The long-term modality effect: In search of differences in processing logographs and alphabetic words*

In-mao Liu

Department of Psychology, Chinese University of Hong Kong, Shatin, NT, Hong Kong

Ying Zhu

Peking University, Beijing, China

Jei-tun Wu

National Taiwan University, Taipei, Taiwan

Abstract

Liu, I.-m., Zhu, Y., and Wu, J.-t., 1992. The long-term modality effect: In search of differences in processing logographs and alphabetic words. Cognition, 43, 31–66.

The visual superiority effect (a reverse long-term modality effect) has been consistently found with Chinese logographs. For its explanation in terms of script differences, it has been believed that lexical access is more direct or quicker for Chinese logographs than for alphabetic words. It has also been believed that Chinese logographs are more unique in shape or more discriminable than alphabetic words. Finally, Chinese logographs have been considered to facilitate recall through their graphic features that classify Chinese words into categories. The results of Experiments 1–5 show that these three assumptions can be ruled out. The results of Experiments 6–10, on the other hand, support the long-term priming interpretation of the visual superiority effect, which explains (a) why the visual superiority effect can be consistently obtained for recall of Chinese words by

Chinese subjects, (b) why the effect cannot be consistently obtained for recall of English words by Western subjects, (c) why the effect can be also obtained for recall of English words by Chinese subjects, (d) why the effect can be easily obtained for recall of a set of words, but not for recall of a different set of words by Chinese subjects, and (e) why the effect can be easily obtained from Chinese subjects speaking a dialect that is different from Mandarin.

Introduction

The modality effect refers to the typical finding in immediate free recall or serial recall that the last few items in the list are better recalled if presentation had been auditory rather than visual (Conrad & Hull, 1968; Corballis, 1966; Craik, 1969; Crowder & Morton, 1969; Murdock & Walker, 1969; Murray, 1966). The long-term modality effect will be used in this paper to refer to the auditory or visual superiority over the other modality obtained for pre-recency items. In this respect, the long-term modality effect found with recency items by Gardiner and Gregg (1979) and Glenberg (1984) using a special paradigm will not be considered. There are three accounts of the long-term modality effect as follows: (a) the temporal distinctiveness, (b) the differential script, and (c) the long-term priming (differential frequency).

Temporal distinctiveness interpretation

As Penney (1989b) remarked, up to the late 1970s, there was no evidence for consistent modality effects in either long-term memory tasks or in the non-recency part of the serial position curve in short-term memory tasks. For example, no modality effects were found for non-recency items in the study by Engle and Mobley (1976). It is only more recently that Conway and Gathercole (1987; Gathercole & Conway, 1988) have reported the auditory superiority effect in a long-term recognition memory task. In their studies, auditory memory advantage was present throughout list positions. They invoked the concept of temporal distinctiveness proposed by Gardiner (1983) and elaborated by Glenberg (1984; Glenberg & Swanson, 1986) for explaining their results as follows. The temporal distinctiveness account assumes that the memory trace preserves temporal information about the time of presentation more accurately for auditory than for visual presentation. They reasoned that the recognition test cues the subject to particular parts of the list, and that the finer grained representation of acoustic information ensures better recognition than the wider grained representation of visual information.

More recently, Penney (1989b) showed that in both delayed free recall and

delayed recognition tests visual presentation of items produced better performance than did auditory. The finding of a visual superiority in delayed free recall and recognition provides strong evidence against the temporal distinctiveness view.

Differential script interpretation

A cross-language view of the long-term modality effect was first proposed by McGinnies (1965). In his study of persuasion through printed versus spoken communication, he found that Japanese subjects were more apt to be persuaded by printed than by spoken messages, contrary to findings frequently obtained with Western subjects (Hovland, 1954; Klapper, 1961). Since Japanese script includes Chinese characters, McGinnies reasoned that the visual superiority effect in persuasion may be due to script differences. Subsequently, Turnage and McGinnies (1973) had their American and Chinese subjects study a list of 15 words in a serial-learning paradigm. The input modality of the stimulus presentation was manipulated. Their major finding was that Chinese subjects learned the list of two-character words faster when it was presented visually, whereas American subjects learned the list of words faster when it was presented auditorily. Their explanation of this difference was that the Chinese script system contains more characters with similar sounds but different meanings than is the case for English. A problem with Turnage and McGinnies' interpretation is that their conclusion was not based on the results obtained by manipulating sound similarity.

Fang (1982) presented a list of nine items (either Chinese characters or two-character words) visually or auditorily at a speed of 1.5 s per item to a group of 19 Chinese subjects. Their recall results showed that more primacy items were recalled in the visual than in the auditory presentation mode. Fang's interpretation of her findings was the same as that of Turnage and McGinnies (1973); that is, the Chinese script system contains more characters with similar sounds, although in her experiments she avoided using different characters with similar sounds and similar meanings.

Although experimental details were not reported, Tzeng and Wang (1983) obtained a similar finding. In their study their subjects recalled lists of nine Chinese characters according to their position in the series. They also found that their Chinese subjects recalled the beginning items in the series consistently better when the lists were presented visually than auditorily. No such memorial superiority of visually presented items was obtained from their English subjects. Recently, Hue, Fang, and Hsu (1990) replicated Tzeng and Wang's findings that the visual superiority effect is quite robust when to-be-recalled items are Chinese logographs.

Assuming that the script difference is responsible for the long-term modality effect, there are several views that attempt to account for the visual presentation

superiority obtained with Chinese logographs: direct-image hypothesis, discriminability hypothesis, and graphic-feature hypothesis. Let us consider each hypothesis in turn.

Direct-image hypothesis

This is a widespread view that Chinese logographs map more directly onto meaning than alphabetic words (e.g., Aaronson & Ferres, 1986). A passage from Wang (1973) reads as follows: "To a Chinese the character for 'horse' means horse with no mediation through the sound ma. The image is so vivid that one can almost sense an abstract figure galloping across the page." This passage has been frequently cited to support theories (e.g., Biederman & Tsao, 1979) without receiving rigorous experimental tests.

If Chinese logographs directly give rise to images or are more like pictures, then the visual presentation superiority of Chinese logographs follows, because the superiority of pictures to words in free recall and recognition is a well-documented result (see a review in Crowder, 1976). A developmental study (Hochberg & Brooks, 1962) also showed that pictures are not linguistic devices but rather accurate depictions of reality, and can be recognized without mediation as well as the objects themselves are recognized.

Theoretically, if Chinese logographs directly give rise to images, they will be coded both verbally and imaginally according to Paivio's dual coding theory (1971, 1986). For those who propose the direct-image hypothesis, an assumption has to be made. That is, they have to assume that alphabetic words are coded verbally and, to a lesser extent, imaginally. Thus, an added image trace is assumed to result in better memory for Chinese logographs than for alphabetic words, if they are presented visually.

Why does an added image trace result in better memory? The dual coding theory assumes that the image system is specialized for processing spatial and synchronous information whereas the verbal system is specialized for sequential processing. That dually coded items are remembered better than unitarily coded items follows simply from the additivity of independent verbal and image components of a memory trace (e.g., Paivio, 1986). Experiments 1 and 2 were designed to test the direct-image hypothesis.

In the following, instead of using the term "verbal" a more unambiguous term "phonological" will be used, because Paivio considered "visual words" as verbal. Therefore, any memory traces with visuo-spatial features will be considered as being visually coded and any memory traces with sound features as phonologically coded. With this terminology, the direct-image hypothesis explains the visual presentation superiority of logographs by assuming that, in addition to phonological traces, logographs leave traces of object images directly, whereas alphabetic words do not, or only indirectly (referentially, according to Paivio, 1971), leave

traces of object images. Of course, an added image trace is the source of the visual superiority effect.

Discriminability hypothesis

According to the discriminability hypothesis, as in the direct-image hypothesis the visual presentation superiority of logographs is explained by their visual traces. However, these visual traces represent character shapes. Logographs do not necessarily give rise directly to object images. Since logographs and alphabetic words leave both phonological and visual traces in this case, the visual superiority effect of logographs is now explained by logographs being more unique and distinctive than alphabetic words, when they are presented visually. This interpretation is also prevalent (e.g., Tzeng & Hung, 1988). In the input stage, if Chinese characters leave unique traces via their unique shapes, then they would be much easier to retrieve at the time of testing. Experiment 3 was designed to test the widely held hypothesis that Chinese characters are visually more unique or discriminable than English words.

Graphic-feature hypothesis

The third hypothesis also attributes the visual presentation superiority of logographs to coded visual traces. This interpretation asserts that the obtained long-term modality effect for Chinese characters is primarily due to meaning components being frequently associated with graphical features in Chinese characters. That most Chinese characters contain an explicit meaning-conveying component called a *radical* is well known. Contrary to a widely held hypothesis that inter-character distinctiveness is responsible for the visual presentation advantage of Chinese characters, the third hypothesis asserts that inter-character visual relatedness is mainly responsible for it.

Most Chinese words consist of two characters. There are many two-character words with the same second characters to indicate that a set of words belongs to the same category, such as *orchid-flower*, *peach-flower*, etc. At the level of characters, the graphic-feature hypothesis asserts that the visual presentation superiority of logographs is explained by an abundance of Chinese words with the same character endings. Experiments 4 and 5 were designed to test the graphic-feature hypothesis by directly manipulating graphic features.

