Reversal of the rise time cue in the affricate fricative contrast, an experiment on the silence of sound

Heuven, V.J.J.P. van

REVERSAL OF THE RISE-TIME CUE IN THE AFFRICATE-FRICATIVE CONTRAST: AN EXPERIMENT ON THE SILENCE OF SOUND*

Vincent J. van Heuven
Dept. of Linguistics/Phonetics Laboratory, Leyden University, P.O. Box 9515, 2300 RA Leiden, The Netherlands

1. INTRODUCTION

1.1. Cues underlying the affricate-fricative contrast

Traditionally, linguists distinguish true consonants (or: obstruents) along a three term manner of articulation dimension: stop, fricative, and affricate. In this paper we shall be concerned with the contrast between two of these categories: affricate versus fricative, as in the word pair chop - shop. This contrast has multiple acoustic cues, which are listed in table I:

Table I, acoustic cues involved in the affricate-fricative contrast

(1) Duration of the preconsonantal vowel (Isenberg, 1978)
(2) Decay rate of the preconsonantal vowel amplitude (Debrock, 1977)
(3) Duration of the pre-burst silent interval (e.g., Kuipers, 1955; Truby, 1955)
(4) Formant transitions of the preconsonantal vowel (Isenberg, 1978; Dorman, Raphael & Isenberg, 1980)
(5) Rise time of the friction noise amplitude (e.g., Gerstman, 1957)
(6) Duration of the friction noise (e.g., Gerstman, 1957)
(7) Rise time of the post-consonantal vowel amplitude (Debrock, 1977)
(8) Presence/absence of a release burst (Dorman et al., 1980)

Perceptual relevance has not been established for all of these acoustic correlates, let alone in a single experiment, but over the years the research has focused on three of them, and their trading relations: noise amplitude rise time (henceforth: rise time), noise duration (henceforth: duration), and the duration of the silent interval separating the preceding vowel and the friction noise (henceforth: interval). So far the following effects have emerged:

*I thank Peter Vroege for running the experiment described in this paper.
The waveform editing programme for the DEC Micro PDP-11/23 was developed in our laboratory by Ing. J.J.A. Pacilly and Drs. A.F.E. van der Horst.

Table II, Effects of cues involved in the affricate-fricative contrast

	fricative	affricate
rise time	gradual	abrupt
duration	relatively long	relatively short
interval	short (absent)	long

The research on trade-offs between these cues has had a long history, beginning with Gerstman (1957). A re-analysis of his data (van Heuven, 1979) has shown that duration and rise time can be traded *only* in the very narrow range between 90 and 130 ms noise duration, indicating that duration is the primary cue to the contrast, at least for stimuli with the contrast in initial position.

Tradings involving the interval can only be examined in a context with a segment (typically a vowel) preceding the contrast. Though there are several studies manipulating interval as a single parameter (e.g., Kuipers, 1955, Truby, 1955), trading research has been quite limited. Two-parameter studies involving the interval (against duration) are described by Repp, Liberman, Eccardt & Pesetsky (1978), and by Dorman et al. (1980) interval versus release burst, duration, and formant transitions. Regular trading relations were established for the interval parameter in all of these studies: the effect of a longer interval cueing affricate could be offset by a more fricative-like value for the competing parameter.

Tradings involving the rise time parameter in context have been investigated least of all. Yet, in one study rise time, too, was found to trade regularly with interval (Dorman, Raphael & Liberman, 1979, experiment also described in Dorman et al., 1980). Here the affricate-fricative contrast was examined in word-final position (ditch - dish) the effect of slower rise time cueing fricative could be counteracted by a longer interval (37 vs. 57 ms interval for cross-overs at 0 and 35 ms rise time, respectively).

Van Heuven (1983) studied the effects of rise time and duration in word-initial position (chop - shop) for isolated stimuli, and for the same tokens embedded in a carrier Why don't you say ... again? For isolated words I obtained results that were essentially similar to Gerstman's (1957), with rise time and duration trading in the regular fashion. However, when in try-outs listeners heard the same tokens in context, they reported exclusively affricates for silent intervals exceeding 20 ms. When, in the final experiment, the interval was fixed at an even shorter value of 15 ms, convincing cross-overs were obtained through manipulating rise time and/or duration, but as the results revealed, the contribution of the rise time had been reversed relative to its function in isolated tokens. in this context longer rise time contributed to affricate.

