Research Article

SPONTANEOUS ATTENTION TO WORD CONTENT VERSUS EMOTIONAL TONE: Differences Among Three Cultures

Keiko Ishii,[1] Jose Alberto Reyes,[2] and Shinobu Kitayama[1]

[1]Kyoto University, Kyoto, Japan, and [2]De La Salle University-Manila, Manila, the Philippines

Abstract—*A Stroop interference task was used to test the hypothesis that people in different cultures are differentially attuned to verbal content vis-à-vis vocal tone in comprehending emotional words. In Study 1, Americans showed greater difficulty ignoring verbal content than ignoring vocal tone (which reveals an attentional bias for verbal content); but Japanese showed greater difficulty ignoring vocal tone than ignoring verbal content (which reveals a bias for vocal tone). In Study 2, Tagalog-English bilinguals in the Philippines showed an attentional bias for vocal tone regardless of the language used, suggesting that the effect is largely cultural rather than linguistic. Implications for culture-and-cognition research are discussed.*

Many Americans who interact with Asians such as Japanese and Filipinos for the first time often feel perplexed because their Asian friends do not seem to mean quite the same thing when they say "yes" as Americans mean when they say the same word (e.g., Barnlund, 1989). Conversely, many Asians also feel perplexed because their American friends often fail to "get it." In the current work, we suggest that underlying this occasional mishap in intercultural communications is a cultural variation in spontaneous attention to different aspects of utterances. Whereas Americans attend primarily to verbal content, Asians pay closer attention to vocal tone and other contextual information.

It has been proposed that in many Western, independent cultures and the languages used therein (e.g., European-American cultures and languages such as English), a greater proportion of information is conveyed by verbal content than by contextual cues (Ambady, Koo, Lee, & Rosenthal, 1996; Hall, 1976; Kitayama, 2000; Markus & Kitayama, 1991). Correspondingly, contextual, nonverbal cues such as vocal tone are likely to serve a relatively minor role. Hall (1976) referred to these cultures and languages as low-context. In contrast, in many Asian, interdependent cultures and the languages used therein (e.g., the cultures of Japan, the Philippines, Korea, and China and languages such as Japanese, Tagalog, Korean, and Chinese), the proportion of information conveyed by verbal content is relatively small and, correspondingly, contextual and nonverbal cues are likely to play a relatively larger role. These languages and cultures are called high-context.

These cross-culturally divergent practices of communication are not a superficial overlay on the basic cognitive processes involved in speech comprehension. To the contrary, by routinely participating in different practices of interpersonal communication, individuals are likely to develop correspondingly divergent modes of cognitive processing (Nisbett, Peng, Choi, & Norenzayan, 2001). The low-context practices foster an allocation of attention primarily to verbal content, whereas the high-context practices encourage attention to be allocated more to contextual information. One contextual cue that always exists in speech communications is vocal tone (Kitayama, 1996).

In a recent series of experiments, we used a Stroop task and provided initial evidence that native English speakers spontaneously attend more to verbal content than to vocal tone, whereas native Japanese speakers spontaneously attend more to vocal tone than to verbal content (Kitayama & Ishii, 2002). Both Japanese speakers and English speakers were presented with a number of spoken words, one at a time, in their native languages. These single-word utterances differed in both emotional word meaning and emotional vocal tone. The respondents judged either how pleasant the vocal tone of each utterance was while ignoring its verbal content or how pleasant the verbal content of each utterance was while ignoring its vocal tone.

The results showed cross-culturally divergent patterns of interference in the two judgments. Americans showed a strong interference effect in vocal-tone judgments. The response time for these judgments was much longer if the attendant verbal content was incongruous than if it was congruous. But a comparable interference effect in word-meaning judgments was negligible. This is the evidence that attention was spontaneously allocated to word meaning in lieu of vocal tone. Japanese respondents showed a contrasting pattern. For them, the size of interference was somewhat larger in word-meaning judgments than in vocal-tone judgments.

Although consistent with the hypothesis that Americans are attentionally attuned more to verbal content and Japanese are attuned more to vocal tone, this study was compromised by the particular stimulus materials used. Two issues deserve careful attention. First, the emotional valence of verbal content was more extreme than the emotional valence of vocal tone. If research is to yield unequivocal evidence for an attentional bias that favors either vocal tone or verbal content, it is important for the polarity of emotional verbal content and emotional vocal tone to be equated. Second, the American respondents in the study showed no interference effect in the word-meaning judgment. Taken at face value, this result implies that Americans pay no attention at all to vocal tone. The absence of interference, however, may be attributable to the fact that the vocal tones used in this study were quite weak. It is therefore highly desirable to test both Americans and Japanese with more explicitly emotional tones of voice. We expect that with such materials even Americans might show reliable interference by vocal tone in the word-meaning judgment, although the effect would likely be less strong than for Japanese speakers.