Long-term priming interpretation

A difficulty with the differential script interpretation is that the visual superiority effect can be sometimes obtained with alphabetic words (Penney, 1989b). A plausible interpretation cannot be, therefore, dependent only on modality-depen-

dent features (Nairne, 1988), nor on the visual or auditory streams (Penney, 1989a) alone. It must depend on an operation that can reverse the relative effectiveness of visual versus auditory experiences. In other words, a viable interpretation must be capable of explaining the presence of the visual superiority effect in a specifiable condition and its absence in another specifiable condition.

There is considerable evidence that the prior presentation of a nameable item facilitates its subsequent processing even when the two occasions are separated by several hours or days (Scarborough, Cortese, & Scarborough, 1977). Kirsner and Dunn (1985) refer to this facilitatory effect as long-term priming in order to distinguish it from other priming effects that appear to have different characteristics and to be confined to much smaller lags. Jacoby and Dallas (1981) found that repetition of a word during study enhances recognition memory and that switching the modality of presentation between study and test reduces recognition memory performance. The finding of long-term priming has been explained as due to the lowered threshold of recognition units in semantic memory (Morton, 1969) or as due to the recapitulation of previously engaged analytic procedures (Kolers, 1979).

According to this long-term priming interpretation, whether visually presented items are more recallable than auditorily presented materials depends on the differential availabilities of visual versus auditory traces. If some Korean words are more frequently written in Chinese logographs than in Hungul, then the visual traces for these Chinese logographs become more available than those for Hungul, as was obtained by Park and Arbuckle (1977). If a set of words is encountered more frequently in readings than in speech, then their visual traces will become more available than their auditory traces, giving rise to a memorial superiority of visually presented items over auditorily presented items, as was found by Turnage and McGinnies (1973) from their Chinese subjects but not from their American subjects.

The long-term priming interpretation assumes that there are two types of frequency: (a) frequency of a verbal item as it appears in print and (b) frequency of a verbal item as it appears in speech. Visual frequency is responsible for the availability of visual traces and auditory frequency for the availability of auditory traces. Whenever there is a large discrepancy between visual and auditory frequencies in favour of the former, the visual presentation superiority will be obtained. It is perhaps because of this reason that availability was found to be an important predictor of recall but that the classically important variable of word frequency was not (Rubin & Friendly, 1986).

The long-term priming interpretation or differential-frequency account has the following implications. First, if a list of English words is presented visually or auditorily to Chinese subjects for free recall, the visual superiority effect will obtain because English is after all a foreign language for Chinese subjects and English words have been experienced more frequently in print than in speech. This implication was tested in Experiments 6 and 7.

As for a second implication of the differential-frequency account, it may be noted that some set of Chinese words is experienced more frequently in print than in speech while another set of Chinese words is experienced more frequently in speech than in print. The differential-frequency account, then, predicts that at least a larger visual superiority effect may be obtained from the former set of words than from the latter. Experiments 8 and 9 were designed for testing this prediction.

There are many different varieties of Chinese known as "dialects" (see, for example, Li & Thompson, 1981), even though they may be different from one another to the point of being mutually unintelligible. Thus, it is often pointed out, for example, that Cantonese and Mandarin (national language spoken frequently at formal occasions and in schools) differ from each other roughly as the Romance "languages" Portuguese and Romanian do. Now, it is possible to find a group of subjects who speak the same dialect (Mandarin) at home and in school. It is also possible to find another group of subjects who speak Mandarin in school and use their own dialect at home. For the former group, although the auditory frequency will accumulate across school and home situations, it will not for the latter group. The differential-frequency account predicts that a larger visual superiority effect will be obtained for the latter group than for the former. This prediction was tested in Experiment 10.

Experiment 1

If Chinese characters are also visually encoded to stand for images, how are these images measured? Following Frege (1960), the meaning of a word is identified by its sense or connotation and reference or denotation. The denotational meaning of an expression is the thing that it stands for. The connotational meaning of an expression is the content of the expression, and specifies whether the expression has that object as its reference.

The question "what does *horse* mean?" asks the connotational meaning of a horse. A partial answer would be that a horse is an animal. The connotational meaning of a word is therefore different from associations. For example, *black* may be an associate to *white*, but *black* is not a connotational meaning of *white*. The strategy of Experiment 1 is then as follows. First, connotational meanings of a set of Chinese characters and their English translations were obtained from a group of subjects. For each Chinese character or its English translation, the response most frequently given by the subjects as representing its connotational meaning was then selected as the dominant meaning component. Second, each dominant meaning component was presented as a prime in a reaction time task in which another group of subjects was timed for judging whether a Chinese character or English word presented after the prime contains the prime as its

dominant meaning. If Chinese characters are coded both verbally and imaginally, then the subject should contact images directly for immediate responses. Since English words are coded predominantly verbally, the subject has to rely on the two routes to find a match or mismatch between the prime and test word: either reaching the dominant meaning component through association from the verbal code or generating an image indirectly (referentially) from the verbal code. Either route should be much slower for English words than a direct route for Chinese characters according to the direct-image hypothesis.

In line with the prediction from the direct-image hypothesis, there are a number of studies conducted for comparing semantic classification of pictures and words (Guenther & Klatzky, 1977; Guenther, Klatzky, & Putnam, 1980; Pellegrino, Rosinski, Chiesi, & Siegel, 1977; Potter & Faulconer, 1975; Smith, Bolzano, & Walker, 1978). This task requires a subject to indicate whether or not a presented item contains a target feature or in other words, is a member of a given target category. It was generally found that pictures are classified faster than words.

There are two types of Chinese characters: phonograms, which contain a phonetic component as a clue to the sound of characters, and pictograms, which are pictographic in origin. The direct-image hypothesis may apply to pictograms more than phonograms. Accordingly, the direct-image hypothesis would predict that semantic processing should be fastest for Chinese pictograms, next for Chinese phonograms, and slowest for English words. The reason for using Chinese characters instead of two-character words is simply because characters are units of the Chinese script system. It would be impossible to choose a sufficient number of two-character words, each consisting of two pictograms or two phonograms. Moreover, many Chinese characters, being morphemes themselves, can be used as words. Most Chinese characters selected in the present study also function as words.

The plan of Experiment 1 was then as follows. A set of Chinese pictograms and phonograms was chosen. These pictograms and phonograms were translated into English words. A dominant meaning component of each item was determined by testing with a group of subjects. For example, *water* is considered as a dominant meaning component of *river*. Then, another group of subjects was tested for making a semantic decision of whether a Chinese character or English word contained a given dominant meaning component. The preparatory study conducted to determine a dominant meaning component for each Chinese character and English word will be described only briefly.

In the preparatory study, 32 characters were chosen from Tung Lai (1978). Half the characters were pictograms and half were phonograms. Their English translations are listed in Table 1. Two comparable high school classes were chosen, each having 40 second year students. Hong Kong students studied English from primary school. Therefore, their English is on the average much better than

Table 1. *English translations and their dominant meaning components*

Pictogram		Phonogram	
Translation	Dominant meaning component	Translation	Dominant meaning component
BLOOD	RED	ATTACK	WAR
CAR	DRIVE	BOARD	WOOD
CLOTHES	KEEP WARM	COLD	WEAR CLOTHES
COW	ANIMAL	EMPTY	NOTHING
EAR	LISTEN	FINISH	END
EAST	DIRECTION	FLOWER	BEAUTIFUL
FIELD	PLANT	NEAR	SHORT DISTANCE
FIRE	HEAT	OIL	LIQUID
FRUIT	FOOD	PULL	FORCE
HAIR	BLACK	RIVER	WATER
MEAT	ANIMAL	SHOP	GOODS
RAIN	WATER	SIMILAR	SAME
SHEEP	WHITE	SISTER	GIRL
STEP	FOOT	TASTE	SWEET
TREE	LEAVES	THING	SHAPE
WHITE	CLEAN	WORD	WRITE
Mean Strokes	5.875		7.625
Mean Frequency	520		614

Note: Mean strokes and mean frequency refer to Chinese pictograms and phonograms.

that of students from the other Asian countries where English is usually taught from high school. A set of booklets containing the 32 characters in various randomized orders was used for one class of students, and another set of booklets containing the 32 English translations used for another class of students. For each item printed on a separate page of a booklet, the students were allowed 10 s to write down at most three meanings that came to their mind instantaneously. In Table 1 is shown the dominant meaning component for each English translation.

Experiment 1 attempted to measure how long it takes for the subject to verify whether an item (Chinese or English) contains a dominant meaning component (Chinese or English).

Method

Subjects and materials

The subjects were 20 freshmen enrolled in an introductory psychology course at the Chinese University of Hong Kong. They participated in the experiment to fulfil a course requirement.

The stimulus materials were 32 Chinese characters and their English translations. The dominant meaning component for each item was obtained through a preparatory study as was described. The 32 English translations and their dominant meaning components are presented in Table 1. The classification in Table 1 is based on the Chinese characters. Thus, the mean number of strokes for the 16 phonograms was significantly higher than that for the 16 pictograms, $t(30) = 5.03$, $p < .01$. The mean word frequencies were, however, not significantly different from each other, $t < 1$.

Procedure

The subject was timed for pressing either of two keys to indicate whether a target item contained a dominant meaning component presented as a prime 2 s prior to the target. A target was either a Chinese character or its English translation. Targets were further subdivided into pictograms and phonograms or their English translations. The present design was then a 2 (language of primes) × 2 (language of targets) × 2 (pictogram or phonogram) within-subjects factorial.

