

Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations

MONIQUE RADEAU and PAUL BERTELSON
Université libre de Bruxelles, B-1050 Brussels, Belgium

While "recalibration by pairing" is now generally held to be the main process responsible for adaptation to intermodal discordance, the conditions under which pairing of heteromodal data occur in spite of a discordance have not been studied systematically. The question has been explored in the case of auditory-visual discordance. Subjects pointed at auditory targets before and after exposure to auditory and visual data from sources 20° apart in azimuth, in conditions varying by (a) the degree of realism of the context, and (b) the synchronization between auditory and visual data. In Experiment 1, the exposure conditions combined the sound of a percussion instrument (bongos) with either the image on a video monitor of the hands of the player (semirealistic situation) or diffuse light modulated by the sound (nonrealistic situation). Experiment 2 featured a voice and either the image of the face of the speaker or light modulated by the voice, and in both situations either sound and image were exactly synchronous or the sound was made to lag by 0.35 sec. Desynchronization was found to reduce adaptation significantly, while degree of realism failed to produce an effect. Answers to a question asked at the end of the testing regarding the location of the sound source suggested that the apparent fusion of the auditory and visual data—the phenomenon called "ventriloquism"—was not affected by the conditions in the same way as adaptation. In Experiment 3, subjects were exposed to the experimental conditions of Experiment 2 and were asked to report their impressions of fusion by pressing a key. The results contribute to the suggestion that pairing of registered auditory and visual locations, the hypothetical process at the basis of recalibration, may be a different phenomenon from conscious fusion.

The extensive work on sensory rearrangement carried out in recent years has led to rather general agreement that one of the main causes of perceptual plasticity is a process of "recalibration by pairing" which is set up when "two variables of stimulation yield discrepant information about a distal property" (Epstein, 1975). The notion was first formulated by Wallach and Karsh (1963) for the situation of intramodal conflict between two depth cues. It was later extended to cases of intermodal discordance (Rock, 1966; Wallach, 1968) and of sensory-motor discordance (Howard, 1968).

The extension raised a problem which seems to have gone largely unnoticed. It concerns the criteria for pairing. How does the system decide whether two different values available in two separate modalities relate to a single object and not to different objects? The problem does not arise for intramodal discordance. In Wallach and Karsh's situation, for instance, there was no way in which the dynamic and stereophonic cues provided by their rotating

wire shape could be related to different objects. But take the familiar "terminal feedback" pointing condition used in many prism adaptation experiments. The subject extends his arm under a screen towards the visible target and, because he observes the screen through prisms, he sees his finger emerge at the other end of the screen some distance to the side of its felt position. Discordance exists only under the hypothesis that the visible finger is the subject's. Under the alternative hypothesis that it is another finger, there is no discordance.

Recalibration by pairing can only become a predictive principle if the rules of pairing are specified. By analogy with cases of intramodal patterning, we could expect pairing to be influenced by both *structural* and *cognitive* factors. By structural factors, we mean abstract properties of the sensory input like those cited in the Gestalt rules of figure segregation—proximity, similarity, common fate, etc.—which can be applied even to completely unfamiliar constellations of data. We use the term "cognitive" to designate features of the situation which communicate knowledge about the distal situation from which the data originate, and so indicate whether a unitary interpretation is correct or not. Such features can consist of verbal instructions or of a context more or less evocative of situations known to produce correlated sense data in the two modalities. The

The present work was carried out while the first author held a grant as "Chargé de Recherches" from the Belgian "Fonds national de la Recherche scientifique." The help of this university's "Centre universitaire du Film scientifique" (CUFS) in preparing the video recordings is gratefully acknowledged. The authors' address: Laboratoire de Psychologie expérimentale, 117, avenue Ad. Buyl, 1050 Bruxelles, Belgium.

structural-cognitive distinction is of course not always clear-cut, but it was felt to have sufficient heuristic value for an exploratory study.

An example of the effect of structural variables on recalibration is provided by the study by Held, Efsthathiou, and Greene (1966) which showed that delaying the visual feedback from the hand made adaptation impossible. Synchronization of visual and efferent (or proprioceptive) data is apparently important to achieve a correct correlation. The operation of cognitive determinants has, on the other hand, been well demonstrated by Welch (1972) in an experiment where he substituted his own finger for the subject's, in the manner considered in the last paragraph. The subject was exposed to an adaptation condition where he pointed under a screen towards a visible target, and the finger of the experimenter appeared at the other end of the screen 11° to the side of the location reached by his finger, which was made invisible, with the correct timing. Adaptive shifts were much larger in subjects who were told that what they saw was their own fingers than in others who were given a correct description of the situation. Explicit knowledge about the situation is thus an effective factor. On the other hand, the subjects who got accurate information nevertheless showed significant adaptation. This interesting finding may mean either that they did not all believe the story about finger substitution or that structural properties of the situation were sufficient to create an effective pairing part of the time, or in some of the subjects, in spite of contrary knowledge.

It is important to make the distinction between *pairing* and impression of identical origin, or *phenomenal fusion*. Adaptation can occur in cases where the subject does not report apparent fusion, and is thus aware of the discordance. In such cases, evidence of same origin is sufficient to produce pairing, hence adaptation, but not fusion. In other cases, adaptation occurs without the subject's becoming aware of the discordance. Pairing and fusion are thus two phenomena which both involve cross-modal matching, but are not necessarily coexistent. Although it is tempting to assume that they are influenced by the same factors and represent two different levels on a single continuum, there is practically no relevant evidence. Studies are needed where adaptation and fusion would be studied in parallel.

