UvA-DARE (Digital Academic Repository)

Identification of deleted consonants

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

DOI

[10.1121/1.382100](https://doi.org/10.1121/1.382100)

Publication date

1978

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. H. (1978). Identification of deleted consonants. *The Journal of the Acoustical Society of America*, 64, 1333-1337. <https://doi.org/10.1121/1.382100>

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.

Identification of deleted consonants

L. C. W. Pols and M. E. H. Schouten

Institute for Perception TNO, Soesterberg, The Netherlands
(Received 17 March 1978; revised 2 August 1978)

VC transitions have often been found to be better cues for the identification of deleted consonants than CV transitions. The lower score for CV transitions could very well be due to an interfering "click sensation" caused by the abrupt beginning of these transitions. In the present investigation the abrupt onset was eliminated by replacing the deleted portions of the consonants with noise bursts; this also caused the incomplete syllables to sound more natural. The results show that the identification of deleted initial voiceless plosives is greatly improved by the addition of noise. The original difference between initial and final plosive transitions disappears almost completely.

PACS numbers: 43.70.Dn, 43.70.Ve

INTRODUCTION

A well-known method of measuring the perceptual effects of coarticulation is to delete the consonant burst and varying portions of the vocalic transition from an utterance, and to see how well the deleted consonant can still be identified. The utterance is usually a consonant-vowel (CV) or a vowel-consonant (VC) combination, and the listener is presented with only the vowel plus a part of the vocalic transitions from or to the consonant.

In some studies, such as Kuehn and Moll (1972), Fisher-Jørgensen (1972), LaRiviere, Winitz, and Herri-man (1975), and Dorman, Studdert-Kennedy, and Raphael (1977), only initial consonant identification, after removal of the release burst, was studied. However, whenever final consonants were included, it was found that identification of final consonants without release bursts was considerably better than that of similarly treated initial consonants: Sharf and Hemeyer (1972), Sharf and Ostreicher (1973), Sharf and Beiter (1974), and Ostreicher and Sharf (1976). Although a few hypotheses were put forward, no satisfactory reason was given why this should be so. Our own *spectral* measurements (Schouten and Pols, 1978a and 1978b) pointed the other way: We found that consonant-vowel transitions were more consonant specific than vowel-consonant transitions, and should therefore be expected to provide better cues for consonant identification. (Briefly, most of the consonant-vowel transitions studied pointed backwards to a fairly well-defined consonant locus positions in the principal-components vowel space, whereas the vowel-consonant transitions trailed off in the same general direction, regardless of the consonant involved.)

A possible explanation for the discrepancy between our spectral measurements and the perceptual results obtained by Sharf and his collaborators could reside in the differences between Dutch and English: Dutch voiceless stops have, for example, zero voice onset time (see Slis and Cohen, 1969). We decided to investigate this discrepancy by presenting listeners with Dutch syllables from which initial or final stop consonants had been deleted (details of the deletion process are described below).

In a pilot test we found for Dutch, too, that final plosives from which portions had been removed were considerably better identified than initial ones. We could

not help noticing however, that in spite of our strenuous efforts to avoid clicks, the abrupt start of the vowel often produced what we called a "click sensation"—the inevitable spectral consequence of short rise times. We felt uncertain about the extent to which this might affect the perception of the mutilated stimuli: It might mask the onsets or offsets of the stimuli, or it might even cause a new plosive consonant to be perceived.

Meanwhile, a new article by Ohde and Sharf (1977) showed us the way to our new experiment: As can be seen in Table I, Ohde and Sharf found that when only the vocalic transitions were presented (i.e., with both noise burst and steady-state vowel deleted), identification of initial stop consonants was hardly affected, but identification of final stop consonants decreased by about 20%, reaching the identification level of the initial consonants.

This led us to hypothesize that cutting away the vowel preceding the final transition had introduced the same abrupt onset and its consequences at the beginning of that transition, thus canceling the advantage it held, according to Ohde and Sharf, over the initial transition. Conversely, we expected that if we could manage to cancel the "click sensation" at the beginning of the CV transition, identification of deleted initial consonants should improve to the identification level of the final consonants. Replacing the deleted consonant portions with noise seems to be a way of avoiding abrupt onsets, and of presenting more natural stimuli to subjects: The segmented syllables appear to emerge from or vanish into noise. The possibility must not be excluded that this effect might be akin to the findings about continuity (e.g., "phoneme restoration") by Warren and Warren (1970), Warren and Obusek (1971), and Powers and Wilcox (1977).

TABLE I. Percent correct identification of initial and final consonants deleted from CV and VC syllables (adapted from Ohde and Sharf, 1977).