All stimuli were presented by an IBM PC. Chinese characters were generated by a chip manufactured by Kuo Chiao. Reaction times were measured in units of milliseconds. Sixteen practice trials preceded 64 experimental trials. Type of target (Chinese or English) was mixed. For half the subjects, Chinese primes preceded English primes. The order was reversed for the remaining subjects. For each type of prime, a dominant meaning component preceded a target on half the trials ("true" trials). On the remaining trials ("false" trials), meaning components and targets were randomly re-paired with a restriction that they were not apparently related. Half the targets were used on positive trials, and the remaining half used on negative trials.

Results and discussion

The mean reaction time for identifying the presence of a dominant meaning component in a Chinese character or English word is listed in Table 2 for each type of prime and each type of target. An analysis of variance showed that reaction times were faster for English primes than the Chinese primes (810 vs. 830 ms), $F(1, 19) = 4.43$, $p < .05$. Reaction times were not faster to pictograms than to phonograms (814 vs. 825 ms), $F < 1$. Chinese characters were also not responded to faster than English words (817 vs. 822 ms), $F < 1$. All two-way and three-way interactions were not significant. An analysis of variance performed on the error data showed that more errors were obtained for Chinese primes than for English primes (12.3% vs. 9.5%), $F(1, 19) = 5.15$, $p < .05$. All other effects were not significant. The data obtained on the negative trials were not specifically

Table 2. *Semantic decision time in milliseconds to a Chinese character or English word*

Target	Prime	
	Chinese	English
Chinese		
Pictogram	812	797
	(11.6)	(7.2)
Phonogram	838	820
	(10.6)	(13.1)
English (translation)		
Pictogram	826	820
	(11.6)	(8.4)
Phonogram	842	801
	(15.3)	(9.4)

Note: The percentage of errors is indicated within parentheses.

analysed, because the same tendency was obtained. The only difference was that the absolute reaction times on the positive trials were faster than those on the negative trials.

The Chinese and English meaning components used as priming stimuli were not identical. Reaction times were faster for English primes perhaps because a fewer number of meaning components were obtained for English words than for Chinese characters from the first group of Chinese subjects in the preparatory study and because the obtained meaning components were more clustered for English words than for Chinese characters. These observations are consistent with the finding that when words are used more frequently they tend to acquire more meanings (Johnson-Laird, 1983; Miller, 1951). For the Chinese subjects, a Chinese word tends to arouse more meanings than its English translation because the former has been used more frequently.

The mean number of strokes was higher for phonograms than for pictograms. However, this difference would have only made the subjects recognize pictograms faster than phonograms (Yeh & Liu, 1972). In spite of this difference that favoured pictograms over phonograms on semantic decision time, pictograms were not responded to faster than phonograms. As a whole, the results of Experiment 1 did not support the direct-image hypothesis.

Experiment 2

Although reaction time measures of Experiment 1 were found to be sensitive enough to detect a minor difference between Chinese and English primes, a rejection of the direct-image hypothesis cannot be made on the basis of no

difference in the semantic decision times. Moreover, semantic decision may not be faster for pictures than for words in some cases (e.g., Paivio & te Linde, 1980; te Linde, 1983). Therefore, in Experiment 2 subjects were not only timed for judging whether Chinese pictograms or phonograms contained some meaning components but were also timed for judging whether pictures drawn for the pictograms and phonograms contained the same meaning components. In this situation, the direct-image hypothesis would predict that semantic decision times for pictograms should be as fast as for pictures and that they should be faster for pictograms than for phonograms.

Method

Subjects and materials

The subjects were 21 freshmen enrolled in an introductory psychology course at the Chinese University of Hong Kong. They had not served in Experiment 1, and participated in the experiment for fulfilling a course requirement.

Twelve pictograms and 12 phonograms of Experiment 1 which could be represented by pictures were used. The size of pictograms and phonograms was 2.5 cm wide by 1.2 cm high on an IBM PC monitor. The size of a picture was slightly larger and spanned a space of about 3–5 cm wide by 2 cm high.

Design and procedure

The design was a 2 × 2 × 2 within-subjects factorial. The first factor refers to type of target, a character or picture; the second to source of target, either based on a pictogram or phonogram; and the third to type of response, a match or mismatch between a target and a prime. Following 12 practice trials, each subject received 48 experimental trials. The set of primes and targets used on the practice trials was entirely different from those used on the experimental trials. On each trial, a dominant meaning component was presented as a prime for 2 s. Then, either a target character or picture followed. The subject verified whether the target contained, or was consistent with, the dominant meaning component. Each target was presented only twice: once on a positive trial and once on a negative trial. Positive and negative trials were counterbalanced between subjects, and the order of trials was randomized for each subject. The other procedural details were the same as in Experiment 1.

Results and discussion

The mean reaction times measured from the target onset are listed in Table 3. An analysis of variance showed that reaction times were faster to pictures than to

Table 3. *Semantic decision time in milliseconds to a character or picture*

Type of response	Character		Picture	
	Pictogram	Phonogram	Pictogram	Phonogram
True	757	762	736	714
	(10.3)	(4.7)	(15.9)	(7.9)
False	854	901	800	826
	(7.1)	(12.7)	(6.3)	(7.9)

Note: The percentage of errors is indicated within parentheses.

Chinese characters (769 vs. 819 ms), $F(1, 20) = 7.22$, $p < .025$. Reaction times were not faster to pictograms than to phonograms (787 vs. 801 ms), $F < 1$. Positive responses were faster than negative responses (742 vs. 845 ms), $F(1, 20) = 19.77$, $p < .001$. All two-way and three-way interactions were not significant.

A separate analysis of variance performed on the error data showed that each main effect was not significant, $F < 1$. The source of target by type of response interaction was significant, $F(1, 20) = 20.24$, $p < .001$. This interaction was obtained apparently because pictures for representing some pictograms were poorly drawn owing to a limitation in computer graphics. Similarly, the type of target by type of response interaction was significant, $F(1, 20) = 4.50$, $p < .05$, also indicating that some pictures for representing characters were poorly drawn.

The obtained results present strong evidence that the direct-image hypothesis should be rejected. Chinese pictograms do not give rise to images directly as pictures. In other words, even Chinese pictograms can be coded only indirectly to give rise to images.

Experiment 3

If Chinese characters have more unique shapes than English words, then it follows that Chinese characters are more discriminable from one another than English words. One of the simplest ways to find out whether such differential discriminability exists between the two languages is to make use of triadic judgments. On each trial a triangular array of three words is presented. The subject's task is to indicate, by pressing a right or a left key, which of the bottom words, the left one or the right one, is identical to the top word. In this kind of triadic judgment, it is a well-known fact that the more similar are two comparison stimuli the slower is the reaction time. If Chinese characters are visually more unique than English words, the subject's triadic judgments would be faster when any two Chinese characters are used for comparison than when any two English words are used.

An important variable that affects stimulus discriminability is stimulus complexity. Stimulus complexity for English words can be identified as number of letters in a word. Number of letters in an English word is comparable to number of strokes in a Chinese character, so far as visual stimulus complexity is concerned. In Experiment 3, the variable of stimulus complexity was also manipulated to see whether it has comparable effects on the two languages.

Method

Subjects and stimulus items

The subjects were 26 freshmen enrolled in an introductory psychology course at the Chinese University. Their English might be as nearly good as their Chinese. They participated in the experiment to fulfil a course requirement.

It was reasoned that Chinese characters with numbers of strokes from 5 to 6 are comparable to English words with three letters in terms of stimulus complexity (A category). Similarly, numbers of strokes from 7 to 8 are comparable to four letters (B category), and numbers of strokes from 9 to 10 comparable to five letters (C category).

The above reasoning is based on the following observations. When counting the number of "strokes" for each English letter written in the lower case as a Chinese character is counted, it is readily seen that "a" has two strokes, "b" has two strokes, "c" has one stroke, etc. The average number of strokes for an English letter is therefore between 1.5 and 2.0.

One hundred characters/words each of A, B and C categories were randomly selected from high-frequency characters (words) (above 50 per million) in Cheng (1982) and Thorndike and Lorge (1944).

Design

A 3×2 within-subjects factorial design was used. The first factor refers to discriminability between two comparison characters/words used for triadic judgments. There were three degrees of discriminability: 0, 1, 2, to indicate the distance between two comparison characters (words) belonging to two categories. If two characters (words) were from the same category, the distance was 0. If two characters (words) were from two adjacent categories, the distance was 1. Finally, if two characters (words) were from categories A and C, then the distance was 2. The second factor refers to language: Chinese or English.

Procedure

Half the subjects were tested with Chinese characters first and then tested with English words next. The order was reversed for the remaining subjects. For each

language, following 12 practice trials each subject received 72 experimental trials: 24 trials each under each discriminability condition. Depending on which discriminability condition a trial was assigned, two characters/words were randomly chosen from categories A, B and C without replacement. On half the trials the right key was correct, and on the other half the left key was correct. The order of trials was block randomized.

Results and discussion

The mean reaction time obtained under each discriminability condition for each language is presented in Table 4. An analysis of variance showed that type of language was not significant (635 for Chinese vs. 632 ms for English), $F < 1$. Discriminability was a significant source of variance (652, 638 and 611 ms), $F(2, 50) = 17.75$, $p < .001$. The discriminability by type of language interaction was not significant. It can also be clearly seen from Table 4 that the obtained error rates under various conditions were very low. An analysis of variance performed on the error data showed that no variable was a significant source of variance.

