

# DICHOTIC BACKWARD MASKING OF COMPLEX SOUNDS

C. J. DARWIN†

*Haskins Laboratories, New Haven*

In the first experiment subjects identified a consonant-vowel syllable presented dichotically with a known contralateral masking sound at a stimulus onset asynchrony of  $\pm 60$  msec. When the mask followed the target syllable, perception of place of articulation of the consonant was impaired more when the mask was a different consonant-vowel syllable than when it was either a steady-state vowel or a non-speech timbre. Perception was disturbed less when the mask preceded the target, and the amount of disruption was independent of which mask was used. Greater backward than forward masking was also found in the second experiment for the identification of complex sounds which differed in an initial change in pitch. These experiments suggest that the extraction of complex auditory features from a target can be disrupted by the subsequent contralateral presentation of a sound sharing certain features with the target.

## Introduction

The traditional task in experiments on the temporal course of auditory masking has been the *detection* of a target presented in close temporal proximity to a mask. This paradigm has shown only small effects when target and mask are presented to opposite ears (dichotically). Moreover, these effects have been found only over very brief stimulus-mask intervals. Elliott (1962), for example, found virtually no forward masking of a brief tone by contralateral white noise, and only slight backward masking extending out to an inter-stimulus interval of about 15 msec.

Recently Studdert-Kennedy, Shankweiler and Schulman (1970) have reported an experiment requiring *identification* of two stop-consonant syllables presented dichotically with a temporal offset between them. They found that for offsets between about 15 and 120 msec the lagging syllable was reported more accurately than the leading syllable. Their result has since been confirmed both in the original paradigm (Berlin *et al.*, 1970; Lowe *et al.*, 1970) and in a slightly different paradigm in which only one sound has to be reported on a single trial (Kirstein, 1970, 1971). No advantage for the lagging over the leading sound, however, was found in binaural presentation (with both syllables coming to both ears) even when the duration of the vowel portion of each syllable was drastically curtailed to eliminate temporal overlap between the two sounds (Porter, 1971). Such curtailing did not influence the dichotic effect.

In terms of masking, these experiments have shown that under dichotic presentation stop-vowel syllables are more effective as mutual backward than as mutual forward maskers, whereas under binaural presentation, provided they do not temporally overlap, any masking that occurs is symmetrical.

† Now at University of Sussex.

In the visual modality, dichoptic masking is essentially a contour interaction (Schiller, 1965; Kahneman, 1968), which is asymmetrical so that backward masking is greater than forward. This asymmetry supports theories which emphasize the interruption of perceptual processes by the mask, rather than a temporal summation of mask and target (Kahneman, 1968; Spencer and Shuntich, 1970). A similar explanation seems appropriate for the auditory case (Studdert-Kennedy, Shankweiler and Schulman, 1970), although for stop-vowel syllables the effect is confined to dichotic presentation, whereas in vision monoptic contour interactions can be obtained (Schiller, 1965).

The present study pursues the analogy between dichoptic and dichotic masking. In the auditory experiments reviewed above it is not clear whether the more effective backward than forward masking is confined to a particular type of mask, since only syllables have been used to mask syllables. The first experiment therefore examines the relative extent of forward and backward masking for a number of different masks on a stop-vowel target set.

### Experiment I

The masks used in this experiment were chosen to have certain properties in common with the target set. Three were speech and the fourth a non-speech timbre. The three speech sounds were (a) a steady-state vowel different from that used in the target syllables, (b) the same vowel as used in the target syllables and (c) a stop-vowel syllable with the same vowel as the targets but a different stop consonant.

### Method

The targets used in this experiment were the four stop-vowel syllables /bε, dε, pε, tε/. These four consonants give two values each on the traditional phonetic dimensions of place of articulation and voicing. The four masks were /gε, ε, /ɔ and a non-speech steady-state timbre, which had three formants at 894, 2910 and 3698 Hz respectively. The two steady-state vowel masks and the five syllables were all highly intelligible. All the sounds were synthesized with three formants on the Haskins parallel formant synthesizer. Each sound lasted 100 msec and all the sounds had the same intonation contour and were equated for peak amplitude on a VU meter. On each trial of the experiment a subject heard one of the target sounds in one ear and one of the masks in the other. He always knew which mask would occur since this was held constant over a block of 48 trials and was played to him before each block, but he did not know into which ear the target would come. His task was simply to identify which of the four targets had been presented; he did not have to say into which ear it had come. The sounds on the two ears were always temporally offset by 60 msec. Whether the target or the mask led was randomly determined with the restriction that within each block of 48 trials each target item led six times and lagged six times. 16 subjects each took eight blocks of 48 trials in a Latin square design which counter-balanced the order in which the four masks were heard. The subjects were given a binaural demonstration of the set of target items before taking the dichotic test. Before each block the mask for that block was played three times binaurally.

