The Cue Summation Theory Tested with Meaningful Verbal Information

Donald R. Cushman

No previous studies pertaining to Hartman's Cue Summation Theory ("Redundant information simultaneously presented by the audio and the visual channels results in greater efficiency in learning than does the same information in either channel alone.") could be found which utilized meaningful verbal information as stimulus material. Students were presented material fitting this criterion and were tested for affective responses and cognitive recall. Comparisons of results were made between three separate presentation procedures: audio alone, visual alone, and combined audiovisual. Comparisons of affective responses did not indicate significant differences, but analyses of cognitive data showed the combined audiovisual treatment resulting in half the number of errors as either the audio or the visual treatments alone.

As new technological developments are introduced into the instructional milieu, students are exposed to an ever-widening range of mediated instruction involving sight and sound. It becomes important, therefore, that more be known about the technology, about its utilization, and especially about the sensory channels involved in communication.

Hartman's research (1961) dealt specifically with learning in the aural and visual channels. He posited that students could learn better if cues or "bits" of information were presented to them simultaneously in both channels rather than in either the aural or visual channels separately. From these ideas and researches he formulated the Cue Summation Theory, which states that redundant information simultaneously presented in the audio and the visual channels results in greater efficiency in learning than does the same information presented in either the audio or the visual channel.

Visible Language, VII 3 (Summer 1973), 247–260.

Author's address: R. R. #4, Box 300, Mooresville, IN 46158.

Definition of Terms

Hartman's theory explicitly identifies the audio and the visual communication channels. *Channels* in this connotation means the sensory vehicle through which information is received by an individual. That is to say, the eyes receive information in the visual channel, while the ears are the receptors for the aural channel.

Communication through these channels involves different modalities, i.e., the verbal and the non-verbal. These terms were more specifically labeled by Knowlton (1964) as *digital* and *iconic*. Digital refers to any verbal information, either spoken or printed; iconic information is a non-verbal representation of an object. Thus, the picture of a cow or the sound "moo" would represent a cow in the iconic mode, and the representations would be in the visual and the aural channels respectively. The printed word "c-o-w" (visual channel) and the spoken word "cow" (aural channel) are both classed in the digital mode. Conway (1968) grouped this information for research as in Figure 1.

		Channels	
		Audio	*Visual*
Modes	Iconic	Sound "moo"	Picture of a cow
	Digital	Spoken word "cow"	Printed word "c-o-w"

Figure 1. Classification of Channels and Modes

Where *learning* is referenced in this study, it will be considered in both the cognitive and the affective domains. In the cognitive sense it will refer to subjects' ability to respond correctly to a test of the specific information contained in the stimulus material. Learning will also be studied in the affective domain by determining the existence of and the degree to which the information creates an emotional response in each subject.

The efficiency referred to by Hartman implies the maximum information gain with the least time involved in presenting the information.

Theory and Statement of the Problem

The Cue Summation Theory refers specifically to the presentation of redundant information. The use of the word "redundant" gives rise to arguments for and against the existence of a condition of redundancy in different channels. It can be argued that separate channels or modalities cannot convey redundant information. If, for example, a person reads the word "cat" and hears someone else say the word "cat," he may be receiving the same contrived verbal message, but some particular inflection or other vocal treatment in the voice of the person speaking the word may supersede redundancy and add connotative meaning that may not exist in the sterile word "cat." Thus, there is a summation of both similar and dissimilar cues.

On the other hand, it can be argued that if the second channel adds no new cues to the first channel, there can be no summation. Those arguing in support of this point contend that this redundancy does not transmit additional information to the receiver.

Since its proposal by Hartman, the Cue Summation Theory has been challenged, tested, supported, and refuted by various researchers with varying combinations of channels and modalities. Support for it has come from Loveless, Brebner, and Hamilton (1970) who found that bi-sensory presentation results in a higher detection rate than is obtained from presentations in either unisensory channel. Van-Mondfrans and Travers (1964) partially supported the theory by finding that the audio channel was inferior to the visual channel or a combined presentation, but they did not find significant differences between the visual and the combined presentations.

Refutations of the Cue Summation Theory have come from several sources. Some refutations are the result of failures to support hypotheses related to the theory, while others come from studies closely related to Hartman's work with opposing conclusions. Notable among researchers who did not support the theory is Severin (1967) who found only a limited application for the theory. He states that the theory applies only when there is an actual summation of cues. Conway (1968) is another researcher who does not support the Cue Summation Theory on the basis of his findings.