It is a surprise to find that Chinese characters were discriminated from each other no better than English words by the Chinese subjects. The concept of familiarity cannot be invoked to explain the present finding, because it only favours discriminability of a more familiar language (e.g., Ambler & Proctor, 1976; Egeth & Blecker, 1971). This means that Chinese characters should be more discriminable to the Chinese subjects than English words. Furthermore, in view of a small difference in stimulus complexity producing a significant effect in the present experiment, the discriminability hypothesis should be rejected to the effect that Chinese characters are no more discriminable among themselves than English words.

It is possible to check the correctness of the way in equating the complexity of the Chinese and English stimuli in the light of the obtained results. If more strokes or fewer strokes of a Chinese character have to be equated with an English letter,

Table 4. *Discriminability in milliseconds as a function of language and distance*

Language	Distance		
	0	1	2
Chinese characters	646	643	617
	(3.0)	(2.7)	(1.1)
English words	656	633	605
	(2.7)	(1.6)	(1.6)

Note: The percentage of errors is indicated within parentheses.

then an interaction between category distance and language will be obtained. Since the interaction was not significant, the rationale of equating 1.5–2.0 strokes with one English letter may be still justifiable. Even though the interaction had been significant, the present design would have allowed a legitimate comparison between the two languages by inspecting the difference obtained for the case of 0 distance (first column of Table 4).

Experiment 4

If it is not true that Chinese characters have more unique shapes than English words, and if it is also not true that Chinese characters are coded directly to give rise to images, then how do we explain the visual presentation advantage found with Chinese characters (Fang, 1982; Turnage & McGinnies, 1973)? A plausible hypothesis is that the obtained visual advantage is primarily due to meaning components being frequently associated with graphical features in Chinese characters. Most Chinese characters contain an explicit meaning-conveying component called a *radical*. Contrary to a widely held hypothesis that inter-character distinctiveness is responsible for the visual advantage of Chinese characters, the graphic-feature hypothesis claims that it is because many Chinese characters often share the same graphic features.

According to the graphic-feature hypothesis, logographs and alphabetic words will leave visual as well as phonological traces. It is with the visual traces left by Chinese characters that the latter can be more effectively used for retrieving the encoded characters. Although the visual traces left by alphabetic words are as distinctively different from each other as those left by Chinese characters, the former are not very helpful because spelling is only highly correlated with sound. In other words, the visual traces left by alphabetic words do not provide anything more than phonological information.

The plan of Experiment 4 was then as follows. In the radical-blocked condition, a group of subjects studied lists of words with the same radicals presented in blocks in the visual and auditory modes. In the radical-mixed condition, another group of subjects studied lists of words with the same radicals randomly mixed with others in the visual and auditory modes. In the control condition, a further group of subjects studied lists of words with distinct radicals (every radical appeared only once in some words) in the visual and auditory modes.

Method

Subjects and list items

The subjects were 150 students enrolled in an introductory psychology course at

National Taiwan University. They participated in Experiment 4 for fulfilling a course requirement.

It was impossible to find a sufficient number of two-character words with both characters of each word containing the same radicals. Therefore, the only requirement was the first characters of a set of two-character words being composed of the same radical. Four lists of seven sets of two-character words each were selected from a dictionary as follows. Each set consisted of four words with the same radical. All the sets in the four lists had distinct radicals. This means that all selected words were composed of the first characters with 28 different radicals. The four lists (A, B, C and D) were approximately equated for their mean frequency counts (Liu, Chuang, & Wang, 1975). A word was randomly selected from each set of four words with the same radical to obtain four control lists (C1, C2, C3 and C4), each of which consisted of 28 words with distinct radicals.

Design and procedure

A 2×3 mixed design was used. Modality (visual or auditory) was a within-subjects variable, while list composition (radical-blocked, radical-mixed or radical-distinct) was a between-subjects variable.

The subjects were assigned to three groups in order of signatures, and tested in a small group of five subjects at a time. Presentation of words was controlled by an IBM-compatible PC. In the auditory mode, digitized sounds were used. For the subjects assigned to the radical-blocked group, they studied two radical-blocked lists, one in the visual and another in the auditory mode. In the radical-blocked lists, each list contained seven sets of words. In each set, there were four words which all had the same radical. The words were presented in seven blocks, grouped according to their radicals. The subjects who were assigned to the radical-mixed group also studied two out of the four lists, A, B, C and D: one list in the visual and another in the auditory mode. In the radical-mixed lists, the words were the same as those in the radical-blocked lists, but the 28 words in each list were mixed up and presented in some random order. Finally, the subjects who were assigned to the radical-distinct group studied two out of the four control lists: one in the visual and another in the auditory mode. The control lists (radical-distinct lists) each contained 28 words, and each word had a distinct radical.

The subjects received a short practice list to familiarize themselves with the procedure before studying two assigned lists. The order of receiving visual and auditory lists as well as the choice of two lists for study was counterbalanced between subjects for each group. Immediately after presentation of a list, the subjects engaged in the distractor activity of solving two-digit number addition problems for 20 s before writing down as many words as they could still remember within a 5-min period.

Table 5. *Mean numbers of words correctly recalled as a function of presentation mode and list composition*

Input mode	List composition		
	Radical–blocked	Radical–mixed	Radical–distinctive
Visual	14.1	11.9	10.7
Auditory	11.2	10.0	7.6

Results and discussion

The mean number of words correctly recalled under each input condition for each group is presented in Table 5. It is clear from the table that recall was higher in the visual than in the auditory mode when words with the same radicals were presented in the blocked or mixed condition. More importantly, recall was also higher in the visual than in the auditory mode when a list consisted of words with distinct radicals. An analysis of variance showed that input mode was a significant source of variance (visual 12.2 vs. auditory 9.6), $F(1, 147) = 81.44$, $p < .001$. List composition was also significant (12.7, 11.0 and 9.2), $F(2, 147) = 18.56$, $p < .001$. The Scheffe's S test showed that recall was higher when words with the same radicals were presented in the same lists than when words with distinct radicals were included in the lists (11.9 vs. 9.2), $p < .01$. The interaction between input mode and list composition was not significant, $F(2, 147) = 1.46$, $p > .20$.

The finding that a list of words with the same radicals was recalled better than a list of words with distinct radicals is easily accounted for in terms of category clustering. A set of words with the same radical often implies that the words are from the same category. However, the finding that recall was higher in the visual than in the auditory mode when radicals were distinct as when radicals were blocked or mixed indicates that the graphic-feature hypothesis cannot explain the results. This is because the graphic-feature hypothesis would predict that the visual presentation advantage should be larger when the words in a list have the same radicals because the shared graphic-features can serve as retrieval cues.

Experiment 5

According to the second version of the graphic-feature hypothesis, the visual presentation superiority of Chinese logographs is due mainly to graphic features at the character level.

Method

Subjects and list items

The subjects were 240 students enrolled at Peking University. They participated in the experiment voluntarily.

Four lists of two-character words were selected from the Liu, Chuang, and Wang (1975) word book as follows. Each list consisted of eight sets of four words each. For one list (C1), four words in each set had an identical first character. For another list (C2), four words in each set had an identical second character. For still another list (S1), four words in each set had an identical first syllable but different characters. For the final list (S2), four words in each set had an identical second syllable but different characters. There is a sufficiently large number of words for each list in Chinese. Four control lists (Cn) were obtained by randomly choosing one word from each set in each of the C1, C2, S1 and S2 lists. Each control list, therefore, consisted of 32 two-character words with no two characters nor two sounds in common.

Design and procedure

The design was a $2 \times 5 \times 2$ factorial. The first factor refers to modality, and the second to type of list, that is, C1, C2, S1, S2 or Cn. These two were between-subjects factors. The last factor was within-subjects, and refers to trial (trial 1 or 2). Excluding the Cn list, the design can be considered as a $2 \times 2 \times 2 \times 2$ factorial.

The subjects were randomly assigned to ten groups of 24 subjects each according to the order of appearance in the laboratory, and tested in a subgroup of six subjects at a time. The ten groups will be referred to as Groups C1-A, C1-V, C2-A, C2-V, S1-A, S1-V, S2-A, S2-V, Cn-A and Cn-V. The group notations stand for type of list and input mode (auditory or visual).

In the visual mode, a list of 32 items was presented at a speed of 2.5 s per word through a computer terminal. A practice list consisting of four words preceded a test list to familiarize the subjects with the procedure. After going through a test list for one trial, eight addition problems were presented at a speed of 3 s per problem before a recall test. The subjects wrote down an answer to each two-digit addition problem within 3 s, and then as many items as they could remember on the same sheet of paper within a 5-min period. The subjects received a second trial in the same way. The order of items within a list was always randomized. In the auditory mode, the same procedure was used except that the list items were presented by a speaker.