The present paper reports a first exploration of conditions for pairing in the particular case of auditory-visual discordance. The auditory and visual modalities are particularly well suited for this sort of study, because they offer wide possibilities of experimental manipulation. Kinesthetic or tactile inputs are, for instance, much more difficult to control.

Adaptation to auditory-visual discordance has been observed in conditions simulating well-known events yielding correlated visual and auditory data.

Canon's (1970) subjects, for instance, wore pseudophones and looked through prisms at a target which involved an illuminated panel and an electric bell, and they could see the clapper of the bell move at the time they heard the sound. Weerts and Thurlow (1971) used a still simpler situation where the subject saw a loudspeaker in one location while hearing sound delivered through a hidden loudspeaker in another location. In such situations, the context evokes a familiar pairing of visual and auditory data. But adaptation has also been obtained in the situation we used in our previous work, where the inputs were reduced to a single luminous point flashing in synchrony with sound pulses, and both when the discordance was produced by having the subject look at the light through a prismatic device (Radeau, 1973, 1974; Radeau & Bertelson, 1969, 1974) and by separation of the sources in actual space (Radeau & Bertelson, 1976). The fact that adaptation was observed in those situations show that contextual suggestion is not necessary for effective pairing to occur. The structure of the situation was apparently sufficient to bring it about. The situation was arranged under the reasonable (but so far untested) assumption that synchronization is the critical structural condition. On the other hand, even if contextual suggestion is not a prerequisite for adaptation to occur, it could still be a facilitating condition.

There is some evidence regarding the effect of structural and cognitive variables on reports of fusion of discordant auditory and visual data—the phenomenon which has come to be called *ventriloquism*, since it is believed to be one basis of the illusion created by the ventriloquist. Ventriloquism is apparently strongly dependent on synchronization of auditory and visual data. This has been shown both in a purified situation, involving only point flashes of light and sound pulses (Choe, Welch, Gilford, & Juola, 1975) and in a semirealistic situation involving the face of a speaker presented on a TV screen and his voice delivered through a hidden, displaced loudspeaker (Jack & Thurlow, 1973). Relevant also are the results obtained by the latter authors in some interesting exploratory experiments where they showed on their TV screen the face of a puppet and presented in the displaced loudspeaker a voice counting 1 to 10. The impression of ventriloquism, recorded by having the subject press a key whenever he experienced it, was much stronger when the head or the lower jaw of the puppet was made to simulate talking movements than when it was left motionless (Jack & Thurlow, 1973; Thurlow & Jack, 1973).

The evidence regarding cognitive influences is perhaps less straightforward. Jackson (1953) had subjects indicate the origin of a sound in the presence of more or less suggestive visual events. Showing steam coming out of one of several steam kettles while hearing the noise produced by a hidden whistle

produced acceptance of the visually suggested position in spite of much larger discrepancies than using an arbitrary pairing such as lights and bells. One problem with the kind of situation where subjects are asked to provide a judgment of location is that the contextual information may well influence the response directly, without necessarily creating or favoring an impression of fusion. In Jackson's case, knowledge of the principle of steam whistles may have lead the subjects to choose the steam-emitting whistle in spite of a contradictory perceptual impression. The difficulty is presumably avoided in the experiments reported by Jack and Thurlow (1973) and Thurlow and Jack (1973), where subjects were explicitly asked to report their impressions of fusion, not their beliefs regarding the external situation. Two sets of findings show the influence of cognitive factors. On the one hand, much less ventriloquism was reported in a situation where the face of a speaker counting was seen on the screen while synchronized tones were heard in the loudspeaker than in more realistic situations involving in one the same face and a voice counting and in the other the tones and the sight of a hand pushing a button. On the other hand, in the puppet situation, elimination of the eyes and nose of the puppet reduced the amount of ventriloquism. The size of the reduction however varied greatly across three series of experiments, and was significant in only one of them.¹

The present study was designed primarily to examine the influence on adaptation to auditory-visual discordance of semirealistic contexts simulating situations known to produce correlated auditory and visual stimuli. An easy way to manipulate the visual context was provided by a video recorder. Two main situations were used. One consisted of seeing on the screen hands playing the bongos while hearing the sound in a different place. The other situation featured the face of a speaker. These situations are called "semirealistic" because, while they involve familiar sequences of events, the presentation is through artificial devices. The context created in that way may be less convincing than a real-life situation, since the subjects are allowed to suspect many kinds of manipulations. In the present experiments, no particular care was applied to avoid such suspicion. In particular, the displaced loudspeaker, while not directly visible, was not really hidden.

The influence of the structural property which seemed most likely to affect pairing, synchronization of auditory and visual data, was also examined in the second experiment.

The design of the two first experiments consisted of having the subject point at auditory targets before and after a period of exposure to auditory-visual discordance. Independent groups of subjects were allotted to different conditions. This offered the advantage that new conditions could be added as a function of what was found.

Adaptation was studied in auditory localization only. In our previous work on adaptation to auditory-visual discordance (Radeau & Bertelson, 1969, 1974, 1976), evidence was found for recalibration of both visual and auditory localization. But auditory recalibration was generally both larger and faster, which made it a more convenient object for an exploratory study.