1.2. Possible causes for the reversal

In my (1983) paper I explained this curious reversal of the rise time cue in context as an effect of forward masking. This seemed a reasonable hypothesis, since the pre-burst vowel had been recorded in an utterance preceding a chop token, i.e., with a relatively abrupt intensity off-ramp as is characteristic of pre-stop vowels (Debrock, 1977). Plomp (1964), among others, has shown that the human ear is relatively insensitive to auditory stimulation shortly after the abrupt termination of a high intensity acoustic event. A subsequent low intensity sound will not be heard during this period of masking, or will at least appear weaker than when presented in isolation. The masking period may extend for as long as 250 ms, though the effect rapidly decays over time.

Due to masking, then, our listeners might have heard a brief interval of silence (or reduced energy), suppressing and replacing the low intensity noise onset of the friction sound. The smoother the noise onset, the longer the noise would remain below the masking threshold, creating a longer perceived gap as well as a shorter perceived noise duration, which two illusions then conspired to cue affricate.

However, in spite of the prima facie attractiveness of this account, there are complications that may force us to reconsider. For masking to occur it is necessary that the frequency distribution of the masker (here: vowel) coincides or overlaps with that of the probe (here: friction noise). Thus, in Plomp (1964) pure tones were masked by white noise (see further, Resnick, Weiss & Heinz, 1979). In my vowel-noise sequences it seems unlikely that the masker and probe frequency distributions were sufficiently similar to cause strong masking effects the vowel /ei/ has most of its energy below 2500 Hz, whereas the /sh/ noise has its energy concentrated above this value (Heinz & Stevens, 1961).

As an alternative explanation for the reversal of the rise time cue in context I now propose the following: Let us assume that it is a necessary condition for listeners to perceive an affricate that the stimulus contain a brief interval of silence (or reduced energy) immediately preceding the friction noise, so as to reflect the presence of an articulatory closure (cf. Dorman et al., 1979). If such a silent interval is absent from the physical stimulus, the listener may have to reinterprete the acoustic signal trying to satisfy the condition for a silent interval. It is conceivable, then, that he will consider the low energy portion in the smooth noise onset to be the silent interval.

Notice that this account is fundamentally different from the masking hypothesis. In the latter case it is assumed that the gradual noise onset is obscured due to a peripheral mechanism whereby the abrupt termination of acoustic energy cannot be resolved by the human ear. In the alternative view the ear is perfectly capable of resolving the intensity envelope of the vowel-friction sequence, but reinterpretes the available cues at a more central level. The purpose of the present paper is to choose between these two competing explanations.

1.3. Approach

If the cue reversal is indeed a matter of reinterpreting the available cues so as to perceive a silent interval, it should not matter to the listener whether the low intensity portion of the stimulus is located in the noise onset, or in the offset of the preceding vowel, as long as the energy dip occurs at the VC-boundary, i.e., at a point in time where a stop-closure can be located.

Therefore, changing the vowel offset from abrupt to gradual should increase the number of affricate judgments when the vowel is closely followed by the friction sound; moreover, gradual vowel offset and noise onset should reinforce one another, since both manipulations can be reinterpreted as the silent interval. Masking, on the other hand, should disappear when the vowel intensity decays gradually, and no effect of masking should obtain at all when the vowel off-ramp is longer than some 70 ms (cf. van den Broecke & van Heuven, 1983). As a result, the noise rise time cue should reverse after an abrupt vowel termination, but function normally (i.e., as in stimulus initial position) when the vowel decay is smooth. Thus, varying noise tise time versus vowel decay time in a two-parameter study will allow us to choose between the competing explanations.

2. METHOD

A male native speaker of R.P.-English recorded the sentence Why don't you say shop again observing normal intonation and timing. The audiosignal was digitised (10 kHz, 12 bits, 4.5 kHz LP cut-off) and stored in computer memory (DEC Micro PDP-11/23). Using a digital tape editing program a new utterance was created by concatenating parts of the digital record, as follows.

The utterance was truncated after the word say at a positive going zero crossing at the end of the glottal period whose intensity was no less than 90 percent of the peak value reached throughout the vowel. As it happened, this eliminated the final three glottal periods from the vowel. To this abrupt vowel termination were appended a 10 ms silent interval and a 60 ms stretch of steady state [sh]-noise, gated out from the centre portion of the original fricative. This in turn was followed by the final 20 ms of friction before the CV-boundary in shop and the remainder of the original utterance. As a result the friction noise had an abrupt onset, and lasted for 80 ms, including the transition into the following vowel. The particular temporal organisation was chosen so as to create a stimulus that could be interpreted as either shop or chop.