A review of the Cue Summation Theory research indicates no instance in which the researchers used meaningful stimulus material in their studies. Typical stimuli have been nonsense syllables, digits,

and simple isolated words. The dependent variables have all been cognitive measures, which tested recall, recognition, and detection of some stimulus. An examination of the literature has pointed to a great inconsistency in the types of measures or the design of the different studies. The lack of standardization in the types of stimulus materials and in the measures and dependent variables emphasizes the inequalities in these studies. A lack of immediate relevance and basis for generalization to an instructional situation is another characteristic of some previous studies.

Previous research indicates there is a need to test the theory with the use of meaningful stimulus material. This would facilitate testing in both the cognitive and the affective domains. It would also provide data that may offer tentative general guidance for instructional settings.

This researcher found no reference to affective measuring in any of the previous studies. The possibility exists, it was reasoned, that some of the differences which may have occurred in the subjects' learning in those experiments may have occurred in the affective domain. Therefore, measurement of an affective variable was introduced.

Questions arising out of classroom experiences and problems generated by research in education have served as the basis for developing the two hypotheses posed for this study. These hypotheses are as follows:

Hypothesis 1: Subjects presented meaningful verbal information in a combined audiovisual presentation will respond more accurately to questions regarding the content than subjects receiving the same information in either the audio or the visual channel alone.

Hypothesis 2: Subjects will display greater affective responses for information received through the combined audiovisual presentation than will subjects who receive the same information through the individual audio and visual channels alone.

METHODOLOGY

Introduction

The purpose of this study was to determine whether or not significant differences in learning would occur in subjects exposed to the same

meaningful verbal stimulus material via different combinations of audio and visual treatments. The stimulus materials were selected to provide as near equivalence as possible between the separate selections that were presented to the subjects in the various treatments. The three treatments involved were reading (visual), listening (aural), and reading and listening simultaneously (audiovisual).

In the design for presentations subjects were selected randomly to receive the stimulus materials in one of nine combinations and sequences.

Two sets of tests, one cognitive and the other affective, were given at the conclusion of each presentation. Comparisons were based on (1) the number of errors to questions in the cognitive portions of the tests, and (2) the measures of scores on the semantic differential scales which quantified the affective responses.

The Stimulus Material

The main consideration used in selecting the stimulus material was that it be meaningful to the subjects. It was decided to use a poem that was appropriate for the junior high school student population. Some of the factors involved in selecting the particular poem were as follows:

First, poetry is typically written with a high affective content. Since affective learning was to be a dependent variable in this study, a stimulus was needed that would have sufficient affective content to provide a response that could be measured.

Second, since each subject was to be exposed to three separate treatments of the material, it was necessary to have a near equivalence in the content of the three separate presentations. Thus, differences in learning, if any existed, could be attributed to the presentation treatment and not to varying content differences in the stimulus material.

Third, poetry is written (in a traditional sense) with regular meter and consistent line or stanza length, and this regularity was desired for control. A six-stanza poem, "The Pardon," provided this constancy and could be divided into three pairs of stanzas (Fig. 2). The poem also provided an equivalence of content of material as explained above.

Exposure time is an important factor in the comprehension of verbal materials. For example, a person who slows down his reading

THE PARDON

My dog lay dead five days without a grave
In the thick of summer, hid in a clump of pine
And a jungle of grass and honeysuckle-vine.
I who had loved him while he kept alive

Went only close enough to where he was
To sniff the heavy honeysuckle-smell
Twined with another odor heavier still
And hear the flies' intolerable buzz.

Well, I was ten and very much afraid.
In my kind world the dead were out of range
And I could not forgive the sad or strange
In beast or man. My father took the spade

And buried him. Last night I saw the grass
Slowly divide (it was the same scene
But now it glowed a fierce and mortal green)
And saw the dog emerging.

I confess
I felt afraid again, but still he came
In the carnal sun, clothed in a hymn of flies,
And death was breeding in his lively eyes.
I started in to cry and call his name,

Asking forgiveness of his tongueless head.
. . . I dreamt the past was never past redeeming:
But whether this was false or honest dreaming
I beg death's pardon now. And mourn the dead.