Results and discussion

The mean number of words correctly recalled on each trial for each condition is

Table 6.   *Mean number of words correctly recalled as a function of type of list, input mode, and trial*

Type of list	Input mode			
	Visual		Auditory	
	Trial 1	Trial 2	Trial 1	Trial 2
C1	17.50	26.04	17.58	24.92
C2	23.13	29.17	21.21	27.54
S1	13.92	22.29	10.88	18.63
S2	13.46	23.42	13.17	20.83
Cn	13.83	22.25	12.38	20.17

presented in Table 6. An analysis of variance showed that more words were recalled in the visual than in the auditory input mode (20.50 vs. 18.73), $F(1, 230) = 13.16$, $p < .001$. Type of list was a significant source of variation, $F(4, 230) = 46.40$, $p < .001$. The modality by type of list interaction was not significant, $F < 1$, indicating that the modality effect was quite uniform with respect to type of list. Trial was significant (15.71 vs. 23.53), $F(1, 230) = 1241.39$, $p < .001$ and interacted with modality ($F(1, 230) = 4.03$, $p < .05$) and type of list ($F(4, 230) = 3.86$, $p < .01$). The trial by modality interaction indicates that the modality effect became larger on the second than on the first trial. The trial by type of list interaction indicates that the differences among lists became smaller on the second than on the first trial.

When the data obtained from the control lists were excluded from consideration, an analysis of variance performed on the remaining data from all the experimental lists showed that visual presentation superiority was again obtained (21.12 vs. 19.35), $F(1, 184) = 11.72$, $p < .001$, that the character-same condition produced higher recall than the sound-same condition (23.39 vs. 17.08), $F(1, 184) = 148.94$, $p < .001$, and that recall was higher for the identical character or sound in the second position than in the first position (21.49 vs. 18.97), $F(1, 184) = 23.75$, $p < .001$. The interaction between type of identical unit and its position was significant, $F(1, 184) = 5.65$, $p < .025$, indicating that recall was higher for the second characters being identical than for the first characters being identical, but that such tendency was not obtained in the case of sounds being identical.

The present results may be explained mainly in terms of the effect of category clustering on recall, because the case of the second characters being identical generally produces better categories than the case of the first characters being identical. It is known that the first characters of two-character words usually function as adjectives, as in *orchid-flower*, and the second characters as nouns. Although not significant, the observation that recall of the S1 lists tended to be inferior to recall of the control lists only in the auditory condition supports Turnage and McGinnies' (1973) conjecture that the auditory inferiority for logographs was produced by the script system containing more characters with

similar sounds but different meanings. Their view is, however, inconsistent with the finding that the modality effect was also obtained for the control lists in which no two characters shared the same sound.

Experiment 6

Experiments 1–5 tested the hypotheses that some unique characteristics of Chinese logographs are responsible for their visual input superiority in recall. A more direct test of all these hypotheses would be to have Chinese subjects study lists of English words. If the visual input superiority disappears in this case, it is worthwhile to search further for some characteristics of Chinese logographs responsible for the modality effect. If the visual superiority effect is still obtained, then some general factor other than scriptal factors may be responsible for the effect.

Method

Subjects and list materials

Fifty students enrolled in an introductory psychology course at the Chinese University of Hong Kong participated in the experiment for fulfilling a course requirement.

Four lists of nine English words each were randomly selected from Kucera and Francis (1967), with the restrictions that their frequency count was above 100 and that they were of two syllables. These restrictions were imposed for the sake of minimizing ambiguity when words were presented auditorily.

Procedure

The subjects were tested in small groups of four to six subjects. Following a short practice list, half the subjects studied two lists presented visually and then two lists presented auditorily. The order was reversed for the remaining half of the subjects. A list of nine words was presented through an overhead projector visually or spoken by the experimenter at a speed of 1.5 s per word. Following presentation of a list the subjects wrote down all words they could recall in a free-recall format.

Results and discussion

The serial position functions obtained in the visual and auditory input modes are shown in Figure 1. It is clear from the figure that visual presentation was superior
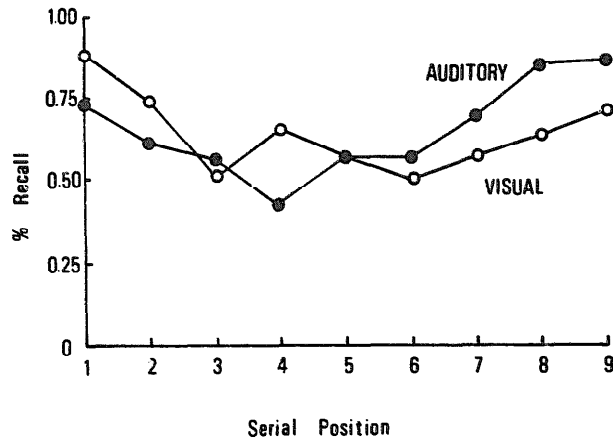
Figure 1.   *Recall as a function of input mode and serial position.*

to auditory presentation for primacy items but inferior for recency items. An analysis of variance showed that modality was not significant, $F < 1$, but interacted with serial position, $F(8, 392) = 5.08$, $p < .001$. Serial position was significant, $F(8, 392) = 11.49$, $p < .001$. Planned comparisons using a t statistic showed that the visual presentation superiority was obtained for the first two primacy items ($t(392) = 2.33$, $p < .05$ and $t(392) = 2.02$, $p < .05$, respectively), but the auditory presentation superiority was obtained for the last two recency items ($t(392) = 3.41$, $p < .01$ and $t(392) = 2.48$, $p < .02$, respectively). This pattern of results, which is similar to that obtained by Fang (1982) using Chinese logographs, has been frequently obtained in our laboratory, when Chinese two-character words were used.

The finding from Western subjects that modality differences at the beginning serial positions are small or non-existent is well documented (e.g., Murdock & Walker, 1969, for free recall; Conrad & Hull, 1968, for serial recall). It is only recently that both auditory and visual superiority effects were obtained from pre-recency items when lists of English words were studied by English-speaking subjects. It may be concluded, however, that the visual superiority effect is rather rarely obtained for English words from English-speaking subjects. In contrast to these findings, the present result of superior visual presentation for the primacy items obtained with English words from Chinese subjects is surprising, if the effect turns out to be robust. This result, however, supports the differential-frequency interpretation of the visual superiority effect, because Chinese students have experienced English more in print than in speech.

Since most Hong Kong college students had studied English as a subject in a primary school for six years and been trained in an English medium secondary school for seven years, an alternative explanation of the present findings in terms of representation of linguistic information in fluent bilinguals has to be consid-

ered. According to the single-code (concept mediation) models (e.g., Potter, So, Von Eckart, & Feldman, 1984), bilinguals represent words in a supralinguistic code that is independent of the language in which the words occurred. The visual superiority effect obtained with English words by Chinese subjects is then explained by the single-code models, because English words and Chinese words have common conceptual representations. There are two problems for this type of explanation. First, the modality effect in this experiment was obtained within one language (English) and a review by Kirsner and Dunn (1985) showed that the amount of repetition priming is virtually nil across a change in language. Second, although the single-code models may be able to account for the modality effect of one language through that of another language, it will be difficult to explain why the modality effect arises in the latter.

Experiment 7

As is well known, evidence from a single experiment is hardly convincing. It is desirable to test Chinese subjects with a long-term recall procedure by presenting long lists of English words. The present experiment also attempted to replicate the results obtained with Chinese words in Experiment 5 by using English words. Because of the morphological differences, it would be hardly possible to manipulate the first and second morphemes of two-morpheme English words as in Experiment 5. It is not difficult, however, to find a sufficient number of English words with their first syllables being identical and a sufficient number of English words with their final syllables being identical. The purpose of Experiment 7 was, then, to find out how input mode affects long-term recall of English words with their first or final sounds being identical.

Method

Subjects and list materials

Thirty-four students enrolled at the Chinese University of Hong Kong served in the experiment voluntarily.

Two types of list, S1 and S2, were constructed as follows. There were four S1 lists, each consisting of four sets of four two-syllable English words each, starting with the same consonant–vowel pronunciation, for example, *meaning*, *meantime*, *meeting* and *meanwhile*. There were also four S2 lists, each consisting of four sets of four two-syllable English words each, ending with the same vowel or vowel–consonant pronunciation, for example, *publish*, *furnish*, *vanish* and *cherish*. Since it was difficult to find a sufficient number of two-syllable words satisfying the

condition of ending with the same vowel or vowel–consonant pronunciation, several three-syllable words were included in the S2 lists, such as *momentum*, *minimum*, *pendulum* and *platinum*. A pilot study showed that all college students could recognize the words selected for the S1 and S2 lists.

Since there were altogether 16 sets of four initial-sound-same words within the four S1 lists, it was possible to construct four S1 control lists by picking up one word each from each of the 16 sets. Similarly, four S2 control lists were constructed from the four S2 lists.

Design and procedure

A $2 \times 2 \times 2$ factorial design was used. The first factor was a between-subjects variable (control vs. experimental lists); the second, a within-subjects variable (visual vs. auditory input); and the third, also a within-subjects variable (type of list, i.e., S1 vs. S2 lists).

The subjects were assigned to the control and experimental groups according to the order of appearance in the laboratory. They were tested in subgroups of 4–5 subjects. The subjects in the experimental group studied the four S1 lists and four S2 lists, while those in the control group studied the four S1 control lists and four S2 control lists. Half the lists of each type were presented visually through an overhead projector and the remaining half spoken by the experimenter. Each list was presented at a speed of 2 s per word. The order, and lists selected, for visual and auditory presentations were systematically counterbalanced between subgroups. Following each list presentation, the subjects solved eight two-digit addition problems for 24 s (3 s per problem) before writing down all words they could remember within a 4-min period.