	Vocalic transition + whole vowel	Vocalic transition only
Initial consonants	39.3	43.9
Final consonants	65.4	45.7

FIG. 1. Examples of the waveforms of initial voiceless (a), initial voiced (b), and final voiced (c) Dutch plosives. The vertical lines mark the segmentation points, which define the six (initial consonants) or five (final consonants) experimental conditions; everything to the left (initial consonants) or to the right (final consonants) of these lines was deleted. In Fig. 1(b) segmentation point 1 is not shown.

I. METHOD

A. Stimuli

The stimuli were 24 Dutch CVC words pronounced in isolation by three native speakers of Dutch. Of these words, 15 consisted of /p, t, k, b, d, / + /i, u, a/ + /t/, and were used for the initial consonants; the other 9 words consisted of /t/ + /i, u, a/ + /p, t, k/, and were used for the final consonants (Dutch contains neither /g/, nor final /b/ and /d/). The words were stored digitally (10 bits, 12 500 Hz), after which the waveform of each word was displayed on a screen, so that the beginning or end of the vowel could be determined by visual and auditory inspection. Figure 1 shows typical examples of initial voiceless, [Fig. 1(a)]; initial voiced, [Fig. 1(b)]; and final voiceless, [Fig. 1(c)] Dutch plosives. Segmentation point 1 left the whole utterance intact. In Fig. 1(b) this point is not shown, its position being considerably to the left of the figure: Dutch initial voiced plosives contain a considerable voice bar. The end of the release burst/beginning of the vocalic transition was always taken to be a positive-going zero crossing; with voiced consonants this was the beginning of the first period that contained no noise [segmentation point 3 in Fig. 1(b)], whereas with voiceless consonants it had to be the beginning of

periodicity [segmentation point 2 in Fig. 1(a)]. As in Fig. 1(a), periodicity always started with a short period; if, as Holmes (1976) suggests, complete closure of the glottis coincides with the highest peak in the waveform of a period, this short period can probably be interpreted as the first closing of the glottis, which had, up to that time, been open. We felt, consequently, that this short period had to be included in the vowel transition from an initial voiceless consonant. The silent gap between vowel and final consonant made segmentation there very simple [segmentation point 2 in Fig. 1(c)].

Great care was taken to make every stimulus as click-free as possible, by selecting segmentation points as close as possible to the zero line. Since our step size was 80 μ s, whereas the accuracy of tape splicing or erasure (Sharf and Ostreicher, 1973) has to be expressed in ms, we had much better control over our actions than the studies cited above.

Three or four more stimuli were manufactured from each of the recorded words by removing more and more of the vocalic transition; these segmentation points (numbered 3 and higher) are also indicated by means of vertical lines in Fig. 1, and were the same (in terms of periods) for each word. In initial voiced consonants segmentation point 2 was the end of the voice bar and the beginning of the noisy part [see Fig. 1(b)]. Consequently, not all six conditions for voiced and voiceless initial consonants are strictly comparable. With final consonants, only three extra segmentation points were determined. As a result, there were six conditions for initial, and five conditions for final consonants.

B. Subjects

In each experiment ten male native speakers of Dutch took part as subjects; some, but not all, participated in both the initial- and the final-consonant experiments, but the same subjects were used for the same consonants both with and without noise. Subjects ranged from experienced listeners to completely inexperienced ones, but there were no clear systematic differences among them.

C. Procedure

The investigation consisted of four experiments:

- initial consonant identification with noise preceding the stimuli,
- initial consonant identification without noise,
- final consonant identification with noise following the stimuli, and
- final consonant identification without noise.

The noise in experiments (a) and (c) was a 300-ms burst of -3 dB/octave-(pink) noise, presented at a relatively low level (probably too low for "phoneme restoration"), and intended to cancel the abrupt onset effect; it replaced the deleted portion of the signal, or rather: In experiment (a) it ended at the point where the stimulus began and in experiment (c) it began where the stimulus ended. The stimuli themselves were presented over headphones. For a schematic drawing, see Fig. 2.

FIG. 2. Presentation of the stimuli (R=response; hatched area = noise). For final consonants, noise and stimulus are reversed; the noise could be switched off for the “no-noise” experiments.

The subject was seated in front of a computer terminal console with a screen; he heard the stimulus twice and immediately after that saw the response set on the screen: PTKBD in experiments (a) and (b), and PTK in experiments (c) and (d). He had to record his response by pressing the appropriate key on the keyboard; if he wanted to hear the stimulus again, which rarely happened, he could do this as often as he liked by pressing a specific key. One second after a response had been given, the next stimulus was presented. In every single experimental session only initial, or only final consonants were involved.

The stimuli from the different conditions (segmentation points) were generated in a random order all in one session; this meant that (3 speakers) × (5 consonants) × (3 vowels) × (6 segments) = 270 stimuli were presented in experiments involving initial consonants, and (3 speakers) × (3 consonants) × (3 vowels) × (5 segments) = 135 in experiments involving final consonants.