### Results

Three different scoring methods were used: (a) the response had to be entirely correct (both place of articulation and voicing); (b) only place of articulation need be correct; and (c) only voicing need be correct. The results according to these

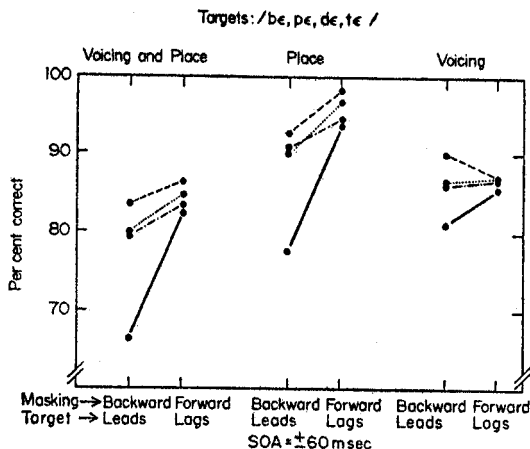


FIGURE 1. Mean percents correct for stop-vowel syllables dichotically opposed by a mask at  $\pm$  60 msec offset. The three columns refer to three different scoring criteria  
Mask: g $\epsilon$  —  $\epsilon$  ····;  $\circ$  — ·—·. Non-speech — —.

three methods are shown in Figure 1. The slope of each line indicates the difference between the target leading and target lagging conditions for the various masks. A line with a positive slope indicates that the target is better perceived when it lags the mask than when it leads it.

In the scores where both place of articulation and voicing had to be correct, analysis of variance showed a significant interaction of the lead/lag factor with mask ( $F(3, 105) = 3.98, P < 0.01$ ). However, since an analysis of variance on the results for place of articulation and voicing separately showed a significant difference between these two features for the interaction of lead/lag and mask ( $F(3, 45) = 3.16, P < 0.05$ ) as well as a significant interaction between the feature analysed and lead/lag condition ( $F(1, 15) = 23.8, P < 0.001$ ), the results will now be described separately for these two features.

For place of articulation, as with both features combined, there was a significant interaction of mask with whether the target led or lagged the mask ( $F(3, 45) = 12.5, P < 0.001$ ). However, as is clear from the figure, this interaction is mainly due to the case when the target leads the mask (i.e. to the backward masking case). This was confirmed in analysis of variance which showed a highly significant effect of mask on a preceding target ( $F(3, 45) = 18.6, P < 0.001$ ), but only slight variation when the target follows the mask ( $F(3, 45) = 2.75, 0.1 > P > 0.05$ ). Thus for perception of place of articulation the amount of forward masking is virtually independent of the mask, but the amount of backward masking is much greater when the mask is another stop-vowel syllable than when it is one of the other masks ( $P < 0.001$ ). However, the three steady-state masks do show significantly greater backward than forward masking ( $P < 0.001$ ) although the amount of backward masking is very much less than for /g $\epsilon$ /.

For the extraction of voicing, however, there was no overall advantage for the lagging over the leading target ( $F < 1.0$ ) and only a slight interaction of lead/lag condition with mask ( $F(3, 45) = 2.46, 0.1 > P > 0.05$ ). Thus the perception of voicing shows no more backward than forward masking for the masks used here.

In summary this experiment shows that for stop-vowel syllables dichotically opposed by a mask at temporal offsets of  $\pm 60$  msec: (1) forward masking is roughly constant for the four masks used, for both place of articulation and voicing; (2) backward masking is greater than forward with place of articulation although not with voicing for all the masks; but (3) this difference is considerably greater when the mask is another stop-vowel syllable than when it is either the same vowel, a different vowel or a non-speech timbre; (4) these last three masks do not differ significantly in any condition.

### Discussion

The amount of backward masking, at least for the perception of place of articulation, is clearly dependent on the mask used. Dichotic masking is thus a potentially useful tool for describing features in auditory perception. The sharp discontinuity between the effects of the three steady-state masks and the  $/g\epsilon/$  mask argues against any general continuum of similarity being important, for if it were we might have expected the  $/\epsilon/$  mask to have been closer in its effect to the  $/g\epsilon/$  mask. Rather, we are led to suppose that the  $/g\epsilon/$  mask contains specific features which are particularly effective at interrupting the perception of the preceding target. This interpretation is strengthened by the absence of any mask specificity in the forward masking case, although this may be at least partly due to the very high performance leaving little room for improvement.