EXPERIMENT 1

The main purpose of this experiment was to examine if a semirealistic context, simulating a real-life situation, would increase adaptation to auditory-visual discordance above the level obtained in the more purified kind of situation used in our previous studies. The semirealistic situation that was chosen consisted of showing, on a TV screen, hands playing a percussion instrument and delivering the corresponding, synchronized sound through a displaced loudspeaker. A partially comparable nonrealistic situation was obtained by substituting light modulated by the sound for the image of the hands.

Method

Apparatus. The subject sat on an adjustable stool, in front of a large table (200 × 100 cm), covered with a black cloth. His head was held in a bite fitted with dental impression material, 32 cm above table top. A semicircular area, 45 cm in radius and centered on the bite, was bordered by a 1-cm-thick plastic rod. That rod served to limit the excursion of the hand during pointing movements. It was lined on the outer side by a slip of paper carrying a graduation in degrees. A small loudspeaker (Philips, AB 2070/Z4), 6 cm in diameter, fitted on a metal stand, could be moved by hand along the graduation. For the adaptation phase of the experiment, it was placed in a straight-ahead position under a black cloth cover, shaped in the form of a small tent 20 cm long and 10 cm high. During the same phase, a Sony TV monitor (PVM 76 CE) with a 14 × 11 cm screen was placed on the table, immediately behind the graduation, with its axis 20° either to the right or to the left of the median plane. Video signals and sound were played on a Sony Videocorder AV-3620 CE.

The whole session was run in dim illumination.

Materials. A talented amateur drummer was videotaped while improvising short solos on a pair of bongos. The camera was situated directly in front of the player, 20° above the horizontal, and recorded an area which, in the plane of the bongos, spanned about 70 × 55 cm and comprised the instruments and the hands of the player, but not his face. For pre- and posttests 15 5-sec pieces, separated by 10-sec silent intervals, were recorded (the video images were not used); for the adaptation phase of the experiment, 12 60-sec pieces, separated by 15-sec silent and black intervals, were recorded.

Subjects. Twenty volunteers, mostly university students, participated in two sessions each. They were paid (100 Belgian francs per session) for their services.

Procedure. A session consisted of three phases: (1) 15 auditory pointing trials (pretests), (2) 12 1-min periods of exposure to auditory-visual discordance (*adaptation phase*), and (3) 15 pointing tests (posttests) identical to the pretests.

During auditory pointing trials, the subject was blindfolded, using skin-diving goggles fitted with opaque material. The TV monitor and the cover of the loudspeaker were removed. Before each trial, the loudspeaker was positioned by the experimenter in

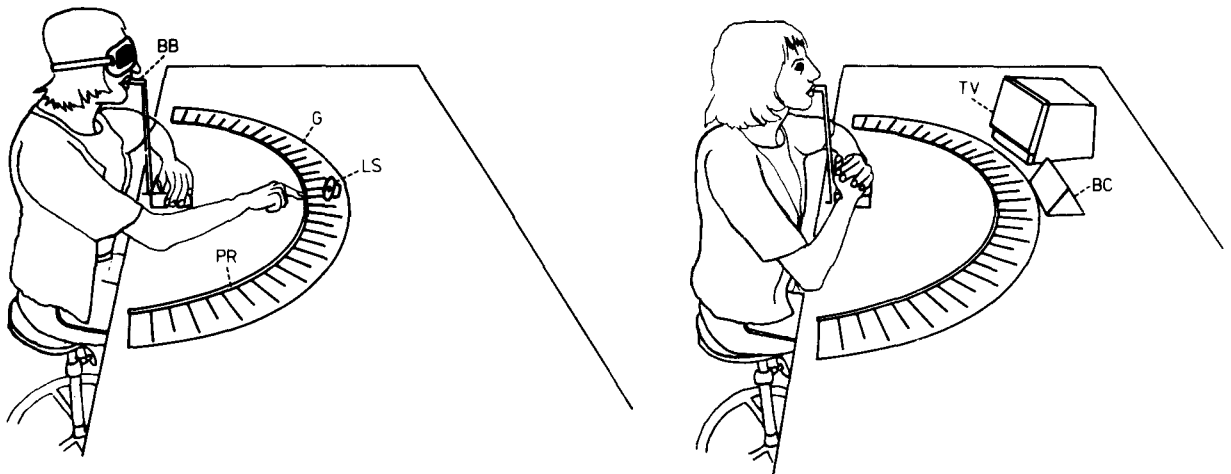


Figure 1. General view of the experimental situation for Experiments 1 and 2. Left: pointing tests (BB = biteboard, G = graduation, LS = movable loudspeaker, PR = plastic rod). Right: adaptation phase (BC = black cloth cover).

one of five positions, immediately behind the semicircular plastic rod, in the median plane or 3° or 6° to the right or to the left. The subject was instructed to indicate the direction from which the sound appeared to come by putting his right index finger against the plastic rod in the corresponding position. Between pointings, he brought his right hand back to a resting position on top of the left hand, which held the stand of the bite.

For the adaptation phase, the loudspeaker was placed in the median plane, under the black cover, and the TV monitor was centered on a position either 20° to the left or 20° to the right, immediately behind the plastic rod. The subject was invited to take the goggles off and to look at the screen. Two conditions were run. In condition B-H (for bongos-hands), both tracks of the video tape were used, so that the hands were seen on the screen while the corresponding sound came from the loudspeaker. In condition B-ML (for bongos-modulated light), only the sound track was used. It was played on the loudspeaker and simultaneously used as the source of video signals, to modulate the light on the screen. This produced a rather complicated pattern of changing horizontal lines, but then the screen was fitted with translucent tracing paper and the desired effect was obtained of blots of diffused light appearing at the rhythm of the bongos. The brightness was set at a level such that one flash corresponded to each stroke.