II. RESULTS AND DISCUSSION

The overall results are summarised in Fig. 3; in Fig. 3(a) the percentage correct identifications of the initial voiceless plosives is shown; Fig. 3(b) shows the same for the initial voiced plosives, and Fig. 3(c) for the final (voiceless) plosives. In each case, the horizontal axis marks the deletion conditions; continuous graph lines indicate stimuli with added noise, while dashed lines indicate stimuli without noise; chance level is marked with a horizontal line at 20% or 33.3%.

Of the curve pairs in Fig. 3, only the two curves in Fig. 3(a) are significantly different: $F = 52.64 (1, 29)$, $p < 0.01$. This means that adding noise only affects the

TABLE II. Identification scores of deleted initial and final voiceless plosives in condition 2 (only the burst deleted). The middle column shows the scores when “voiced” responses are added to the “voiceless” ones for initial consonants, reducing everything to a triple-choice place-of-articulation identification task.

	Initial	Initial b = p, d = t	Final
Without noise	p 66	p 66 + 22 = 88	p 79
	t 14	t 14 + 34 = 48	t 63
	k 13	k 13 = 13	k 66
	93 : 34%	149 : 55%	208 : 77%
With noise	p 75	p 75 + 7 = 82	p 83
	t 56	t 56 + 14 = 70	t 57
	k 46	k 46 = 46	k 68
	177 : 66%	198 : 73%	208 : 77%

perception of deleted initial voiceless plosives; perception of voiced initial plosives and final plosives remains unchanged. An analysis of variance furthermore revealed that there were significant differences between the data for the three speakers, for the consonants, and, of course, for the six (initial) or five (final) deletion conditions.

Identification of deleted final consonants is not affected by cancellation of abrupt offsets; if there is a “click sensation” at the end of a transition, it does not disturb perception of that transition. This is in agreement with the finding by Ohde and Sharf (1977) that cutting away the vowel following the transition belonging to an initial consonant (which, in our view, is tantamount to introducing a “click sensation” at the end of that transition), does not affect consonant identification much (see Table I).

Initial consonants show a different pattern; in Fig. 3(a), the important condition 2 [see Fig. 1(a)] shows an improvement of 32% with the addition of pink noise. However, there is a significant ($F = 7.32 (1, 29)$, $p < 0.05$) difference of 12% the other way for the initial voiced consonants [Fig. 3(b)] in condition 2, but no difference in condition 3 [see Fig. 1(b)]. Apparently, replacing a voice bar with an aperiodic noise burst introduces conflicting voicing cues. Nevertheless, the overall effect is quite clear: cancellation of the abrupt onset at the beginning of a transition significantly improves perception of that transition, as we predicted on the basis of Ohde and Sharf’s data. The overall effect for both initial and final voiceless plosives in condition 2 is shown in Table II; the middle column contains the results recalculated in terms of a place-of-articulation identification experiment, on the assumption that a /b/ response is equivalent to a /p/ response, and a /d/ response to a /t/ response. This was done in order to make identification scores of initial and final consonants more com-

FIG. 3. Overall percentage correct identification, as a function of segmentation condition. Parameter is the presence (continuous line) or absence (dashed line) of noise.

FIG. 4. Percentage correct identification for each of the three speakers. Parameter is the presence (continuous line) or absence (dashed line) of noise.

parable: Dutch has no final voiced consonants; neither is there a voiced place-of-articulation equivalent to /k/. It is clear that final stop consonant identification remains at 77% when noise is added, but that identification of initial consonants, which stood at 55%, is brought up to nearly the level of final-consonant identification by the addition of noise. That was the effect we sought to demonstrate.

That the added noise could have a slight masking effect, may be deduced from Fig. 3(a): in condition 1 (presentation of the whole word, including the plosive burst), the addition of noise causes identification to deteriorate by almost 4% ($F = 7.44$ (1, 29), $p < 0.05$).

The differences between the three speakers were significant for initial voiceless consonants and for final voiceless consonants ($p < 0.01$ in both cases), but not for initial voiced consonants. Figure 4 shows the differences between the speakers for the curve pairs that were found to differ significantly; it is clear that the trends seen in Figs. 3(a) and 3(c) are the same for all three speakers.

The only vowel effect we found was the fact that /a/ increased the positive effect of the addition of noise on initial /t/ and /k/, not only in condition 2, but also in conditions 3–6, where response to /ta/ and /ka/ was now well above chance.