Results and discussion

The mean number of English words correctly recalled is presented as a function of list type under each input condition in Table 7. An analysis of variance showed that input mode was significant (visual 14.02 vs. auditory 10.93), $F(1, 32) = 17.33$, $p < .001$, and did not interact with group, $F < 1$. More items were recalled from

Table 7.   *Mean number of English words correctly recalled*

Input mode		Experimental list	Control list
Visual	S1	14.65	14.53
	S2	14.00	12.88
Auditory	S1	12.24	10.41
	S2	9.71	11.35

the S1 lists than from the S2 lists (12.96 vs. 11.99), $F(1, 32) = 6.47$, $p < .05$, probably because the latter contained longer words and thus were harder to recall. Recall of items in the sound-clustered lists did not differ from recall of items in the sound-distinctive lists, $F < 1$, apparently because the clustered sounds did not carry any similar meaning.

The present experiment essentially replicates the major finding of Experiment 5 that visual presentation superiority was obtained in long-term recall. The finding that the same initial or final sounds did not produce higher recall also replicates the observation of Experiment 5 and is in line with the view that long-term recall is based on searching for semantic rather than phonological retrieval cues.

Experiment 8

If the visual presentation superiority cannot be explained in terms of any characteristic of a script, what is the underlying mechanism responsible for it? A possible mechanism would be frequency operated through the long-term priming (Jacoby & Dallas, 1981; Kirsner & Dunn, 1985; Scarborough et al., 1977). According to this frequency interpretation, the visual presentation superiority is obtained because the frequency of experiencing some items in the visual input mode is higher than that of experiencing the same items in the auditory input mode.

The differential-frequency interpretation can account for the visual presentation superiority of Chinese logographs (Turnage & McGinnies, 1973) by noting that there is a discrepancy in kinds of speech many Chinese people use in school and at home. The type of speech (Mandarin) used at formal occasions is different from a dialect used at home for many Chinese people. Moreover, although there is evidence for speech recoding of Chinese logographs in reading (e.g., Tzeng, Hung, & Wang, 1977), it may be less prevalent in Chinese readers than in English readers because the former are often not confident about how to pronounce low-frequency characters correctly. For instance, Zhou (1978) estimated that the relationship between graphic cues and sounds in actual characters is very low (30%).

The finding of superior long-term recall of visually presented over auditorily presented English words obtained in Experiments 6 and 7 with Chinese subjects can be easily accounted for by the frequency interpretation as follows. For most Chinese people, English is after all a foreign language. They access English more frequently in print than from speech. They also learn to retrieve English words more from English readings than from English speech. A natural consequence is that the visual presentation superiority will be obtained for Chinese subjects.

The idea that visual frequency is different from auditory frequency may be basic to the effect of associative frequency or availability on recall. The concept of

availability is defined as the relative ease with which an isolated word comes to mind. It is operationalized in terms of associative frequency, which is measured by the number of times a word is given as an associate to a sample of stimulus words. Although the variable of availability is long known to affect recall (Asch & Ebenholtz, 1962; Dale, 1967; Deese, 1965; Leicht, 1968; Tversky & Kahneman, 1973), it is only through a series of studies by Rubin (1980, 1983) and Rubin and Friendly (1986) that availability is found to be one of the best predictors of which words are best recalled.

It is not difficult to see that availability directly reflects auditory frequency, because availability is defined in terms of how many times a word is produced by a group of persons to a set of stimulus words. However, the latter is simply a measure of auditory frequency, if the set of stimulus words is sufficiently large. It is then clear why visual frequency defined by how many times a word appears in print is so ineffective in predicting recallability (Rublin & Friendly, 1986). The fact is that whenever there is a discrepancy in word frequency and auditory frequency, the former becomes less important because words in print are after all produced in a less straightforward way than speech.

The Chinese people in Hong Kong speak one of the seven major dialect groups in Chinese (Cantonese), which differs from the standard Chinese (Mandarin). There are some Cantonese words that appear neither in dictionaries of standard Chinese nor in formal readings. They are written and pronounced uniquely, and appear only in local newspapers and magazines. We will refer to them as dialectic forms. Therefore, there are two types of Chinese words: (a) one type of word that can be written in two different forms, that is, dialectic and standard forms, and (b) another type of word that can be written only in a standard form, that is, the written form is the same whether one writes in standard Chinese or in a dialect. The majority of the Chinese words belong to the latter type. With respect to those words in double form, they are used in a dialectic form more frequently in everyday speech than in print.

Let us consider those words that can be written in the two different forms. Since a word can be written in the two different forms, relative to the other type of word, the long-term priming effect will accumulate for the latter but not for the former. Consequently, the visual superiority effect should be observed from those words that can be written only in a single form, but not from those words that have two different forms. Experiment 8 was designed to test this prediction.

Method

Subjects and list items

Sixty freshmen, enrolled in an introductory psychology course at the Chinese

University of Hong Kong, participated in the experiment for fulfilling a course requirement. All of them can speak fluent Cantonese.

Forty two-character words belonging to a Cantonese dialect were chosen. They are frequently spoken in everyday speech, and written in a double form: a dialectic form and a standard form. Therefore, two lists of 20 words each were constructed in dialectic forms, and by rewriting these 40 words in standard forms there resulted in two lists of 20 words each in standard forms.

Another set of 80 two-character words were selected from Huang and Liu (1978). They can be used only in standard forms, and will be referred to as single-form words. Half the words (40) had high imagery values of about 5.0 on a 6-point rating scale, and the remaining half had low imagery values of below 2.5. By randomly dividing each set of 40 words into two halves, two lists of 20 words each were constructed.

Procedure

The subjects were randomly divided into two groups according to the order of appearance in the laboratory, and tested in a small group of 4–6 subjects at a time. Following two short practice lists, one in the visual and another in the auditory input mode, the subjects in one group studied (a) two double-form lists of words in a standard form, and (b) two single-form lists of high-imagery words, one of each pair of lists in the visual input mode and another in the auditory input mode. The subjects in another group studied (a) two double-form lists of words in a dialectic form, and (b) two single-form lists of low-imagery words, also one of each pair of lists in the visual input mode and another in the auditory input mode. The order of administrating four lists was systematically counterbalanced between small groups.

Each list of 20 words was presented visually through an overhead projector or spoken in the auditory input mode by the experimenter at a speed of 1.5 s per word. Following presentation of an entire list, the subjects solved eight two-digit number addition problems as in Experiment 7 before writing down all they could remember within a 4-min period.

Results and discussion

The mean number of words correctly recalled for each type of list is presented in Table 8. As is clear from the table, more high-imagery words were recalled than low-imagery words irrespective of the visual or auditory input mode. Our main interest is, however, in relative magnitudes of the modality effect obtained for the double-form and single-form words. An analysis of variance showed that modality was significant (9.50 vs. 8.06), $F(1, 58) = 5.35$, $p < .025$. The interaction between

Table 8.   *Mean number of words correctly recalled in each condition of word form*

Modality	Double form		Single form	
	Standard	Dialect	High imagery	Low imagery
Visual	10.03	9.50	10.10	8.37
Auditory	8.93	8.87	8.07	6.37

modality and type of word (double vs. single form) was also significant, $F(1, 58) =$ 4.85, $p < .05$, indicating that the visual presentation superiority was larger for single-form words than for double-form words. With respect to single-form words, planned comparisons using a t statistic showed that the visual superiority effect was obtained for high-imagery and low-imagery words, $t(58) = 3.40$, $p < .01$ for the former and $t(58) = 3.35$, $p < .01$ for the latter. As for double-form words, the visual superiority effect was not significant, $t(58) = 1.84$, $p > .05$ for standard words and $t(58) = 1.06$, $p > .20$ for dialectic words, although the auditory superiority effect was originally expected for the latter words.

The finding that the visual superiority effect was significant only for single-form words supports the frequency interpretation that the effect was produced by a large discrepancy between visual and auditory frequencies. There is a question as to why the auditory superiority effect could not be obtained in spite of the fact that some words such as dialectic double-form words are experienced more frequently in the auditory than in the visual mode. A possible answer is as follows.

Let us call those words that are experienced more frequently in the auditory than in the visual mode *auditory* words. Similarly, we have *visual* words. Now, in order to demonstrate the auditory presentation superiority, we have to compare the condition of presenting *auditory* words visually with the condition of presenting *auditory* words auditorily. When *auditory* words are presented visually, however, they will be recoded phonologically (auditorily). Therefore, presenting *auditory* words visually does not differ significantly from presenting them auditorily, resulting in the absence of the auditory superiority effect.

The situation is entirely different when *visual* words are presented auditorily. In this case, auditorily presented *visual* words will not leave characteristic visual traces which are the consequence of presenting *visual* words visually.

Experiment 9

This experiment considers only single-form words: the words written in a single form without an additional dialectic form. There are many words that frequently appear in school textbooks but rarely in newspapers, such as *observation, length,*

procedure, measure, etc. There are also many words that frequently appear in newspapers but rarely in school textbooks, such as *police, telephone, market, game,* etc. Relatively speaking, the former may be considered as high in visual but low in auditory frequency, while the latter may be considered as high in auditory but low in visual frequency. The frequency interpretation predicts that the visual superiority effect should appear in the former but not in the latter.

Method

Subjects and list items

The subjects were 38 freshmen enrolled in an introductory psychology course at the Chinese University of Hong Kong. They served in the experiment for fulfilling a course requirement.