In both conditions, the subjects were given the task of monitoring the images and the sound of occasional discrepancies. It was explained that although the two generally corresponded, sounds had occasionally been added or deleted. The subjects were invited to say, after each 1-min observation period, whether or not they had detected an omission or an addition. No alteration was actually introduced, yet some were reported by all subjects in condition B-ML and even by 9 out of 10 subjects in the realistic condition B-H.

Posttests immediately followed the adaptation phase. Between the end of the last exposure period and the beginning of the first posttest, 15 sec elapsed, during which the subject put the goggles on and the experimenter arranged the setup for the pointing situation.

Ten subjects were run in each condition. Each subject did one session with the screen to the left of the loudspeaker and one with it to the right. The order of the two directions of discrepancy was balanced across subjects.

At the end of the second session, each subject was asked where the sound came from while they were looking at the TV.

Results

The aftereffect was measured as usual by subtracting mean error in posttests from mean error in pretests. Shifts were called positive when they went in the direction of the discordance, i.e., away from the median plane.

As shown in Table 1, significant aftereffects are observed in both conditions. They are slightly larger in condition B-ML, but the difference is nonsignificant ($t = .32$; $df = 18$; $p > .10$).

To the question regarding the origin of the sound, the majority of subjects in both conditions answered that it came from the monitor (Table 2). Those who mentioned a different origin pointed to the black cover, except for one subject (in condition B-H), who said the sound came from a direction at his back.

Discussion

The present results give no support to the notion that realism of the sensory context promotes adaptation. Seeing the image of the hands actually producing the sounds did not cause more adaptation than seeing synchronized flashes of light. Moreover, the aftereffects obtained in the present experiment are of comparable size to those obtained previously with

Table 1
Experiment 1—Overall Results

Condition	Mean Aftereffect	σ	p^*	N
B-H	3.04°	2.50°	.005	9
B-ML	3.49°	3.07°	.005	8

Note—N = number of subjects (in 10) showing adaptive aftereffect. B-H = bongos-hands. B-ML = bongos-modulated light.
*One-tailed t test.

Table 2
Experiments 1 and 2: Answers to Question About Direction
of the Sound During Adaptation Phase

Experiment	Condition	Number of subjects locating sound		
		in TV	in other place black cover	behind
1	B-M	7	2	1
1	B-ML	6	4	-
2	V-F	6	3	1
2	V-ML	5	5	-
2	DV-F	-	9	1
2	DV-ML	4	5	1
2	V-Vg	-	9	1

punctual flashes and sound pulses (Radeau & Bertelson, 1974, 1976).

EXPERIMENT 2

This experiment had two main objectives: (1) the first objective was to control, in another situation, the previous conclusion that realism of the sensory context is not an important condition of adaptation. The new semirealistic condition consisted of seeing the face of a speaker while hearing his voice, and a corresponding nonrealistic condition was obtained, exactly as in Experiment 1, by substituting flashes, modulated by the voice, for the face of the speaker. (2) The second objective was to examine the role of synchronization. This factor could not be studied in the bongo situation, because effective desynchronization proved hard to achieve with such highly periodic inputs. In the voice-face situation, acceptable desynchronization can be obtained by delaying the speech relative to the facial movements. The same manipulation was applied to the voice-modulated light situation so that the influence of asynchrony could be examined at the levels of both the realistic and the nonrealistic conditions. It must be noted that synchronization and realism cannot be conceived as two orthogonal factors, because desynchronization necessarily makes the voice-face situation unrealistic.

At this step of our exploration, it was felt necessary to control for a possible effect of simply paying attention to visual stimuli in one position while being exposed to auditory stimuli from a different location. Weerts and Thurlow (1971) found that simply fixating 20° to one side while listening to sounds coming from a centrally positioned hidden source produced small adaptive shifts in auditory localization. This effect did not reach significance, but it seemed, nevertheless, important to check for its occurrence in the present situation, where it might affect the size of the adaptation observed. A fifth condition was thus run where the subject listened to the voice coming from the frontal loudspeaker and

at the same time monitored a source of infrequent visual signals 20° to one side.

Method

Apparatus. The apparatus was the same as for Experiment 1.

Materials. A gentleman with very clear pronunciation (a professional phonetician) and wearing neither mustache nor beard was videotaped while reading passages of prose or poetry. The camera was situated in front of him, in the horizontal plane of his mouth, and recorded an area which at the distance of his face spanned about 21 × 16 cm and comprised the face up to the forehead. The recordings involved: (1) 15 four-verse pieces taken from La Fontaine's Fables, lasting about 5 sec each and separated by 10-sec silent intervals, to be used for pre- and posttests; (2) 12 60-sec pieces of prose, taken from several contemporary sources—magazines, novels, essays; the pieces were separated by 15-sec silent and blank intervals.

For use in the delayed-voice condition (DV-F) to be described later, one copy of the tape was made where the sound signal was relayed through the record and playback heads of a taperecorder (Revox A77) running at 9.5 cm/sec, which created a 0.35-sec delay. The video signal was copied directly, without an intermediate tape. To equalize as far as possible the effects of the copying operation on the quality of the recordings used in the different conditions, another copy was made, where the original tape was copied without an interpolated tape recorder, and was used in the synchronous conditions (V-F and V-ML). On the other hand, a tape-recorder copy of the sound track was prepared using the same Revox recorder running at 9.5 cm sec. The signal was copied directly on one track, and then transferred through the playback head of the channel onto the second track, which established a constant 0.35-sec lag between the two tracks. This tape-recorder copy was used in condition DV-ML.