Figure 2. The stimulus material.

pace considerably in order to try to comprehend more of what he reads is distorting the measure by his increased exposure time (VanMondfrans and Travers, 1964). In an attempt to keep the exposure times as consistent as possible, a pilot study was run to determine the approximate amount of time required for a subject to read one pair of stanzas. The times required by each subject in the pilot study were quite consistent and the range of times was narrow.

The total time of the narrative tape recording was matched with the time required by the subjects to read the stimulus.

Process and Measures of the Experiment

Two types of measures were taken of the subjects in this study, cognitive and affective. The methodology for obtaining these measures was as follows:

Each subject was tested individually in one of nine possible treatment combinations. Before the tests began, each subject was instructed in the use of the semantic differential scale and was given a practice exercise. The practice exercise involved a different poem from that used in the actual test. Following the exposure to the sample poem each subject filled out the semantic differential scale. This practice exercise served to familiarize the student with the terms in the scale. It also afforded practice in completing the scale and helped make the subject feel at ease before the actual test began.

After the first set of two stanzas from the poem had been presented and the subject had been tested over it, the process was repeated a second and a third time to complete the presentation and testing of the six-stanza poem. At the conclusion of the six-stanza presentation and testing each subject was asked additional questions. The twenty-four subjects who received all three different treatments were asked to state their preferences for the types of treatments to which they had

GOOD	—:—:—:—:—:—:—	BAD
BEAUTIFUL	—:—:—:—:—:—:—	UGLY
SWEET	—:—:—:—:—:—:—	SOUR
KIND	—:—:—:—:—:—:—	CRUEL
PLEASANT	—:—:—:—:—:—:—	UNPLEASANT
BITTER	—:—:—:—:—:—:—	SWEET
HAPPY	—:—:—:—:—:—:—	SAD
SACRED	—:—:—:—:—:—:—	PROFANE
NICE	—:—:—:—:—:—:—	AWFUL
FRAGRANT	—:—:—:—:—:—:—	FOUL

Figure 3. The semantic differential scale.

been exposed. The purpose of this question and response was to provide data for a correlation of student performance to preference.

Sixteen subjects preferred the combined treatment over the audio and the visual treatments alone. Six placed the audio alone treatment first, and two rated the visual alone treatment as their preference.

The Pearson Product-Moment Correlation was applied to each set of choices, correlating the preferences with the actual performances of the subjects. For first presentations the correlation of preference with performance was +.41. For the second and third presentations the correlations were, respectively, +.10 and +.13. Subjects were more positive of their first choice than of their second or third choices. The relative indecision in the second and third choices was regarded as a factor in the low correlations.

The second question was asked to determine the subjects' opinions regarding the use of the audio channel for the transmission of the affective elements in the poem. Twenty-three of twenty-four subjects indicated that they felt the audio treatment added more meaning than was present without it, and the twenty-fourth stated that there were "about one-half and one-half" of the instances when the voice would help and when it wouldn't.

Notations were kept of any non-verbal expressions exhibited by subjects during the presentations. Seen occasionally were frowns (either of puzzlement or of distaste), winces, shrugs, tongues being stuck out, and wrinkled noses. In general there were few overt reactions such as these. When they did occur, however, the majority of them were during presentations involving the audio channel.

Statistical Design

For the cause of statistical reinforcement and to determine if an ordering and/or improvement-through-practice effect (Design # 2) would occur, two statistical designs of twenty-four subjects each were selected for this study. Subjects were randomly assigned to treatment types and sequences.

In the first design, involving twenty-four subjects, each received all three treatments—the audio, the visual, and the audiovisual—in one of the following sequences: A–V–AV, V–AV–A, V–A–AV, AV–A–V, AV–V–A, or A–AV–V. A one-way analysis of variance was used to determine statistical significance between treatments.

With this design it was possible to compare all subject performances at the end of the first two-stanza treatment. For example, it was possible to compare all the audio presentations with all the video presentations and with all the combined presentations for all twenty-four subjects.

The second design utilized in this study was a repeated measures design for testing each subject in only one treatment in all three presentations; i.e., each subject was presented all three pairs of stanzas via either A–A–A, V–V–V, or AV–AV–AV.

In addition to the repeated measures analysis which compared differences in subjects' scores, a comparison was made to determine the equivalence of the scores from the three two-stanza presentations; i.e., the equivalence of the difficulty of questions for presentation #1 compared to #2 and compared to #3. The test showed a high degree of equivalence.

The semantic differential scores were assigned to the same types of statistical design and analysis as explicated in the previous discussion. The results of the analyses of the findings follow.

