Some Maskinglike Phenomena in Speech Perception*

M. F. Dorman, L. J. Raphael, A. M. Liberman, and B. Repp

ABSTRACT

To study maskinglike effects in the perception of continuous speech, listeners were presented two-formant syllables /bëb/ or /bëg/ followed, at intervals from 0 to 150 msec, by /de/. The subjects were instructed to identify the syllable-final consonant. An 80-msec intersyllable interval was required for recognition of the syllable-final consonant to reach asymptote. To determine the level at which this "effect" occurred, we repeated the experiment, but with the second-formant transitions of the first syllable presented alone for judgment as rising or falling chirps. Recognition of the isolated transitions (chirps) was essentially unaffected. These data suggest that the "masking" in the initial study was due to the elimination of a necessary cue—in this case, a silent interval, corresponding to the stop closure between the syllables—and not to backward masking of the auditory information. A third study found that changing voice between the syllables from male to female also eliminated almost all the "masking." This reinforces the conclusion of the first two studies, indicating in this case, that the effect is not to be attributed to interruption of information processing.

It has been suggested recently that forward and backward masking may constrain the perception of phonetic segments. Thus, Massaro (1972) has proposed that "...the redundancy of a vowel in normal speech...protects it from later speech until processing has been completed" and, in the same spirit, that "...if a consonant-vowel transition was followed by a speech sound that could not be integrated with it, perception should be disrupted, and backward recognition masking should occur."

*This is a revised version of a paper presented at the 89th meeting of the Acoustical Society of America, Austin, Texas, 7-11 April 1975.

† Also Herbert H. Lehman College of the City University of New York.

‡‡ Also University of Connecticut, Storrs, and Yale University, New Haven, Conn.

§§ University of Connecticut Health Center, Farmington, Conn.

[HASKINS LABORATORIES: Status Report on Speech Research SR-42/43 (1975)]
In this paper we will examine several cases of speech perception that fit the paradigms for forward and backward masking. Our purpose is to see if the underlying processes are, indeed, those of masking, either peripheral (integration of target and mask) or central (disruption of processing).

For the case of forward masking, we have chosen an instance of a type long familiar to students of speech perception: to perceive the stop in a syllable-initial fricative-stop cluster, we must have a period of silence between the fricative noise and the start of the stop transitions. The syllables in our experiment are /pe/ and /ke/. In Figure 1 we have shown a schematic spectrogram sufficient for the perception of /pe/. Seen in terms of the forward-masking paradigm, the noise of the fricative /ʃ/ would be the mask and /pe/ (or /ke/) the target. In our first experiment we varied the silent interval (let us call it the interstimulus interval or ISI) between the /ʃ/ mask and the /pe/ or /ke/ targets. The resulting stimulus patterns were randomized and presented to 14 listeners for judgment as /pe/, /ke/, or /e/. We see in Figure 2 that "masking" did occur: at ISIs of 20 msec or less the listeners reported hearing /e/, not /pe/ or /ke/.

To gain some insight into the processes underlying the failure to hear the stop, we undertook a second experiment to determine if there was, in fact, a masking of the essential acoustic cue—the second-formant transition—for the perceived distinction between /pe/ and /ke/. For that purpose, we followed exactly the procedures of the first experiment except that, in this case, the targets were not the syllables but only their second-formant transitions. These isolated transitions are not heard as speech; rather, they sound like "chirps," which our subjects easily learned to identify as "high" or "low." The outcome of this second experiment is shown in Figure 3. We see that correct perception of the "chirps" was not noticeably affected by the /ʃ/ mask. From that we infer that our subjects' failure to hear the stops in the first experiment was not due to masking in the ordinary sense. That is, the role of the silent interval between the /ʃ/ noise and the stop transitions is not, apparently, to avoid interference between target and mask. For a more reasonable interpretation we should note that in the production of initial fricative-stop clusters, closure must occur after the fricative, and therefore we may suppose that the silence caused by the closure is a essential manner cue for the perception of the stop. On that interpretation, the silence provides information, not freedom from interference.

Let us turn now to the paradigm for backward masking and in that connection consider the perception of the disyllables, /beb dc/ and /beg dc/. A schematic spectrogram sufficient for the perception of /beb dc/ is shown in Figure 4. As a case for backward masking, the syllable-final consonant /b/ in /beb/ is the target and the syllable /dc/ is the mask. To determine, then, whether masking does occur in this case we varied the silent interval between the mask /dc/ and the targets /beb/ or /beg/, randomized the resulting patterns, and presented them to 13 subjects for judgment as /beb dc/, /beg dc/, /be dc/. The outcome is shown in Figure 5. We see that at ISIs of 50 msec or less the subjects reported hearing /be dc/—that is, they did not hear the syllable-final stops /b/ and /g/.

To find out more about the underlying processes, we carried out for this paradigmatic case of backward masking an experiment analogous to the forward-masking experiment with the chirps. That is, we isolated the acoustic cue for
Figure 1: Stylized spectrogram of /ʃpɛ/. 