Subjects. Fifty paid volunteers served for two sessions each.

Procedure. Exactly as in Experiment 1, each session involved 15 pretests, 12 1-min adaptation periods, and 15 posttests. The tests were identical to those of Experiment 1, except that the auditory targets consisted of speech instead of bongo solos.

Five different conditions were used during the adaptation phase, one condition for each group of 10 subjects. Conditions V-F (for voice-face) and V-ML (for voice-modulated light) were, respectively, the realistic and nonrealistic conditions corresponding to conditions B-H and B-ML of Experiment 1. In condition V-F, the face of the speaker appeared on the screen and his voice was heard synchronously in the frontal hidden loudspeaker. In condition V-ML, the light on the screen was modulated by the voice itself. Exactly as in condition B-ML of Experiment 1, the screen was covered with tracing paper to eliminate high-frequency details. The brightness was set at a level such that a distinct flash corresponded to about each syllable. The same discrepancy-monitoring task as in Experiment 1 was given in both conditions. Written instructions were given explaining that auditory and visual events corresponded most of the time, but that sounds had occasionally been added or deleted, and that the task was to say at the end of each 1-min period if one had noticed additions, deletions, or nothing. All subjects, except one in condition V-ML, reported some alterations.

In conditions DV-F (delayed voice-face) and DV-ML (delayed voice-modulated light), the auditory and visual inputs were identical to those of conditions V-F and V-ML, respectively, but the sound was delayed 0.35 sec relative to the visual events. So, in condition DV-F, the subject saw the face of the speaker and heard his delayed voice in the loudspeaker. In condition DV-ML, the subject saw light modulated by the voice, but heard that voice with the 0.35-sec delay in the loudspeaker. Condition DV-F was produced using the videocord and the delayed-voice video tape. Condition DV-ML was produced with the tape recorder and the tape with the two desynchronized recordings of the voice. The

videocorder was nevertheless kept in its usual position and was run during both tests and adaptation phase to equalize as far as possible the acoustic surroundings in the different conditions.

The discrepancy-monitoring task used in the synchronous conditions could not be adapted to the two desynchronized conditions. Another task was substituted which also required focusing of attention on the correspondence between visual and auditory events. It was explained (in written instructions again) that, during some 1-min periods, movements and sounds (in cond. DV-F) or lights and sounds (in cond. DV-ML) had simply been desynchronized, while for other periods, unrelated sound and light episodes were being played. The task was to classify each period in terms of that dichotomy. Only two subjects, one in each condition, opted for desynchronization throughout. Besides, one subject in group DV-ML insisted on three occasions that the events were actually synchronized.

Condition V-Vg (voice-vigilance) was the control condition. A small lamp (LED Litronix RL55) mounted on a black chassis was substituted for the monitor, i.e., was placed either 20° to the right or to the left of the median plane, while the voice was played in the loudspeaker. The lamp was lit (for ± 1 sec) 0 to 3 times, at irregular intervals, during each period, and the subject was given the task of reporting at the end of the period, the number of flashes that had occurred.

Each subject did one session with the visual display 20° to the left of the median and one with the opposite arrangement, half the subjects in each group of 10 beginning with one of the two arrangements.

At the completion of the second session, each subject was asked the same question as in Experiment 1 regarding the apparent origin of the sound.

Results

In the control condition, V-Vg, only a negligible 0.06° aftereffect occurred, which of course is non-significant. In all other conditions, significant aftereffects were obtained, which went in the direction of the TV monitor. As appears in Table 3, every subject in all four conditions shows an adaptive mean aftereffect.

The mean individual aftereffects for the four experimental conditions were submitted to a two-way analysis of variance. It yielded a significant effect of synchronization ($F = 6.08$; $df = 1,36$; $p = .025$). The effect of realism is nonsignificant ($F < 1$) as is the Realism by Synchronization interaction ($F = 1.21$; $df = 1,36$).

The responses to the question regarding the direction of the sound source appear in Table 2. In conditions V-F and V-ML, more than half the subjects report a unitary interpretation, a result similar

to that obtained in Experiment 1. No such responses are obtained in condition DV-F, just as in the control condition V-Vg. But four subjects of group DV-ML report an unitary impression. Had it been predicted, the difference in the distribution of responses between groups DV-F and DV-ML would be significant at $p = .05$ by Fisher's exact test (Siegel, 1956).

Discussion

Aftereffects. The fact that no aftereffects occurred in the control condition, V-Vg, rules out the possibility that those observed in the experimental conditions were due to the simple fact of focusing auditory and visual attention in different directions.

As in the first experiment, realism failed to affect adaptation significantly: The apparent position of the voice was not shifted more by exposure to the image of the face of the speaker than to modulated light.

On the other hand, desynchronization of visual and auditory inputs reduced the effects of spatial discordance strongly in both the semirealistic and the nonrealistic situations.

One may ask why significant aftereffects were observed in the two desynchronized conditions. The reason is presumably that the temporal shifting procedure that was used to produce these conditions still allowed many chance coincidences between auditory and visual events, so that synchronization was still experienced part of the time. We have already mentioned that with the highly periodic bongo stimulation, desynchronization was practically impossible to obtain.