Two more points require discussion: the slight though consistently greater effect of backward over forward masking for the three steady-state masks, and the absence of any differences either between masks or between the forward and backward conditions for the perception of voicing. The first point may be attributable to some unspecific auditory effect or perhaps may not even be specific to the auditory modality. A kick on the shins may be an effective backward mask to this extent. The effect is quite small and will probably be difficult to investigate. The absence of any interesting effects in the perception of voicing may reflect the very different acoustic cues underlying the perception of voicing and of place of articulation. For voicing, at least in this experiment, the detection of some aspiration at the beginning of the stimulus would give sufficient information, whereas for place of articulation detailed knowledge of the slope of rapid formant transitions is required. The extraction of this latter information may be particularly sensitive to disruption.

This experiment alone cannot decide whether extraction of those acoustic parameters on which the decision concerning place of articulation is based is being disturbed, or rather whether it is some specifically linguistic process such as the relation of these acoustic features to a phonemic framework. To distinguish between these two hypotheses the next experiment looks at backward masking for stimuli which, like stop-vowel syllables, are distinguished by a rapidly changing initial section, but which are not perceived as falling into different phonemic categories.

### Experiment II

This experiment uses a paradigm introduced by KIRSTEIN (1970). No *a priori* distinction is made between target and mask, both being drawn from the same

stimulus set. The subject attends to one ear and is asked to recall the stimulus presented there.

### Method

Three different sounds were used. They differed in their fundamental frequency contours which are illustrated in Fig. 2. These pitch contours were carried on the steady vowel /e/. Dichotic pairs were made up using the Haskins parallel formant synthesizer and a special computer program (Mattingly, 1968) which ensured perfect timing of the signals on the tape's two tracks. On each trial a subject heard two pitch contours, one in either ear. They were either simultaneous or offset by  $\pm 25$  msec. Subjects attended to one ear for a block of trials and were instructed to identify only the sound that was presented to that ear. They were given training in identifying the three sounds with the first three digits. Half the pairs of sounds they heard were simultaneous and half were temporally offset. 12 right-handed subjects took the experiment in a procedure which counterbalanced ears and attention.

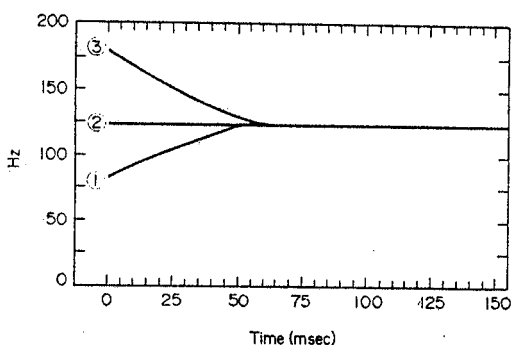


FIGURE 2. Pitch contours of the three sounds used in Experiment II.

### Results

The results are tabulated (Table I) in terms of the asynchrony of the reported stimulus. Thus, if the subject was presented with stimulus 1 to his left ear 25 msec ahead of stimulus 2 to his right and, though asked to report the left ear, in fact wrote "2", a correct response would be entered in the cell where the "right-ear" column under "lagging" and the row for "attend left" intersect. There was a clear advantage for the lagging over the leading condition irrespective of ear or attention condition (12 subjects for, none against). Subjects were

TABLE I

Mean per cents correct in Experiment II according to asynchrony of the reported stimulus

	Asynchrony of reported stimulus					
	Simultaneous		Leading		Lagging	
	Left	Right	Left	Right	Left	Right
Attend left	40.4	37.1	35.2	33.0	48.6	43.6
Attend right	36.6	40.4	29.5	37.8	40.9	48.6
Total	38.5	38.7	32.4	35.4	44.8	46.1

generally poor at selecting the requested ear though there was some indication that selective attention was easier for the staggered pairs than for the simultaneous ( $P < 0.1$ ). There was no difference between the ears in either the simultaneous or the staggered condition ( $P > 0.1$ ).

### *Discussion*

As in the first experiment we find here greater backward than forward dichotic masking for sounds which are distinguished by a rapidly changing initial portion. In the first experiment, rapidly changing formant transitions cued the place-of-articulation distinction in stops whereas in this experiment the sounds have been distinguished by changes in fundamental frequency which did not cue a phonemic distinction. Parsimony suggests that explanations for these effects should be sought at a purely auditory level of analysis rather than supposing that separate explanations are required for both speech and non-speech sounds.

Two subsidiary results of the second experiment call for comment: first, the slightly more efficient selective attention under staggered than under simultaneous conditions bears out a suggestion by Treisman and Riley (1969) to that effect. Second, the absence of any ear difference here contrasts in an interesting way with Haggard and Parkinson's (1971) finding that when rapid pitch changes similar to the ones used here cue the voiced-voiceless distinction in stops (Haggard, Ambler and Callow, 1970) there is an advantage for the right ear under dichotic presentation. The right ear advantage is thus determined by the use to which acoustic information is put rather than to the presence of particular acoustic features (Darwin 1971; Haggard and Parkinson, 1971).