The resulting record was DA converted and passed through analog gates (Grason-Stadler 1284B) which were modified so as to allow continuous adjustment of rise/fall times (van den Broecke & van den Broek, 1978) with a precision of .1 ms. Twenty-five stimulus types were manifactured differing in the vowel offset in say and the friction onset in the ambiguous word shop/chop: five vowel decay times, ranging from smooth (80 ms linear decay) to abrupt (0 ms decay) in steps of 20 ms were generated, and orthogonally combined with 5 noise rise times (same steps). These were recorded onto audio tape in

8 different random orders, preceded by 10 practice items, and separated by 2 s interstimulus intervals (offset to onset).

The entire tape was played in a quiet room over headphones to 5 audiometrically normal adult native English listeners. They were instructed to decide for each stimulus whether they perceived <u>shop</u> or <u>chop</u> with binary forced choice.

3. RESULTS AND CONCLUSION

Percent fricative judgements was determined for each of the 25 stimulus types (N = 40 judgments per stimulus type). Figure 1 plots these percentages as a joint function of vowel offset (horizontal dimension) and noise onset (vertical dimension).

Let us concentrate, first of all, on the results obtained for stimuli with abruptly terminating vowels. Here we notice that abrupt noise onset is associated with fricative, and gradual onset with affricate. This effect runs counter to what is usually claimed in the literature (where gradual noise onset is a cue for fricative), but replicates our earlier results for this parameter with <u>shop/chop</u> tokens presented in a spoken context. Apparently, both the abrupt noise onset and the 10 ms of physical silence are completely masked by the preceding vowel. Both to ourselves and to our subjects the transition of vowel into consonant sounded perfectly smooth.

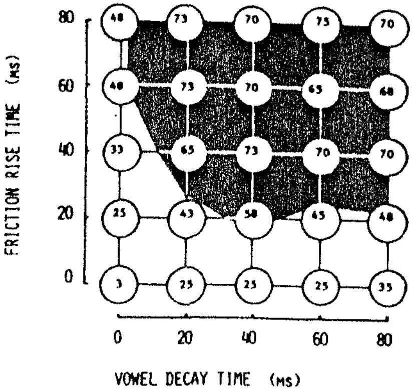

FIGURE 1, Percent <u>chop</u> responses as a function of the rise time of the friction noise and decay time of the preconsonantal vowel. The phoneme boundary separating affricate (shaded) from fricative (open) areas is drawn through the 50% cross-over points, which were determined by linear interpolation between stimuli straddling the boundary.

So far the results are in line with either explanation. However, the reversal of the rise time cue should disappear when the occurrence of forward masking is prevented. To this effect the preconsonantal vowel offset was systematically varied between abrupt (strong forward masking) and gradual (no or very weak masking). Clearly, the results falsify the prediction based on masking: the reversal of the noise onset cue is maintained regardless of the vowel offset characteristic, and more gradual vowel offsets are associated with affricate. The effects of vowel offset and noise onset are roughly additive, at least if we ignore the fact that the affricate scores plateau at 70-75%. These results unequivocally support the alternative hypothesis based on a central reinterpretation of low intensity sound as silence.

·. DISCUSSION

Though the masking hypothesis has been convincingly falsified, should be noted that some measure of masking still persists. It was bserved that the 10 ms silent gap separating vowel and friction sound as inaudible, which effect has to be ascribed to masking. Apart from us, temporal resolution of the intensity envelope was excellent iroughout the experiment, since even the insertion of a mere 20 ms oise on-ramp or vowel off-ramp brought about a 22% increase of ffricate judgments (cf. figure 1). Thus it would seem that the effects f forward masking in the present speech context are extremely mited, which finding is in line with e.g. Slis & van Nierop (1970) and esnick et al. (1979) who showed that vowel onto consonant masking , quite small. Even at masker-probe intervals as short as 25 ms the mount of masking never exceeded -18 dB, while the intensity ifference between adjacent vowels and consonants in normal speech is lways less than this value.

Secondly, our results bear out that a silent interval is indeed a ecessary condition for the perception of the affricate (or stop) anner. When such an interval is physically absent, low intensity ortions of the stimulus flanking the VC-boundary are reinterpreted as ilence. This behaviour seems to support the view that during speech erception the acoustic cues are evaluated in the light of what the istener knows about articulation. One wonders if reinterpretation of ues could be used in a more principled way to examine the relative nportance of multiple cues in phonetic contrasts. Clearly, if one cue an be reinterpreted as an other, but not vice versa, the on-negotiable cue is the stronger of the two. In this light our results ndicate once more that rise time is a manner cue of limited mportance, especially for contrasts occurring in connected speech.