The structural requirements to achieve effective pairing are probably not very stringent. This is suggested not only by the occurrence of adaptation in our desynchronized conditions, but also by the fact that with synchronization, amount of adaptation was not affected by the substitution of modulated light for the hands (Experiment 1) or the face (Experiment 2). This substitution, besides eliminating familiarity, reduces the amount of visual information that can be correlated with the auditory data. The sight of the hand movements gives advance information of the strokes which is not available with the modulated light. In the case of speech, on the other hand, facial movements can be correlated with specific parts of the auditory stream, while light flashes can be correlated with any maxima of sound energy.

Relation to detection of discordance. The answers to the question about the origin of the sound give some information regarding the occurrence of ventriloquism during the adaptation phase. It is recognized, of course, that these answers, requested only at the end of the second of two sessions, may be a rather unreliable reflection of what happened during the sessions.

Those responses are not affected by the conditions

Table 3
Experiment 2: Overall Results

Condition	Mean Aftereffect	σ	p^*	N
V-F	3.45°	.92°	.001	10
V-ML	2.54°	2.48°	.01	10
DV-F	1.70°	1.14°	.005	10
DV-ML	1.87°	1.99°	.025	10
V-Vg	.06°	1.41°	n.s.	5

Note—N = number of subjects (in 10) showing adaptive aftereffects.

*One-tailed *t* test.

in exactly the same way as aftereffects, since here the two asynchronous conditions do not give the same results: Reports of fusions are completely absent with face and asynchronous voice, but occur with modulated light and asynchronous voice. It is as if the poorer information content of modulated light influenced conscious detection, leading to less systematic detection of the discrepancy with asynchronous inputs, but not pairing, since aftereffects are not increased.

Another way of looking at the relation between adaptation and awareness of discordance consists of examining separately the mean aftereffect of those subjects who reported and of those who did not report the discordance. This separation is performed in Table 4 for the results of the two experiments. One can first notice that for all the conditions where the two answers were obtained, those subjects who reported the discordance showed smaller aftereffects than did those who did not report it. There is thus some correlation between adaptation and ventriloquism. But the whole effect of our conditions on adaptation cannot be mediated by its effect on ventriloquism. First, in all experimental conditions, the subjects who report the discrepancy nevertheless show clear aftereffects. Second, the performance of these subjects is still affected by the particular conditions they were exposed to.

EXPERIMENT 3

The relationship between aftereffects and the judgments regarding the direction of the sound source formulated by the subjects of the two preceding experiments suggests that ventriloquism and adaptation may not be affected by our experimental treatments in exactly the same way. On the other hand, we have seen in the introduction that there are data suggesting that ventriloquism may be facilitated by a realistic context. For these reasons, it was felt useful to obtain reports of perceptual fusion, in our various conditions, in a more systematic way than by the single question asked so far.

We, therefore, chose to ask subjects for intro-

Table 4

Experiments 1 and 2: Mean Aftereffect as a Function of the Answer to the Question Regarding the Direction of the Sound Source During the Adaptation Phase

Experiment	Condition	Subjects locating sound	
		in TV	elsewhere
1	B-H	3.48	1.54
1	B-ML	3.98	2.75
2	V-F	3.47	3.42
2	V-ML	2.72	2.36
2	DV-F		1.70
2	DV-ML	1.94	1.83
2	V-Vg		.06

spective reports. Although we appreciated the difficulties associated with this approach, we preferred it to the one consisting of requesting judgments of same vs. different origin. Choe et al. (1975) have correctly noted that such judgments can be affected by response biases. The provision of a suggestive context, for instance, could well increase judgments of same origin directly, without actual perception of fusion. This argument has already been mentioned à propos Jackson's (1953) experiment. The solution advocated by Choe et al. of applying detection measures has, however, been found inadequate (Bertelson & Radeau, 1976). Asking subjects explicitly to report their impressions, not their beliefs, and doing our best to secure their cooperation, seemed the most valid way of approaching fusion.

The subjects were exposed to the four experimental conditions of Experiment 2, and were asked to press a key whenever they experienced fusion. The setup had to be modified to increase uncertainty regarding sound origin. With the loudspeaker constantly in the middle, and covered with the rather obvious black tent, the fusion impression became extremely fragile. As a matter of fact, we observed, when trying the task on ourselves, that fusion tended to vanish as soon as we tried to decide whether to press the key or not.

Method

Apparatus and Materials. The bite was removed. The movable loudspeaker and its cover were replaced by three more effectively hidden units. Each loudspeaker was fixated to the inferior face of the table panel, at the bottom of a circular hole, 2 cm in diameter. The cloth covering the table top made the holes invisible. They were situated immediately behind the graduation, one in the median plane and one 20° to the left and to the right. The experimenter chose the loudspeaker to be used during each episode by means of a three-way switch. The subject reported his impression of fusion by pressing a two-way lever, mounted horizontally in the median position. Keypress duration was recorded on a paper-tape recorder and at the same time accumulated on a millisecond timer.

The conditions involving the face of the speaker (V-F and DV-F) were obtained in the same way as in Experiment 2, with the videorecorder. A tape was prepared with 21 1-min episodes: one episode with synchronous voice and then alternately a pair of episodes with delayed voice and a pair with synchronous voice.

The conditions with modulated light (V-ML and DV-ML) were obtained using the tape recorder. A tape was prepared with the same 21 1-min sound episodes as on the videotape. A pair of episodes with the sound recorded synchronously on both tracks were followed with a pair with sound delayed 0.35 sec on Track 2 relative to Track 1. Track 1 was used as a source of video signals to control the light on the screen, and Track 2 to control the sound.