The differences between the consonants were significant at the 1% level in all three subsets (initial voiceless, initial voiced, and final voiceless). Figure 5 shows the percentage correct recognition for each of the consonants averaged over the three speakers. It is clear that most of the improvement in condition 2 for the initial voiceless plosives caused by adding noise is due to initial /t/ and /k/, which reach chance level as soon

as anything is cut off, but retain a fairly high identification level in condition 2 when noise is added. Initial /p/ exhibits much better overall identification, which declines much more gradually, and has a much smaller noise-addition effect at condition 2. We think that the transition from initial /p/ to a vowel retains consonant cues much longer than do the transitions from initial /t/ and /k/; apart from any information contained in the consonant burst itself, the first short period of voicing seems to be particularly important for the identification of /t/. This did not come entirely unexpected: in Schouten and Pols (1978b) we had found that initial Dutch /p/ is characterized by considerably longer vocalic transitions than is initial Dutch /t/; /k/ was not investigated in that study.

The finding that initial /b/ and /d/ are *not* positively affected, and in condition 2 even slightly negatively affected, by the addition of noise is probably due to the fact that the actual transitions are considerably longer (and consequently contain place cues over a much wider time interval), with voiced consonants (see Schouten and Pols, 1978b), and it points to the possible conclusion that onset masking affects only the very first part of the transition; if the transition is long enough, spoiling its first one or two periods makes little difference.

Like Fischer-Jørgensen (1972), we observed a strong overall response bias towards /p/ for deleted consonants, both in initial and in final position. This bias did not disappear after noise had been added, so the interpretation that the "click sensation" has a /p/-like character, which we entertained for a while, can be rejected.

III. CONCLUSIONS

It is evident that it is very risky to conduct experiments in which deleted speech sounds have to be iden-

FIG. 5. Percentage correct identification for each of the consonants. Parameter is the presence (continuous line) or absence (dashed line) of noise.

tified: The abrupt start of a stimulus is bound to give rise to a "click sensation" which seems to affect the first 10–20 ms of a stimulus, so that if identification depends on what takes place during that time, the listener is thrown completely off course. This is the case with initial /t/ and /k/, and to a smaller extent, with initial /p/. We have shown that this state of affairs could be remedied by substituting noise for the removed speech sounds, although one should not lose sight of the possibility that the noise itself may also have some effect. Identification of initial /b/ and /d/ is not affected by an abrupt beginning, probably because of the much longer vowel transition. An abrupt ending does not seem to matter, at least not for final /p/, /t/, or /k/.

ACKNOWLEDGMENT

The research reported on in this article was made possible by a grant from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

- 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," *Annu. Rep. Inst. Phon. Univ. Copenhagen*, 1972.
- Holmes, J. N. (1976). "Formant excitation before and after glottal closure," *IEEE 1976 Int. Conf. ASSP*, 39–42.
- Kuehn, D. P., and Moll, K. L. (1972). "Perceptual effects of

- forward coarticulation," *J. Speech Hear. Res.* **15**, 654–664.
- LaRiviere, C., Winitz, H., and Herriman, E. (1975). "The distribution of perceptual cues in English prevocalic fricatives," *J. Speech Hear. Res.* **18**, 613–622.
- Ohde, R. N., and Sharf, D. J. (1977). "Order effect of acoustic segments of VC and CV syllables on stop and vowel identification," *J. Speech Hear. Res.* **20**, 543–554.
- Ostreicher, H. J., and Sharf, D. J. (1976). "Effects of coarticulation on the identification of deleted consonant and vowel sounds," *J. Phonet.* **4**, 285–301.
- Powers, G. L., and Wilcox, J. C. (1977). "Intelligibility of temporally interrupted speech with and without intervening noise," *J. Acoust. Soc. Am.* **61**, 195–199.
- Schouten, M. E. H., and Pols, L. C. W. (1978a). "Vowel segments in consonantal contexts: a spectral study of coarticulation, Pt. I," *J. Phonet.* (in press).
- Schouten, M. E. H., and Pols, L. C. W. (1978b). "CV- and VC-transitions: a spectral study of coarticulation, Pt. II," *J. Phonet.* (in press).
- Sharf, D. J., and Beiter, R. (1974). "Identification of consonants from formant transitions played forward and backward," *Lang. Speech* **17**, 110–118.
- Sharf, D. J., and Hemeyer, T. (1972). "Identification of place of consonant articulation from vowel formant transitions," *J. Acoust. Soc. Am.* **51**, 652–658.
- Sharf, D. J., and Ostreicher, H. J. (1973). "Effect of forward and backward coarticulation on the identification of speech sounds," *Lang. Speech* **16**, 196–206.
- Slis, I., and Cohen, A. (1969). "On the complex regulating the voiced-voiceless distinction I and II," *Lang. Speech* **12**, 80–102 and 137–155.
- Warren, R. M., and Obusek, C. J. (1971). "Speech perception and phonemic restorations," *Percept. Psychophys.* **9**, 358–362.
- Warren, R. M., and Warren, R. P. (1970). "Auditory illusions and confusions," *Sci. Am.* **223**, 30–36.