Forty textbook words and 40 newspaper words were selected from Siu and the Chinese Vocabulary Research Committee (1986) as follows. In this word book, for each word, frequency counts are separately obtained from textbook and outside reading materials selected on the basis of a readings survey of Hong Kong secondary school students. The counts are based on approximately 400,000 characters of textbook passages most frequently used and also on approximately 400,000 characters of outside readings consisting mainly of newspapers and magazines. For the sake of exposition, the former will be referred to as a textbook frequency count, and the latter as a newspaper frequency count.

In the selection procedure, 60 pairs of two-character words were selected, starting from the word with the highest total frequency. Each pair was matched on the basis of total frequency counts, with one member high in textbook frequency but low in newspaper frequency and another member low in textbook frequency but high in newspaper frequency. Next, another pool of 15 subjects rated each of the 60 pairs of words with respect to imagery values according to a 6-point scale. Finally, 40 pairs of words were selected by equating approximately for their mean total frequencies and mean imagery values, although the results of Experiment 8 indicated that the effect of imagery on the visual presentation superiority seems very small.

The 40 textbook words (high in textbook frequency but low in newspaper frequency) were randomly divided into halves to obtain two lists of 20 textbook words each. The 40 newspaper words were also randomly divided into halves to obtain two lists of 20 newspaper words each. Table 9 presents the mean textbook and newspaper frequency counts for each type of word.

Design and procedure

The design was a 2×2 within-subjects factorial. The first factor refers to

Table 9.   *Mean frequency count and mean number of words correctly recalled*

Type of word	Type of frequency count		Number of words recalled	
	Textbook	Newspaper	Visual	Auditory
Textbook	84.2	10.6	8.45	7.34
Newspaper	15.3	76.5	9.63	9.03

presentation modality, and the second to type of word (list of textbook words or list of newspaper words). Half the subjects received two lists in the visual input mode and then two lists in the auditory input mode. The order was reversed for the remaining half. The order of receiving type of list was counterbalanced between subjects. All other experimental details were the same as in Experiment 8.

Results and discussion

The mean number of words correctly recalled under each condition is presented in the last two columns of Table 9. An analysis of variance showed that the visual superiority effect was obtained, $F(1, 37) = 7.24$, $p < .025$. More newspaper words were recalled than textbook words, $F(1, 37) = 23.40$, $p < .001$. The interaction between input mode and type of word was not significant, $F < 1$. Planned comparisons using a t statistic showed that the visual superiority effect was obtained for textbook words, $t(37) = 2.47$, $p < .02$, but not for newspaper words, $t(37) = 1.34$, $p > .10$, in support of the proposed frequency interpretation.

Another noteworthy finding from the present experiment is that more newspaper words were recalled than textbook words, although the mean total frequency counts and mean imagery values were approximately equated. This finding points out the ineffectiveness of the classical frequency count in accounting for recallability, and supports Rubin's (1980, 1983) finding that differential availability is an important factor determining word recallability. As is evident from the first two columns of Table 9, the present experiment further shows that it is a large discrepancy in the visual and auditory frequencies that produced the observed difference in recallability.

Experiment 10

The Beijing dialect, Mandarin, is adopted as the national language, and used at formal occasions and in schools. The differential-frequency account of the long-

term modality effect, then, predicts that Beijing natives who speak Mandarin at home will show at most a weak visual superiority effect in comparison with many Chinese college students who speak their own dialect at home. A strong visual superiority effect will be expected from the latter subjects, because the long-term priming effect for speech does not accumulate in this case so well as for the former subjects.

The present pool of the Beijing subjects was different in composition from that used in Experiment 5. This is because, as one of the most prestigious universities in China, Peking University admits secondary school graduates from all over the country. Therefore, the majority of Peking University students are not Beijing natives. In addition, not all Beijing natives speak Mandarin at home. If their ancestors came from other provinces, their children tend to retain the habit of speaking their former dialect, even though they were born in Beijing. Although it varies from year to year, it is estimated that less than 30% of Peking University students are Beijing natives who speak Mandarin at home.

Method

Subjects and materials

The subjects were 24 Beijing natives and 24 Fukien (southern China) natives. The former subjects were enrolled at Peking University, speaking Mandarin at home. The latter subjects were enrolled at Fukien Normal University, speaking a southern Chinese dialect at home. They participated in the experiment voluntarily. Four lists of two-character words were selected from those used in Experiment 5. Each list consisted of 32 words. There was no character shared by two words.

Design and procedure

A 2 × 2 factorial design was used. The first factor refers to modality, and the second to type of dialect used at home. The first factor was a within-subjects variable, while the second a between-subjects variable.

The same procedure of Experiment 5 was used except the following. First, instead of presenting arithmetic problems before a recall test, the subjects wrote down as many items as they could remember immediately after presentation of a list. However, to remove those items that could be recalled from short-term memory, the last five recency items were excluded from consideration. Second, the order of presenting the four lists, two in the visual mode and two in the auditory mode, was counterbalanced within subjects and balanced between subjects. Third, in the auditory mode the same tape recording of a female voice in Mandarin was used for both Beijing and Fukien subjects.

Results and discussion

For the Beijing subjects, the mean percentages of items correctly recalled were 0.49 in the visual condition and 0.46 in the auditory condition. For the Fukien subjects, the means were 0.54 and 0.41 in the visual and auditory conditions, respectively. An analysis of variance showed that the visual superiority effect was significant (0.52 vs. 0.44), $F(1, 46) = 23.92, p < .001$. The Beijing subjects did not differ significantly from the Fukien subjects (0.48 vs. 0.48), $F < 1$. The interaction between modality and type of dialect was significant, $F(1, 46) = 9.16, p < .01$, indicating that the visual superiority effect was stronger for the Fukien than for the Beijing subjects. For the Beijing subjects, a planned comparison showed that the visual superiority effect was not significant (0.49 vs. 0.46), $t(46) = 1.32$, $p > .10$. For the Fukien subjects, the effect was significant (0.54 vs. 0.41), $t(46) = 5.60, p < .001$. In the auditory condition, a planned comparison showed that the Fukien subjects' performance was not significantly lower than for the Beijing subjects (0.41 vs. 0.46), $t(46) = 1.05, p > .25$. This can be used to rule out an explanation of the visual superiority effect obtained with the Fukien subjects by assuming that their comprehension of Mandarin was poor in the auditory condition.

General discussion

An interest in the visual superiority effect for Chinese logographs originated from McGinnies' (1965) study of persuasion through printed versus spoken communication. Since then the consistent visual superiority effect has been well documented for Chinese logographs (Fang, 1982; Hue et al., 1990; Turnage & McGinnies, 1973; Tzeng & Wang, 1983). For alphabetic words, however, there has been no evidence for consistent long-term modality effects. Thus, although Engle and Mobley (1976) found no modality effects for non-recency items, Conway and Gathercole (1987; Gathercole & Conway, 1988) have reported the auditory superiority effect in a long-term recognition memory task. More recently, using one-syllable common English words, Penny (1989b) obtained the visual superiority effect in both delayed free recall and delayed recognition tests.

To account for the consistent visual superiority effect obtained with Chinese logographs, an interpretation in terms of script differences appeared in one form or another. Experiments 1–5 tested the three forms of the differential script interpretation: the direct-image hypothesis, the discriminability hypothesis, and the graphic-feature hypothesis. None of these hypotheses was found to be capable of explaining the consistent visual superiority effect for Chinese logographs.

That Chinese logographs do not directly give rise to images and are no more visually unique than English words is compatible with the observation that, when

words are used more frequently, they tend to acquire more meanings (Miller, 1951; Johnson-Laird, 1983). This means that meanings of words/characters are acquired and come to be dissociated from the way in which words/characters were originally created. To Chinese linguists who are well aware of the origins of characters, the character for "horse" triggers a vivid image of an abstract figure galloping across the page. However, to most Chinese such a vivid image will not be aroused.

If a set of verbal items experienced repeatedly in one modality can make the traces in that modality more available for retrieval through the mechanism of long-term priming (e.g., Jacoby & Dallas, 1981; Kirsner & Dunn, 1985; Scarborough et al., 1977), then this differential-frequency interpretation is capable of predicting the following. First, as a second language for Chinese students, English is experienced more in print than in speech, and as such the consistent visual superiority effect should be obtained with English words for Chinese subjects. Second, if a set of words is experienced more frequently in print than in speech and another set of words is experienced less frequently in print than in speech, then the consistent visual superiority effect should be obtained from the former set of words, but not necessarily from the latter set of words. Third, if a group of Chinese subjects speak a non-Mandarin dialect at home, then a stronger visual superiority effect should be observed from this group than from a group of Beijing natives who speak Mandarin at home, because both groups use Mandarin to read the same script. These predictions have been confirmed in Experiments 6-10.

The differential-frequency account also explains why the visual superiority effect can be consistently obtained for Chinese subjects, but not for Western subjects. Every Chinese college student is in general capable of speaking Mandarin and at least one dialect that is significantly different from Mandarin. Since the Chinese script is essentially the same for various Chinese dialects, the visual superiority effect will be obtained through the operation of long-term visual priming. On the other hand, Western subjects are much the same as Beijing natives who use Mandarin at home as in school and in reading, and there will be no consistent visual superiority effect.