Subjects. Twelve paid volunteers served for one session each.

Procedure. The subject was exposed to 36 1-min episodes identical to those of the experimental conditions of Experiment 2.

He was given a written description of the task, with the following critical passage (in French, of course): "We are interested in the evolution of your impression regarding the relative positions of sound and image. When watching TV, one nearly always has the impression that the sound comes from the image, and even that it moves with the image, although the loudspeaker is

generally located on one side of the set. In the present situation, you are asked to look at the screen, to analyze your impression, and to express it as fast and accurately as possible with the key in front of you. You may feel that the sound comes or does not come from the same place as the image, and this feeling may change in time. During each episode, your task is to press the key every time you have the impression that sound and image come from the same place, to keep it down as long as that impression lasts, and to release it every time the impression stops."

The experimenter checked, by asking a few questions, that the subject had grasped the distinction between reporting his impression and trying to detect the discrepancy. One subject was discarded because his responses implied that the difference escaped him.

The session consisted of two halves, one devoted to the two conditions with the face of the speaker, the other to the two conditions with modulated light. During each half session, the subject was exposed to 3 practice episodes and 16 experimental ones. Pairs of episodes with synchronous voice were alternated with pairs with delayed voice. Half the subjects started with the synchronous condition, and the other half with the asynchronous condition. The different orders of conditions were obtained by starting the tape either on the first or on the third episode. One pair of episodes comprised one episode with the sound to the left of the monitor and one with the sound to the right. This systematic alternation of directions of discordance was adopted to avoid adaptation. Half the subjects always had the sound first to the left then to the right of the monitor, and the other half had the other order. Each direction of discordance was achieved for half the episodes by placing the monitor in the middle and using a lateral loudspeaker; for the other half, it was the other way round. The experimenter moved the monitor and switched on another loudspeaker between episodes.

Results

All subjects reported some fusion under all conditions, except for one subject who reported none under condition DV-F. Mean duration of reported fusion, in percent of total exposure time, is given for each condition in Table 5. A two-way analysis of variance, with subjects and conditions as factors, was applied to mean duration of fusion per subject and condition. But subjects ($F = 4.4$; $df = 11,33$; $p < .005$) and conditions ($F = 26.4$; $df = 3,33$; $p < .005$) were found to be highly significant sources of variation. Subsequent application of Duncan's multiple range test showed that, at $p = .01$, all differences between conditions are significant, except that between V-ML and DV-ML.

Detailed examination of the recordings showed that once they reported fusion, subjects tended to go on till the end of the episode. This all-or-none aspect of the phenomenon is apparent, for instance, when

10-sec blocks of recording are categorized as a function of the time spent responding: 0% and 100% responding form the vast majority, as is shown in Table 5. Subjects tend, of course, to start responding earlier in condition V-F than in the other ones. The influence of time during the session was also examined. The same general pattern is present throughout the session, but the differences between conditions increase with time.

Discussion

The effect of our conditions on reports of fusion gives a rather different pattern from their effect on adaptation.

Desynchronization which affected adaptation both in the conditions with the face of the speaker and in those with modulated light, reduces fusion considerably in the conditions with the face, but has no significant effect in those with modulated light. A parallel observation had been made in Experiment 2 at the level of the judgments of sound location: desynchronization eliminated judgments of same origin in condition DV-F, but not in condition DV-ML.

For realism, the substitution of modulated light for the face of the speaker reduces fusion in the synchronized conditions. No corresponding effect was observed at the level of aftereffects. In the desynchronized conditions, the face of the speaker is actually detrimental and reduces fusion below the level observed with modulated light. The most likely explanation for this effect is the one considered in the discussion of Experiment 2. The richer visual information provided by the facial movements, as compared with the modulated light, could give more opportunities to detect temporal discordances with the voice. This assumption, of course, makes information content, not realism, the responsible factor. Another nonrealistic condition, providing temporal information comparable to that provided by the face, but lacking the realistic features, would be necessary to know what the relevant factor is.

Although the present experiment was addressed primarily to the question of the correlation between ventriloquism and adaptation, it is worth noting that the results confirm the investigations examined in the introduction, where ventriloquism was found to be affected by both synchronization and the realism of the situation.

Table 5
Experiment 3

Condition	Mean Duration of Fusion in %	σ	10-sec Blocks with Percent Fusion		
			0	Intermediate	100
V-F	77.5	16.6	18.2	15.8	66.0
DV-F	21.5	21.4	74.6	9.5	15.9
V-ML	48.6	18.9	46.0	14.1	39.9
DV-ML	41.7	24.5	51.7	15.2	33.1

GENERAL DISCUSSION

The experiments allow two main conclusions.

(1) In the case of auditory-visual discordance, recalibration is affected mainly by the structural condition of synchronization. The degree of realism of the situation was, in the present situation, irrelevant.

Synchronization as a condition of sensory patterning is, of course, reminiscent of the Gestalt law of common fate. Pairing, the hypothetical operation at the basis of recalibration, would involve a mechanism which correlates inputs to different modalities and looks for contingencies. Such a view is in line with suggestions made long ago by Held (Held & Freedman, 1963) for the case of visuomotor discordance.