FINDINGS

Cognitive Results

The data collected in this study provided a basis for comparing the amount of learning exhibited by the subjects in each of three presentation treatments of meaningful verbal stimulus material. Subjects' responses stimulated by reading, listening, or reading and listening simultaneously provided affective and cognitive data for study and interpretation.

Scores from the first presentation of the various treatments provided the only comparative data in which all forty-eight subjects were included. The mean number of errors per treatment was as follows: Audio: 6.44. Visual: 5.94. Audiovisual: 2.88.

A one-way analysis of variance was used to analyze this data. The results of the analysis, shown in Table I, proved to favor significantly the combined presentation over either individual presentation.

The first experiment which was designed so each subject received the stimulus in three separate units, A, V, and AV, yielded the analysis in Table II.

Table I. Analysis of First Presentations for All Subjects

Source of variation	*Sum of squares*	*Degrees of freedom*	*Mean squares*	*F-Ratio*
Between	327.4	2	163.7	88.96*
Within	82.6	45	1.836	
Total	410.0	47		

*p < .001

Table II. Analysis of Variance for Subjects Receiving All Three Treatments

Source of variation	*Sum of squares*	*Degrees of freedom*	*Mean squares*	*F-ratio*
Between	204.749	2	102.375	34.58*
Within	204.126	69	2.96	
Total	408.875	71		

*p < .001

Table III. Analysis of Variance for Repeated Measures

*Source of variation**	*Sum of squares*	*Degrees of freedom*	*Mean squares*	*F-ratio*
Mean	1880.8889	1	1880.8889	
T	148.3611	2	74.1806	23.107†
R:T	67.4167	21	3.2103	
M:T	14.5000	6	2.4167	1.070‡
RM:T	94.8333	42	2.2579	

* This table is to be read as follows:
T = treatments, R = replications per treatment,
M = measures per replication
† p < .001 ‡ n.s.d.

Table IV. Total Scores and Mean Scores on Semantic Differential Scales

	*1st design: one way analysis of variance**			*2nd design: repeated measures analysis of variance*		
SCORES	A	V	AV	A	V	AV
TOTAL	281	279	271	265	281	279
Mean	11.7	11.6	11.3	11.0	11.7	11.6

* Each design included twenty-four subjects

The second experimental design included eight subjects within each of the three treatments, with each subject assigned to only one treatment, audio only, visual only, or audiovisual. The mean number of errors in this design was as follows: Audio: 6.33. Visual: 5.96. Audiovisual: 2.58.

The analysis of this data is presented in Table III.

The analysis in Table III shows that at the .001 level of confidence the combined presentation had significantly fewer errors than either the audio or the visual presentations alone.

Affective Results

The second hypothesis proposed in this study was that subjects would show greater affective learning when the stimulus was presented via the combined presentation than when it was presented via the audio or the visual channels alone. On the basis of data from the semantic differential scales, no significance was found between treatments for the affective data.

DISCUSSION

Data shows that the combined audiovisual presentations were significantly superior to either the audio or the visual presentations alone at the .001 level of confidence. This high level of significance indicated that the Cue Summation Theory is supported when tested with meaningful information.

Further support for the Cue Summation Theory came from the comments made by the subjects during the testing sessions. Most of the subjects (96%) preferred the combined presentation, indicating they felt the treatment was superior. Some validity can be accredited to these subjects' opinions, because there was a +.41 correlation between their preferences and their performances on "best" treatments.

In order to be consistent with Severin's (1967) limitations on the term "cue summation," the stimulus material in the present study utilized only one coding system or modality. Within this verbal mode, the actual text of materials presented was identical in both channels. In the visual form the overt cues to the subjects were the printed words which the subjects read. Any other cues that were utilized by the subjects were covert and supplied by them.

In the audio presentation it was assumed that the narrator's voice cued more information than just the printed text. This assumption was supported by statements from the independent evaluators for this study who agreed that the narrator's inflection was appropriate for the stimulus. Additional evidence that cues were supplying "more" information was reported by the subjects themselves, many of whom stated that the parts with the voice sounded "more emotional" or "sadder."

The above statements which report increased learning from the audio presentations are not supported by the affective data. These data show the audio treatment was consistently inferior to the visual treatment. However, the differences were not statistically significant. The combined treatment also showed no superiority.