Figure 1
Figure 2: Percent identification of /fε/ and /ʃpεʃkε/.
Figure 3: Percent correct identification of "chirps" and /pɛ s/kɛ/.
Figure 4: Stylized spectrograms of /bɛb dɛ/.
Figure 5: Percent identification of /bɛ/ and /bɛb bɛg/.
the perceived distinction between /bɛb/ and /bɛg/-the second-formant transitions that, by themselves, sound like chirps—and, substituting them for the target syllables /bɛb/ and /bɛg/, we added the /dɛ/ mask exactly as it had been added in the experiment where it has, at short ISIs, effectively masked the stop-consonant targets. (The subjects were taught to identify the "chirp" as high or low.) The outcome is shown in Figure 6. We see that the correct perception of the chirps was little affected by the mask. This suggests that the processes underlying the failure to hear the stops was not due to masking of the differential acoustic cue.

Once again we might suppose that, as in the paradigmatic case of forward masking, the role of the necessary silent interval is to provide information, not time for processing. As in the earlier case of forward masking, that supposition is reasonable on the basis that the disyllables /bgɛ dɛ/ and /bɛg dɛ/ can be produced only if the speaker closes his vocal tract (thus producing an interval of silence) between the end of the first syllable and the beginning of the second. In the case of backward masking, however, there remains a masking interpretation to which the results with the chirps are not necessarily relevant: it is possible that the mask (the /dɛ/ syllable) interrupted the phonetic (as opposed to auditory) processing of the target. It is difficult to test that hypothesis directly, but the data of the next experiment do bear on it.

The next experiment was like the backward-masking case just described except that there were three targets—/bab/, /bad/, and /bag/—followed by a single mask /da/. The outcome is shown in Figure 7, where we see that the length of the necessary silent interval is different for the three targets; /bad/ in particular stands out, needing a much longer silent interval than the other two. One can think of no reason why the perception of syllable-final /d/ should require so much more processing time than the other stops, which is the assumption that a masking-process interpretation demands. On the other hand, it is quite obvious that the normal production of the geminate /bad da/ requires a longer vocal-tract closure than /bab da/ or /bag da/ (Delattre, 1971). Thus, this result points once again to the conclusion that silence is here a cue, a source of information. Moreover, it suggests, more compellingly perhaps than the earlier experiments, that phonetic perception is in this case constrained not primarily by what the auditory system can do but by what vocal tracts can do. The auditory system could hear the chirps even at short ISIs, but no vocal tract can produce the geminated stops with such short closures.

On the assumption that perception might here be obeying vocal-tract constraints, we ask next: Whose vocal tract? Common sense suggests that it would not be that of the listener or the speaker nor yet of any other individual, but rather some very abstract conception, which somehow takes account of what can and cannot be done by vocal tracts in general. In that connection we note that, as we have already remarked, a vocal tract cannot produce /bab da/ or /bad da/ without closing between the syllables and, as we have found, a listener cannot perceive both stops unless there is a corresponding interval of silence. But that applies only to a single vocal tract. Given two speakers, one can produce the first syllable and the other the second syllable with no silent interval between. We thought it of some interest to determine experimentally if the constraints that applied in the perception of utterances produced by a single voice would apply equally when perceptibly different voices produce the target syllable and the mask syllable.
Figure 7: Percent correct identification of syllable-final /b d g/ when followed by /d/.
Figure 8: Percent correct identification of syllable-final stops when followed by a syllable in the same voice or in a different voice.
In the last experiment, then, we duplicated the procedures of the earlier experiment in which we had varied the silent interval between target syllables /bab/ and /bag/ and a mask /da/, but in one case both syllables were spoken by a male while in the other the target was spoken by a male and the mask by a female. Two test sequences were produced, one for the same-voice condition, the other for the different-voice condition. Within each sequence the patterns were randomized and presented for judgment of the syllable-final consonant /b/ and /g/, as in the earlier experiments. Ten subjects were told before each condition that the syllables either were or were not produced by the same voice. The outcome is shown in Figure 8. The solid line (labeled "male") is for the same-voice condition. We see much the same result that had been found earlier: at relatively short ISIs the syllable-final consonants are not heard. The results for the different-voice condition are shown by the dashed lines (labeled "female"). We see that eight of the listeners in the different-voice condition correctly perceived the syllable-final consonants, even at the very shortest ISIs. Two of the listeners performed in the different-voice condition exactly as they had in the same-voice condition. (It is worthy of note that one of these two listeners spontaneously commented at the end of the experiment that she had not thought that the voices were different; she had assumed, rather, that it was the same person speaking at two different pitches.)

Though many controls need now to be carried out, we shall tentatively conclude on the basis of the last experiment that if phonetic perception is, in any case, constrained by what vocal tracts can and cannot do, the constraint is a very abstract one indeed. From all the experiments here reported, we shall conclude, somewhat less tentatively, that the role of silence before (or after) a stop is not to avoid interference (as between target and mask) but to provide important information. Putting that less tentative conclusion together with the more tentative one, we might say that the information is important because it tells the listener what a vocal tract is doing.

REFERENCES