The negative result regarding realism should be interpreted with caution. It cannot be taken as implying that pairing cannot be affected by cognitive variables. The results of Welch, showing that knowledge of the source of discordance influences adaptation to visuoproprioceptive discordance, have been discussed earlier, and we, ourselves, have found in a previous study that verbal explanations regarding the origin of conflicting auditory and visual data affected recalibration of both modalities, although in a way which allowed no simple interpretation (Radeau & Bertelson, 1974). The conclusion to be derived from the present result is thus essentially that actual presentation of a realistic context is not important. On the other hand, it is still possible that the factor would become effective in combination with a stronger suggestion of common origin. As was noted in the introduction, no attempt at making such a suggestion was made here. This is due to the fact that our interest in the effect of realism was partly a practical one: We wanted to know if adaptation could not be brought about more easily in future experiments just by simulating a familiar situation. From that practical point of view, the effect was only of interest if it did not depend on particular instructions. The effect of a realistic context should be examined in conjunction with other cognitive determinants before a stronger conclusion can be reached.

(2) Recalibration and fusion were not affected in the same way by our manipulations. This result also should be interpreted cautiously. Fusion and adaptation have been measured separately, under instructions which probably tended to induce different cognitive sets. The difference might thus simply reflect an interaction between cognitive set and our experimental variables. But a dissociation between adaptation and ventriloquism was suggested by the answers to the question regarding sound origin in Experiment 2. In particular, the difference between conditions DV-F and DV-ML was observed

in these answers. We thus have rather serious reasons to suspect that pairing of registered locations and conscious fusion are not two successive states on a continuum of increasing association between visual and auditory data, which would of course have been the most economical view.

REFERENCES

- BERTELSON, P., & RADEAU, M. Ventriloquism, sensory interaction, and response bias: Remarks on the paper by Choe, Welch, Gilford, and Juola. *Perception & Psychophysics*, 1976, 19, 531-535.
- CANON, L. K. Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *Journal of Experimental Psychology*, 1970, 84, 141-147.
- CHOE, C. S., WELCH, R. B., GILFORD, R. M., & JUOLA, J. F. The "ventriloquist effect": Visual dominance or response bias? *Perception & Psychophysics*, 1975, 18, 55-60.
- EPSTEIN, W. Recalibration by pairing: A process of perceptual learning. *Perception*, 1975, 4, 59-72.
- HELD, R., & FREEDMAN, S. J. Plasticity in human sensorimotor control. *Science*, 1963, 142, 455-461.
- HELD, R., EFSTATHIOU, A., & GREENE, M. Adaptation to displaced and delayed visual feedback from the hand. *Journal of Experimental Psychology*, 1966, 72, 887-891.
- HOWARD, I. P. Displacing the optical array. In S. J. Freedman (Ed.), *The neuropsychology of spatially oriented behavior*. Homewood, Ill: Dorsey Press, 1968.
- JACK, C. E., & THURLOW, W. R. Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Perceptual and Motor Skills*, 1973, 37, 967-979.
- JACKSON, C. V. Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology*, 1953, 5, 52-65.
- RADEAU, M. The locus of adaptation to auditory-visual conflict. *Perception*, 1973, 2, 327-332.
- RADEAU, M. Adaptation au déplacement prismatique sur la base d'une discordance entre la vision et l'audition. *Année Psychologique*, 1974, 74, 23-34.
- RADEAU, M., & BERTELSON, P. Adaptation à un déplacement prismatique sur la base de stimulations exafférentes en conflit. *Psychologica Belgica*, 1969, 9, 133-140.
- RADEAU, M., & BERTELSON, P. The aftereffects of ventriloquism. *Quarterly Journal of Experimental Psychology*, 1974, 26, 63-71.
- RADEAU, M., & BERTELSON, P. The effect of a textured visual field on modality dominance in a ventriloquism situation. *Perception & Psychophysics*, 1976, 20, 227-235.
- ROCK, I. *The nature of perceptual adaptation*. New York: Basic Books, 1966.
- SIEGEL, S. *Non-parametric statistics*. New York: McGraw-Hill, 1956.
- THURLOW, W. R. G., & JACK, C. E. Certain determinants of the "ventriloquism effect." *Perceptual and Motor Skills*, 1973, 36, 1171-1184.
- WALLACH, H. Informational discrepancy as a basis of perceptual adaptation. In S. J. Freedman (Ed.), *The neuropsychology of spatially oriented behavior*. Homewood, Ill: Dorsey Press, 1968.
- WALLACH, H., & KARSH, E. B. The modification of stereoscopic depth perception and the kinetic depth effect. *American Journal of Psychology*, 1963, 76, 429-435.
- WEERTS, T. C., & THURLOW, W. R. The effect of eye position and expectation on sound localization. *Perception & Psychophysics*, 1971, 9, 35-39.
- WELCH, R. B. The effect of experienced limb identity upon adaptation to simulated displacement of the visual field. *Perception & Psychophysics*, 1972, 12, 453-456.

NOTE

1. In all these experiments, the degree of synchronization of visual and auditory events depended on the skill of the experimenters at the time they prepared their recordings. This may have introduced unwanted differences between the different conditions which might account for the large scatter of the results obtained in the puppet situation. The same factor might have contributed to the difference in duration of fusion between the face-tones condition on one side and the face-voice and hand-

tones conditions on the other. The authors synchronized the sounds after the video recording had been made, "to equate, as far as possible, timing of visual and auditory stimuli in the different experiments." But this precaution does not eliminate the possibility that achieving an exact synchronization was perhaps easier in the realistic conditions.

(Received for publication January 24, 1977;
revision accepted May 4, 1977.)