sentences, - - -) are combined to give predicted values and are then subtracted from the actual sentence score, yielding the "sentence gain" in the right-most column.

	VCV	Cond. (6)	Predicted	Obtained	Gain
Condition (3) + - +					
p	43.4	12.4	50.4	75.8	25.4
t	40.6	29.4	58.0	85.7	27.6
k	63.6	39.3	77.9	87.4	9.5
b	73.8	31.5	82.0	92.0	9.9
d	52.1	18.5	60.9	93.0	32.0
Condition (4) - - +					
p	20.0	12.4	29.9	48.0	18.1
t	17.3	29.4	41.6	71.6	30.0
k	30.9	39.3	58.0	53.2	- 4.9
b	40.4	31.5	59.1	80.4	21.2
d	36.1	18.5	47.9	84.2	36.2
Condition (5) + - -					
p	19.5	12.4	29.2	27.9	- 1.3
t	12.1	29.4	37.9	55.1	17.1
k	42.2	39.3	64.9	68.8	3.8
b	42.1	31.5	60.3	55.4	- 4.9
d	29.5	18.5	42.6	43.7	1.1

can now calculate the predicted sentence score P from the VCV error score E_{vcv} and the condition (6) error score E_6 , in the following way:

$$P = 100 - (E_{vcv} \times E_6) / 100\%$$

The gain is then equal to the actual score minus the predicted score. The various scores for conditions (3), (4), and (5) are given in Table IV, per consonant. Turning first to condition (3) (+ - +), we see that the sentence gain for /p/, after symmetrization and correction for chance, is surprisingly great at 25.4%. The gains for /t/ and /d/ are also considerable, but those for /k/ and /b/ seem rather small by comparison. In condition (4) (- - +), four of the five consonants show a sentence gain; however, /k/ has a small negative sentence gain. In condition (5) (+ - -), finally, only /t/ seems to benefit from presentation in a sentence.

Although we realize that the above description is only a qualitative one, we feel quite confident about concluding that our hypothesis (that indirect cues in the sentence context improve plosive identification) is confirmed for conditions (3) (+ - +) and (4) (- - +), but not for condition (5) (+ - -), except in the case of /k/. It is much easier to correctly identify a CV transition [condition (4), - - +] presented in its sentence context than it is to identify the same transition presented in isolation. This is also true when both transitions [condition (3), + - +] are present, but does not apply to VC transitions by themselves [condition (5), + - -], except probably in the case of /t/. In an earlier investigation (Schouten and Pols, 1979a,b) we had found that in Dutch VC transitions there is very little consonant-specific information; it is therefore not surprising that the

sentence gain is negligible for VC transitions. In the present experiments, moreover, the VC transition was usually relatively weak, and frequently involved the unstressed /ə/.

IV. CONCLUSIONS

The hypothesis that information about plosive consonant identity contained in the *vocalic transitions* is more accessible when the transitions are presented in a sentence context than when these transitions are presented in isolation, was confirmed for the CV transitions except for /k/. It had to be rejected for the VC transitions, with the possible exception of /t/.

Plosive burst identification, on the other hand, did not benefit greatly from in-context presentation. The only significant difference between isolated presentation and in-context presentation pointed in the other direction; this applied to /b/ and could be interpreted as an effect of forward masking.

These conclusions point to a rather complicated state of affairs: generally speaking, one type of transition (CV) suffers from removal of the sentence context, but the other type (VC) of transition does not; neither does the burst. We think this means that experiments involving the perception of isolated VCV or CVC utterances, whether spoken in isolation or excised from longer utterances, will tend to underestimate the contribution of the CV transition to plosive consonant perception.

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor for going far beyond the call of editorial duty in helping us get this paper into its final shape.

