INFLUENCE OF VOCALIC ENVIRONMENT ON PERCEPTION OF SILENCE IN SPEECH: AMPLITUDE EFFECTS

Bruno H. Repp

Abstract. This experiment investigated the influence of the relative amplitudes of the preceding and following vocalic portions on the perception of silence as a cue for the distinction between single and double stop consonants, both for nonidentical (/b/ vs. /b-g/) and identical (/b/ vs. /b-b/) places of articulation. The effects were generally small and in the direction of increased double-stop responses as amplitude increased. In the case of non-identical places of articulation, only the preceding vocalic portion had a significant effect, whereas both vocalic portions had independent effects in the case of identical places of articulation. These results supplement those of Repp (1979) concerning effects of spectrum and duration of vocalic context; together, they place constraints on the form that a theory of silence perception in speech might take.

INTRODUCTION

In earlier studies (Repp, 1979) I investigated the influence of spectrum and duration of preceding and following vocalic portions on the perception of silence in speech. Their effect was measured in terms of the amount of silence needed to perceive (on half of the trials) a sequence of two stop consonants whose places of articulation were cued by formant transitions into and out of the silent interval. In one condition, loosely referred to as the single-cluster condition, the two sets of transitions conveyed different places of articulation (/b-g/); in the other condition, the single-geminate condition, both were appropriate to the same place (/b-b/). Roughly 70 msec of silence are needed to hear both stops in /b-g/ (only /g/ is heard at shorter silences), and about 200 msec are needed to hear both stops in /b-b/ (only one /b/ is heard at shorter silences). For general introductions and for discussions of these effects, the reader is referred to Dorman, Raphael, and Liberman (1979), Repp (1978), and, of course, Repp (1979), whose studies the present experiment supplements.

The purpose of my earlier experiments (Repp, 1979), as well as of the present study, was not so much to decide between alternative hypotheses concerning the perception of silence in speech—although some tentative

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR05596. Thanks are due to Patti Price who ran the experiment and tabulated the data.

[HASKINS LABORATORIES: Status Report on Speech Research SR-59/60 (1979)]
conclusions could be drawn—as to provide data that would form the basis for a future comparison with psychophysical observations on the perception of silence in nonspeech context. When we know more about the psychoacoustics of silence perception—and our knowledge is quite incomplete in that regard—such a comparison will certainly place a strong constraint on the form that an appropriate theory of the role of silence in speech perception might take.

In addition to spectrum and duration—the factors investigated earlier (Repp, 1979)—amplitude is an important auditory parameter. By investigating its influence on silence perception in speech, the present study extends and complements the earlier experiments. There was an additional motivation of the present study: In the earlier experiment on spectral effects, certain alterations in the amplitudes of the synthetic stimuli were made that may have confounded the results. The present experiment was expected to provide an estimate of the magnitude of such possible confounding, so that the earlier findings might be re-evaluated.

Method

Subjects. Twelve subjects participated. They included ten paid student volunteers with little experience in listening to synthetic speech, a research assistant with some listening experience, and the author, a seasoned subject. The results of all subjects were combined, as no qualitative differences were apparent.

Stimuli and procedure. Stimuli and procedure were identical with those of Repp (1979, Exp. II), except that variations in amplitude replaced variations in duration. Only the most important information will be given here. The stimuli were synthetic and consisted of a VC portion (/ib/, 190 msec long) followed by a variable period of silence and a CV portion (/ga/ or /ba/, 290 msec long). In the single-cluster condition (/ib-ga/), the silent interval varied from 15 to 115 msec in 10-msec steps. In the single-geminate condition (/ib-ba/), it varied from 115 to 315 msec in 20-msec steps. The relative amplitudes of the VC and CV portions were varied orthogonally in three 6-dB steps. The amplitude changes in each portion were +6, 0, and -6 dB relative to the baseline stimuli (Repp, 1979); the changes were implemented in the digitized wave forms before the test sequences were recorded. Each subject heard each of the 99 stimuli in each condition (9 amplitude combinations, 11 silence durations) 8 times in randomized order and identified the stop consonant(s) heard. Single-cluster and single-geminate conditions were presented as separate blocks in counterbalanced order.

Results

Figure 1 shows percentages of double-stop (cluster or geminate, depending on the condition) responses as a function of silence duration. Separate response functions are shown for the nine VC-CV amplitude combinations in the two conditions. The 50-percent cross-overs of these functions—about 70 msec for the single-cluster distinction and somewhat below 200 msec for the single-geminate distinction—are in good agreement with earlier data. The effect of VC amplitude can be seen within each panel, whereas the effect of CV amplitude extends vertically across panels.
Figure 1. Percentage of double-stop responses as a function of silence duration for nine VC-CV amplitude combinations in two experimental conditions.
It is evident that amplitude effects were rather small, especially in the single-cluster condition. Nevertheless, the effect of VC amplitude did reach significance in that condition, $F(2,22) = 4.3, p < .05$. The overall percentage of cluster responses increased slightly with VC amplitude. There was no effect of CV amplitude here, and no interaction between VC and CV amplitude effects.