In conclusion, although it is tempting to attribute the observed consistency or inconsistency of the visual superiority effect to script differences, a much more general principle may apply to all languages to account for the diversified findings at a deeper level of analysis.

References

Aaronson, D., & Ferres, S. (1986). Sentence processing in Chinese-American bilinguals. *Journal of Memory and Language*, 25, 136-162.

Ambler, B.A., & Proctor, J.D. (1976). The familiarity effect for single-letter pairs. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 222–234.

Asch, S.E., & Ebenholtz, S.M. (1962). The principle of associative symmetry. *Proceedings of the American Philosophical Society*, 106, 135–163.

Biederman, I., & Tsao, Y.C. (1979). On processing Chinese ideographs and English words: Some implications from Stroop-test results. *Cognitive Psychology*, 11, 125–132.

Cheng, C.M. (1982). Analysis of present-day Mandarin. *Journal of Chinese Linguistics*, 10, 281–358.

Conrad, R., & Hull, A.J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science*, 10, 135–136.

Conway, M.A., & Gathercole, S.E. (1987). Modality and long-term memory. *Journal of Memory and Language*, 26, 341–361.

Corballis, M.C. (1966). Rehearsal and decay in immediate recall of visually and aurally presented items. *Canadian Journal of Psychology*, 20, 43–51.

Craik, F.I.M. (1969). Modality effects in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 658–664.

Crowder, R.G. (1976). *The principles of learning and memory*. Hillsdale, NJ: Erlbaum.

Crowder, R.G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception and Psychophysics*, 5, 365–373.

Dale, H.C.A. (1967). Response availability and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 6, 47–48.

Deese, J. (1965). *The structure of associations in language and thought*. Baltimore: John Hopkins Press.

Egeth, H., & Blecker, D. (1971). Differential effects of familiarity on judgments of sameness and differences. *Perception and Psychophysics*, 9, 321–326.

Engle, R.W., & Mobley, L.A. (1976). The modality effect: What happens in long-term memory? *Journal of Verbal Learning and Verbal Behavior*, 15, 519–528.

Fang, S.P. (1982). Interaction between mode of presentation and serial position in serial recall of Chinese characters and words. In H.S.R. Kao & C.M. Cheng (Eds.), *Psychological studies of the Chinese language* (pp. 227–243). Taipei: Wen-Ho.

Frege, G. (1960). On sense and reference. In P. Geach & M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege* (pp. 56–78). Oxford: Basil Blackwell.

Gardiner, J.M. (1983). On recency and echoic memory. *Philosophical Transactions of the Royal Society of London*, B302, 267–282.

Gardiner, J.M., & Gregg, V.H. (1979). When auditory memory is not overwritten. *Journal of Verbal Learning and Verbal Behavior*, 18, 705–719.

Gathercole, S.E., & Conway, M.A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory and Cognition*, 16, 110–119.

Glenberg, A.M. (1984). A retrieval account of the long-term modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 231–255.

Glenberg, A.M., & Swanson, N.G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 3–15.

Guenther, R.K., & Klatzky, R.L. (1977). Semantic classification of pictures and words. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 498–514.

Guenther, R.K., Klatzky, R.L., & Putnam, W. (1980). Commonalities and differences in semantic decisions about pictures and words. *Journal of Verbal Learning and Verbal Behavior* 19, 54–74.

Hochberg, J., & Brooks, V. (1962). Pictorial recognition as an unlearned ability. *American Journal of Psychology*, 75, 624–628.

Hovland, C.I. (1954). The effects of mass media of communication. In G. Lindzey (Ed.), *Handbook of social psychology* (pp. 1062–1103). Cambridge, MA: Addison-Wesley.

Huang, J.T., & Liu, I.M. (1978). Paired-associate learning proficiency as a function of frequency count, meaningfulness, and imagery value in Chinese two-character ideograms. *Acta Psychologica Taiwanica*, 20, 5–17.

Hue, C.W., Fang, D., & Hsu, K.Y. (1990). Immediate serial recall of Chinese characters: A study of input modality effect. *Acta Psychologica*, 75, 213–223.

Jacoby, L.L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340.

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge, MA: Harvard University Press.

Kirsner, K., & Dunn, J. (1985). The perceptual record: A common factor in repetition priming and attribute retention? In M.I. Posner & O.S.M. Marin (Eds.), *Attention and performance XI* (pp. 547–565). Hillsdale, NJ: Erlbaum.

Klapper, J.T. (1961). *The effects of mass communication.* Glencoe, IL: Free Press.

Kolers, P. (1979). Reading and knowing. *Canadian Journal of Psychology, 33,* 106–117.

Kucera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Leicht, K.L. (1968). Recall and judged frequency of implicitly occurring words. *Journal of Verbal Learning and Verbal Behavior, 7,* 918–923.

Li, C.N., & Thompson, S.A. (1981). *Mandarin Chinese: A functional reference grammar.* Berkeley and Los Angeles, CA: University of California Press.

Liu, I.M., Chuang, C.J., & Wang, S.C. (1975). *Frequency count of 40,000 Chinese words.* Taipei, Taiwan: Lucky Books.

McGinnies, E. (1965). A cross-cultural comparison of printed communication versus spoken communication in persuasion. *Journal of Psychology, 60,* 1–8.

Miller, G.A. (1951). *Language and communication.* New York: McGraw-Hill.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76,* 165–178.

Murdock, B.B., Jr., & Walker, K.D. (1969). Modality effects in free recall. *Journal of Verbal Learning & Verbal Behavior, 8,* 665–676.

Murray, D.J. (1966). Vocalization-at-presentation and immediate recall, with varying recall methods. *Quarterly Journal of Experimental Psychology, 18,* 9–18.

Nairne, J.S. (1988). A framework for interpreting recency effect in immediate serial recall. *Memory and Cognition, 16,* 343–352.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart & Winston.

Paivio, A. (1986). *Mental representations: A dual coding approach.* New York: Oxford University Press.

Paivio, A., & te Linde, D.J. (1980). Symbolic comparisons of objects on color attributes. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 652–661.

Park, S., & Arbuckle, T.Y. (1977). Ideograms versus alphabets: Effect of script on memory in "biscriptural" Korean subjects. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 631–642.

Pellegrino, J.W., Rosinski, R.R., Chiesi, H.L., & Siegel, A. (1977). Picture-word differences in decision latency: An analysis of single and dual memory models. *Memory and Cognition, 5,* 383–396.

Penney, C.G. (1989a). Modality effects and the structure of short-term verbal memory. *Memory and Cognition, 17,* 398–422.

Penney, C.G. (1989b). Modality effects in delayed recall and recognition: Visual is better than auditory. *Quarterly Journal of Experimental Psychology, 41A,* 455–470.

Potter, M.C., & Faulconer, B.A. (1975). Time to understand pictures and words. *Nature, 253,* 437–438.

Potter, M.C., So, K.-F., Von Eckart, B., & Feldman, L.B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23,* 23–38.

Rubin, D.C. (1980). 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior, 19,* 736–755.

Rubin, D.C. (1983). Associative asymmetry, availability, and retrieval. *Memory and Cognition, 11,* 83–92.

Rubin, D.C., & Friendly, M. (1986). Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronuncability for 925 nouns. *Memory and Cognition, 14,* 79–94.

Scarborough, D.L., Cortese, C., & Scarborough, H.S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 1–17.

Siu, P.K. and the Chinese Vocabulary Research Committee (1986). *A study of Chinese vocabulary for Hong Kong secondary school students*. Hong Kong: Hong Kong Education Department.

Smith, E.E., Bolzano, G.J., & Walker, J. (1978). Nominal, perceptual, and semantic codes in picture categorization. In J.W. Cotton & R.L. Klatzky (Eds.), *Semantic factors in cognition* (pp. 137–168). Hillsdale, NJ: Erlbaum.

te Linde, D.J. (1983). Pictures and words in semantic decisions. In J.C. Yuille (Ed.), *Imagery, memory and cognition* (pp. 117–129). Hillsdale, NJ: Erlbaum.

Thorndike, E.L., & Lorge, I. (1944). *The teacher's wordbook of 30,000 words*. New York: Teachers College Press.

Tung Lai (1978). *Recognizing Chinese characters*. Taipei: Mingjen.

Turnage, T.W., & McGinnies, E. (1973). A cross-cultural comparison of the effects of presentation mode and meaningfulness in short-term recall. *American Journal of Psychology*, *86*, 369–381.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232.

Tzeng, O.J.L., & Hung, D.L. (1988). Cerebral organization: Clues from scriptal effects on lateralization. In I.M. Liu, H.C. Chen, & M.J. Chen (Eds.), *Cognitive aspects of the Chinese language* (Vol. 1, pp. 119–139). Hong Kong: Asian Research Service.

Tzeng, O.J.L., Hung, D.L., & Wang, W.S.Y. (1977). Speech recoding in reading Chinese characters. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 621–630.

Tzeng, O.J.L., & Wang, W.S.Y. (1983). The first two R's. *American Scientist*, *71*, 238–243.

Wang, W.S.Y. (1973). The Chinese language. *Scientific American*, *228*, 50–60.

Yeh, J., & Liu, I.M. (1972). Factors affecting recognition thresholds of Chinese characters. *Acta Psychologica Taiwanica*, *14*, 26–37.

Zhou, Y.G. (1978). To what degree are the "phonetics" of present-day Chinese characters still phonetic? *Zhongguo Yuwen*, *146*, 172–177.