Thirdly, "sound" reinterpreted as "silence" provides a less owerful cue to affricate than physical silence does. We may observe hat affricate judgments plateau at 70-75%, which means that the ffricate end of the stimulus space was not highly convincing. Given he results of other studies much better exemplars of affricates can be generated if a proper period of silence is inserted between vowel and riction burst.

Finally, the reinterpretation hypothesis assumes that the perceived length of silence is much larger than the time interval that energy drops below threshold. If only the below-threshold portion of he energy dip contributed to the perceived silent interval, affricate udgments should not have plateaued at 75%; instead, each smoother ioise onset or vowel offset should have boosted the percentage of affricates. Therefore it might be worthwhile exploring in rather more detail the perceptual equivalence of true silence and sound einterpreted as silence. The adjustment paradigm seems particularly suited to this purpose. Subjects can be asked to adjust the duration of a true silent interval (with sharply defined boundaries) so as to be perceptually equal to a stimulus with smooth energy dips (symmetrically or asymmetrically) distributed on both sides of the VC-boundary. When such experiments are done with non-speech stimuli (e.g., sawtooth-white noise sequences), the adjusted silent interval should equal the time the energy dip remains below threshold. In more speech-like stimuli (e.g., vowel-affricate), subjects will tend to exaggerate the pereceived length of silence so as to make the stimulus fit a stop percept.

REFERENCES

1. Broecke, M.P.R. van den and Broek, D. van der (1978). Matching slopes with electronic switches. Progress Report of the Institute of Phonetics Utrecht 3.2, 24-26.
2. Broecke, M.P.R. van den and Heuven, V.J. van (1983). Effect and artifact in the auditory perception of rise and decay time: speech and non-speech. Perception and Psychophysics 33, 305-313.
3. Debrock, M. (1977). An acoustic correlate of the force of articulation. Journal of Phonetics 5, 61-80.
4. Dorman, M.F., Raphael and L.J., Liberman, A.M. (1979). Some experiments on the sound of silence in phonetic perception, Journal of the Acoustical Society of America 65, 1518-1532.
5. Dorman, M.F., Raphael, L.J. and Isenberg, D. (1980). Acoustic cues for a fricative-affricate contrast in word-final position, Journal of Phonetics 8, 397-405.
6. Gerstman, L.J. (1957). Perceptual dimensions for the friction portions of certain speech sounds. Unpublished Ph.D. dissertation, New York University.
7. Heinz, J.M. and Stevens, K.N. (1961). On the properties of voiceless fricative consonants. Journal of the Acoustical Society of America 33, 589-596.
8. Heuven, V.J. van (1979). The relative contribution of rise time, steady time, and overal duration of noise bursts to the affricate-fricative distinction in English. In J.J. Wolf and D.H. Klatt (eds.), ASA*50 Speech Communication Papers. The Acoustical Society of America, New York, 307-311.
9. Heuven, V.J. van (1983). Rise time and duration of friction noise as perceptual cues in the affricate-fricative contrast in English. In M.P.R. van den Broecke, V.J. van Heuven and W. Zonneveld (eds.), Sound structures, studies for Antonie Cohen. Foris, Dordrecht, 141-157.
10. Isenberg, D. (1978). Effect of speaking rate on the relative duration of stop closure and fricative noise. Haskins Laboratories Status Report on Speech Research SR-55/56, 63-79.
11. Kuipers, A.H. (1955). Affricates in intervocalic position, Haskins Laboratories Quarterly Progress Report 15, appendix 6.
12. Plomp, R. (1964). Rate of decay of auditory sensation. Journal of the Acoustical Society of America 36, 277-284.
13. Repp, B.H., Liberman, A.L., Eccardt, T., and Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, affricate, and fricative manner. Journal of Experimental Psychology: Human Perception and Performance 4, 621-637.
14. Resnick, S.B., Weiss, M.S. and Heinz, J.M. (1979). Masking of filtered noise bursts by synthetic vowels. Journal of the Acoustical Society of America 66, 674-677.
15. Slis, I.H. and Nierop, D.J.P.J. van (1970). On the forward masking threshold of vowels in VC-combinations. IPO Annual Progress Report 5, 68-72.
16. Truby, H. (1955). Affricates. Haskins Laboratories Quarterly Progress Report 11, 7-8.