A concern in the present study was for the so-called "visual learners" and "listeners" and the pace of the presentations. The "visual learners" are those who contend they learn better by seeing than by any other method, while the "listeners" learn better through aural means. Involved in either the visual or the audio preference is pacing. While few comments were made regarding the pacing of the stimulus during this study, some students did mention they prefer to read at their own pace rather than being held to the pace of a narrator. Important to a reader is the control he has over his own pace. He can, if he chooses, revert to points not comprehended during a first reading and accelerate or decelerate at his own pace. Subjects who were presented the visual treatment had this freedom during the study and may have exercised it (or they may have been influenced by the "Hawthorne Effect") even though they were asked to read once through the stimulus at their "normal" reading pace.

By contrast, the students listening to the audio treatment may have been restricted to the pace of the narration. Even though the students indicated that the narration communicated a greater degree of emotion and feeling than the visual presentation, they did not display a greater comprehension of the content as determined by the affective tests.

The semantic differential data did not support the Cue Summation Theory, because of a lack of significant differences between the pairwise combinations of treatments. As evidenced by analyses of the two designs, no method was consistently superior.

Several reasons can be offered for the lack of differences appearing on these scales. First, it might be reasoned that the semantic differential technique is not a very sensitive measuring instrument. Thus, when the affective measures are small, the semantic differential scale is not capable of quantifying the affective data.

A second possibility is that differences in affective learning did not actually occur for the separate treatments. While this suggestion is a possibility, it seems unlikely in view of the subjects' responses during the open-ended questioning. Most of the subjects stated that during the audio presentations their emotions were heightened more than during the visual alone presentations.

A third possibility for the lack of variance in the affective measure between treatments would be the relative lack of experience the subjects had had with the semantic differential scale. No subject admitted having used such a scale prior to this study. This lack of prior experience in the use of the scale made it necessary to have the explanatory and practice sessions for each subject prior to testing. After each subject had completed a scale, the experimenter occasionally asked questions regarding clarification of some of their checkmarks. Their explanations of the checks and elaborations of their feelings indicated that the marks were accurate.

Conclusions

The first hypothesis posited for this study was that a combined audio-visual presentation of redundant meaningful verbal information would result in a greater comprehension of the material than would either the audio or the visual presentations alone. The cognitive data, with a high level of significance, support this hypothesis.

The second hypothesis stated that subjects would display greater affective responses to the stimulus when both channels were used simultaneously for the presentation of redundant meaningful verbal information than when they were used alone. This hypothesis was not supported by data from the semantic differential scale.

This study has shown that bi-sensory communication which combines words in two channels (words aurally and words in print) will result in significantly greater information gains than single-channel communications. This does not support the predictions and findings of Severin (1967), who concluded that multi-channel communications are no better when they are merely redundant.

Summary

For cognitive data, the findings of the present study support the Cue Summation Theory. Students receiving redundant meaningful verbal information through both the audio and the visual channels simultaneously performed significantly better on tests of information recall than did subjects receiving the same information in either the audio or the visual channel alone.

With the same stimuli, there were no significant differences between treatments with affective data from semantic differential scales.

REFERENCES

Conway, Jerome Kenneth. *Differential Memory for Referentially Equivalent Single and Multiple Sign-Vehicle Presentations*, unpublished doctoral dissertation, Department of Mass Communication, Indiana University, 1968, 83 pp., typed.

Hartman, Frank R. "Recognition Learning under Multiple Channel Presentation and Testing Conditions," *Audiovisual Communication Review 9*: 24–43, Spring, 1961.

Knowlton, James Q. *A Socio-and-Linguistic Theory of Pictorial Communication*, U.S. Department of HEW, Office of Education, Contract No. 3–16–019, November, 1964, Indiana University.

Loveless, N. E., Brebner, J., and Hamilton, P. "Bisensory Presentation of Information," *Psychological Bulletin 73*: 161–199, March, 1970.

Severin, Werner. "Pictures as Relevant Cues in Multi-Channel Communication," *Journalism Quarterly 44*: 17–22, 52, Spring, 1967.

VanMondfrans, Adrian P., and Travers, R. M. "Learning of Redundant Material Presented through Two Sensory Modalities," *Perceptual and Motor Skills 19*: 743–751, 1964.

Wilbur, Richard. "The Pardon," Summerfield, Geoffrey, ed., *Voices: An Anthology of Poems and Pictures*. Chicago: Rand McNally, 1969, 127 pp.