In the single-geminate condition, the effect of VC amplitude was somewhat larger and more consistent across different silence durations, as can be seen in Figure 1. It, too, was in the direction of more double-stop (here: geminate) responses as VC amplitude increased. The effect was highly significant, $F(2,22) = 10.2, p < .001$. In addition, there was a significant effect of CV amplitude, $F(2,22) = 7.2, p < .01$, again in the same direction. The VC and CV amplitude effects appeared to be independent, as there was no significant interaction. Note that this implies an increase in geminate responses with overall stimulus amplitude; thus, it was not the relative amplitudes of the VC and CV portions but their absolute levels that mattered.

Since some of these effects are difficult to see in Figure 1, the results are summarized in more concise form in Figure 2. Instead of category boundaries—which are difficult to estimate accurately from response functions with indeterminate asymptotes, such as those in Figure 1—Figure 2 simply plots the percentages of double-stop responses, averaged across all silence durations, as a function of the two amplitude parameters. Apart from the effects just discussed (now more clearly visible), it is evident that the effect of VC amplitude was nonmonotonic: a 6-dB attenuation had a larger effect than a 6-dB amplification.

**DISCUSSION**

These amplitude effects are small compared to those of changes in VC (and CV) spectrum and duration, especially in the single-cluster condition. Therefore, the pre-experimental amplitude changes perpetrated in the stimuli of the spectrum experiment (Experiment I of Repp, 1979) probably had little influence on the results. Moreover, the only substantial changes (9–10 dB attenuation) had been made in /ba/ and /ga/, both CV stimuli. Since the present experiment indicates little or no effect of CV amplitude over a 12-dB range, the earlier results are vindicated.

Consider now the theoretical implications of the present results. In my earlier paper (Repp, 1979), I discussed three hypotheses about the perception of silence in speech. According to the first, the backward masking hypothesis (which really applies to the single-cluster condition only), cluster responses should have increased with increases in the amplitude of the "target" (the VC portion), as indeed they did; however, one might also have expected them to decrease with increases in the amplitude of the "mask" (the CV portion), which they did not. Thus, the evidence is equivocal with respect to the backward masking hypothesis.

According to the second hypothesis, the articulatory hypothesis, the perceptual results should mirror what happens in speech production. Of course, simple amplitude changes of the sort used here hardly ever occur in natural speech production; therefore, the articulatory hypothesis cannot be
Figure 2. Percentage of double-stop responses, averaged over all silence durations, as a function of VC and CV amplitudes in two conditions.
easily evaluated here. Nevertheless, we note that amplitude changes in production are primarily associated with changes in stress, and the present experimental stimuli probably could have been judged by listeners to vary systematically in their stress pattern. However, perceived stress depends, especially when only a single parameter varies, on the relation between two signal portions, and the present results showed that this relation did not influence the subjects' responses. For this reason, the articulatory hypothesis seems not to be supported by the present data.

The third hypothesis, the differentiation hypothesis, claims that the listener's task is the perceptual separation of the auditory-phonetic events preceding and following the silence. One factor that facilitates separation is an increase in the effective (subjective or physical) duration of the silent interval. Increases in the amplitudes of the VC and CV portions presumably increased their distinctiveness as "markers" of this interval. Although it is not clear why this should increase the subjective duration of the silence (and there seem to be no relevant data from psychophysical studies), it might reduce uncertainty about the boundaries of the silence. There is evidence from the auditory literature that increases in the amplitude of brief, burst-like markers increase the discriminability of silent intervals (Abel, 1972; Carbotte & Kristofferson, 1973; Divenyi & Danner, 1977; Divenyi & Sachs, 1978). Thus, in the present study, the perceptual salience of the silent interval may have increased with increased signal amplitude, leading to better perceptual separation of the VC and CV portions, and of their associated phonetic messages. At the same time, the clarity of the signal components themselves may have improved. Thus, of the three hypotheses considered, the differentiation hypothesis seems to be most compatible with the present findings.

These conclusions are necessarily tentative. Psychophysical studies are planned to compare auditory and phonetic perception of silence more directly by matching auditory stimuli to the present VC-CV stimuli in duration, amplitude, and—as closely as possible—spectral characteristics. These studies should reveal whether the perception of silence in speech can be accounted for by auditory principles, or whether specifically phonetic processes must be postulated.

REFERENCES

Repp, B. H. Perceptual integration and differentiation of spectral cues for