- Amerman, J. D., and Parnell, M. M. (1981). "Influence of context and rate of speech on stop-consonant recognition," *J. Phon.* 9, 323-332.
- Blumstein, S. E., and Stevens, K. N. (1980). "Perceptual invariance and onset spectra for stop consonants in different vowel environments," *J. Acoust. Soc. Am.* 67, 648-662.
- Clarke, F. R. (1957). "Constant-ratio rule for confusion matrices in speech communication," *J. Acoust. Soc. Am.* 29, 715-720.
- Cole, R. A., and Scott, B. (1974). "The phantom in the phoneme: invariant cues for stop consonants," *Percept. Psychophys.* 15, 101-107.
- Delattre, P. C., Liberman, A. M., and Cooper, F. S. (1955). "Acoustic loci and transitional cues for consonants," *J. Acoust. Soc. Am.* 27, 769-773.
- Dorman, M. F., Studdert-Kennedy, M., and Raphael, L. J. (1977). "Stop-consonant recognition. Release bursts and formant transitions as functionally equivalent, context-dependent cues," *Percept. Psychophys.* 22, 109-122.
- Fischer-Jørgensen, E. (1972). "Tape cutting experiments with Danish stop consonants in initial position," *Ann. Rep. Inst. Phon. Univ. Copenhagen*, 104-168.
- Fujimura, O., Macchi, M. J., and Streeter, L. A. (1978). "Perception of consonants with conflicting transitional cues: A cross-linguistic study," *Lang. Speech* 21, 337-346.
- LaRiviere, C., Winitz, H., and Herriman, E. (1975). "Vocalic transitions in the perception of voiceless initial stops," *J. Acoust. Soc. Am.* 57, 470-475.
- Lynn, J. M., and Brotman, S. R. (1981). "Perceptual significance of the CID W-22 carrier phrase," *Ear Hear.* 2, 95-100.
- Ohde, R. N., and Sharf, D. J. (1981). "Stop identification from vocalic transition plus vowel segments of CV and VC syllables: A follow-up study," *J. Acoust. Soc. Am.* 69, 297-300.
- Pols, L. C. W. (1979). "Coarticulation and the identification of initial and final plosives," in *ASA *50 Speech Communication Papers*, edited by J. Wolff and D. Klatt (Am. Inst. of Physics, New York), pp. 459-462.
- Pols, L. C. W., and Schouten, M. E. H. (1978). "Identification of deleted consonants," *J. Acoust. Soc. Am.* 64, 1333-1337.
- Pols, L. C. W., and Schouten, M. E. H. (1981). "Identification of deleted plosives: The effect of adding noise or applying a time window. A reply to Ohde and Sharf," *J. Acoust. Soc. Am.* 69, 301-303.
- Schouten, M. E. H., and Pols, L. C. W. (1979a). "Vowel segments in consonantal contexts: A spectral study of coarticulation—Part I," *J. Phon.* 7, 1-23.
- Schouten, M. E. H., and Pols, L. C. W. (1979b). "CV- and VC-transitions: A spectral study of coarticulation—Part II," *J. Phon.* 7, 205-224.
- Schouten, M. E. H., and Pols, L. C. W. (1983). "Perception of plosive consonants—The relative contributions of bursts and vocalic transitions," in *Sound Structures: Studies for Antonie Cohen*, edited by M.P.R. van den Broecke, V. J. J. P. van Heuven, and W. Zonneveld (Foris, Dordrecht, The Netherlands), pp. 227-243.
- Sharf, D. J., and Ohde, R. N. (1981). "Psychological, acoustic, and perceptual aspects of coarticulation: Implications for the remediation of articulatory disorders," in *Speech and Language: Advances in Basic Research and Practice* (Academic, New York), Vol. 5, pp. 153-247.
- Streeter, L. A., and Nigro, G. N. (1979). "The role of medial consonant transitions in word perception," *J. Acoust. Soc. Am.* 65, 1533-1541.
- Wagenaar, W. A. (1968). "Application of Luce's choice axiom to form-discrimination," *Ned. Tijdschr. Psychol.* 232, 96-108.
- Winitz, H., Scheib, M. E., and Reeds, J. E. (1972). "Identification of stops and vowels from the burst portion of /p,t,k/ isolated from conversational speech," *J. Acoust. Soc. Am.* 51, 1309-1317.
- Wolf, C. G. (1978). "Perceptual invariance for stop consonants in different positions," *Percept. Psychophys.* 24, 315-326.