Both experiments reported here have used dichotic presentation. As mentioned in the Introduction, the superiority of backward over forward masking for stop-vowel syllables by similar syllables is not found for monaural presentation (Porter, 1971). A related finding is that by Massaro (1970), who finds slightly larger dichotic than monotic backward masking for the identification of a pure tone followed by a longer masking tone. This generally greater backward masking under dichotic than under monotic conditions may reflect a segmentation problem faced by the auditory system. If the arrival of a particular auditory feature always interrupted the processing of any similar feature which had arrived just previously, the range of sounds which the auditory system could analyse successfully would be severely restricted. On the other hand, there may be good reasons why interruption should occur when the second sound forms part of a different perceptual unit from the first. Misplaced interruptions would be restricted to a minimum, at least in natural situations, if sounds from the same source were treated without interruption. Spatial location could provide a very reliable criterion for determining whether temporally distinct sounds originated from a common source. Spatial location has the added advantage that at least its directional aspect has neurophysiological correlates at a very peripheral level of the auditory system. This information is thus potentially available for guiding the sequential analysis of the auditory input at higher levels. This may not be the only criterion, and indeed Massaro's experiments with pure tones show appreciable monotic backward masking. But different processes may be operating for simple and complex

stimuli, since Massaro also finds backward masking for pure tones to be relatively independent of the similarity of test and masking tones.

This work was partially supported by a grant to Haskins Laboratories from the National Institute of Child Health and Development while the author was supported by a Fellowship from the Commonwealth Fund of New York.

### References

- BERLIN, C. I., WILLETT, M. E., THOMPSON, C., CULLEN, J. K. and LOWE, S. S. (1970). Voiceless versus voiced CV perception in dichotic and monotic listening. *Journal of the Acoustical Society of America*, **47**, 75(A).
- DARWIN, C. J. (1971). Ear differences in the recall of fricatives and vowels. *Quarterly Journal of Experimental Psychology*, **23**, 46-62.
- ELLIOTT, L. L. (1962). Backward masking: monotic and dichotic. *Journal of the Acoustical Society of America*, **34**, 1108-15.
- HAGGARD, M. P. and PARKINSON, A. M. (1971). Stimulus and task factors as determinants of ear advantages. *Quarterly Journal of Experimental Psychology*, **23**, 168-171.
- HAGGARD, M. P., AMBLER, S. and CALLOW, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America*, **47**, 613-7.
- KAHNEMAN, D. (1968). Method, findings and theory in studies of visual masking. *Psychological Bulletin*, **70**, 404-25.
- KOLERS, P. and ROSNER, B. S. (1960). On visual masking (metacontrast). *American Journal of Psychology*, **73**, 2-21.
- KIRSTEIN, E. F. (1970). Selective listening for temporally staggered dichotic CV syllables. *Journal of the Acoustical Society of America*, **48**, 95(A).
- KIRSTEIN, E. F. (1971). Temporal factors in perception of dichotically presented stop consonants and vowels. Ph.D. dissertation, Univ. of Connecticut.
- LOWE, S. S., CULLEN, J. K., THOMPSON, C., BERLIN, C. I., KIRKPATRICK, L. L., and RYAN, J. T. (1970). Dichotic and monotic simultaneous and time-staggered speech. *Journal of the Acoustical Society of America*, **47**, 76(A).
- MASSARO, D. M. (1970). Preperceptual auditory images. *Journal of Experimental Psychology*, **85**, 411-17.
- MATTINGLY, I. G. (1968). Experimental methods for speech synthesis by rule. *I.E.E.E. Transactions*, AU-16, 198-202.
- PORTER, R. J. (1971). The effect of temporal overlap on the perception of dichotically and monotically presented CV syllables. Paper at 81st Meeting of the Acoustical Society of America, Washington, D.C.
- SCHILLER, P. H. (1965). Monoptic and dichoptic visual masking by patterns and flashes. *Journal of Experimental Psychology*, **69**, 193-9.
- SPENCER, T. J. and SHUNTICH, R. (1970). Evidence for an interruption theory of backward masking. *Journal of Experimental Psychology*, **85**, 198-203.
- STUDDERT-KENNEDY, M., SHANKWEILER, D. P. and SCHULMAN, S. (1970). Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables. *Journal of the Acoustical Society of America*, **48**, 599-602.
- TREISMAN, A. M. and RILEY, J. G. A. (1969). Is selective attention selective perception or selective response? A further test. *Journal of Experimental Psychology*, **79**, 27-34.