Moreover, there remains a thorny issue of whether the phenomenon we observed was mediated by cultural processes or linguistic processes. In its most classic form, the linguistic relativity hypothesis (Whorf, 1956) posits that individuals' cognition, perception, and worldviews, or, in short, their "culture," are significantly shaped by the

Address correspondence to Keiko Ishii, Department of Behavioral Science, Bungakubu, Hokkaido University, N10W7, Kita-ku, Sapporo 060-0810, Japan, e-mail: ishii@let.hokudai.ac.jp, or to Shinobu Kitayama, Graduate School for Human and Environmental Studies, Kyoto University, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan, e-mail: kitayama@hi.h.kyoto-u.ac.jp.

language they speak. Although the strongest form of linguistic relativity is hardly justified in view of subsequent work (Brown, 1976), some aspects of language may have penetrating effects on one's psychological processes (Lucy, 1992). Alternatively, it may be primarily culture's practices and meanings that foster psychological differences (Kitayama, 2002). According to this view, it is not language per se but rather culture-dependent ways in which language is used that matter. Evidence for this second view comes from a number of recent demonstrations of cultural differences in nonlinguistic cognitions that are in line with the differences in linguistic cognitions (e.g., Kitayama, Duffy, Kawamura, & Larsen, in press; Masuda & Nisbett, 2001). We return to this issue in Study 2.

STUDY 1: ATTENTIONAL BIAS IN JAPAN AND THE UNITED STATES

The purpose of Study 1 was to carry out a more stringent test of the hypothesis on cultural difference in attention. For this purpose, we developed a stimulus set of emotionally spoken emotional words in both English and Japanese, equating the strength of emotional meanings and the strength of emotional vocal tones both within each language and between the two languages. Furthermore, we used several bilingual speakers to create the stimuli in the two languages. In this way, we equated the vocal quality across the languages.

Method

Respondents and procedure

One hundred thirty-four Japanese undergraduates at a Japanese university (all native Japanese speakers, 61 females and 73 males) and 106 American undergraduates at an American university (all native English speakers, 52 females and 54 males) participated in the experiment. Fifteen Japanese and 11 American respondents evidently misunderstood the instructions and failed to achieve chance-level accuracy. Their data were therefore excluded from the following analyses. The data from the remaining 214 respondents are reported here.

Respondents were informed that the study was concerned with the perception of spoken words. They were presented with utterances in their native language and, depending on the assigned judgment condition, instructed to either judge word meaning as pleasant or unpleasant while ignoring the attendant vocal tone or judge vocal tone as pleasant or unpleasant while ignoring the attendant word meaning.

The entire procedure was computerized. The experiment consisted of 32 trials, preceded by 10 practice trials. The order of the experimental trials was randomized for each respondent. On each trial, following a warning signal on the screen, a word was presented through headphones. The respondents were asked to press one of two response keys that corresponded to the two response options. They were asked to respond as quickly as possible without sacrificing accuracy in judgment. Response time was measured in milliseconds from the onset of each stimulus. There was a 1,500-ms interval between trials.

Materials

Stimulus utterances were developed in four steps. First, we prepared 90 pairs of translation-equivalent Japanese and English words, both nouns and adjectives that vary in emotional meaning. We had 25 Japanese and 27 Americans judge both the pleasantness of the meaning of each of the words (1 = *very unpleasant*, 7 = *very pleasant*) and the frequency of each word's appearance in daily life (1 = *not at all*, 5 = *very frequently*). We used the average pleasantness ratings to choose 30 pairs, with 10 in each of three word-meaning conditions (pleasant, neutral, unpleasant).

Second, four Japanese-English bilinguals (two females and two males) were trained to read all the English and Japanese words in each of three distinct tones of voice, namely, a smooth and round tone (pleasant), a business-like tone (neutral), and a harsh and constricted tone (unpleasant). This yielded a total of 720 utterances (see Kitayama, 1996, and Scherer, 1986, for the validity of this manipulation of emotional vocal tones). Two of the authors listened carefully to the utterances and selected for each word in each of the three tone conditions one utterance spoken by a male speaker and another spoken by a female speaker. All the words were pronounced articulately, and emotional vocal tones were quite distinct and clear.

Third, the resulting set of 360 utterances was low-pass filtered at 400 Hz, which preserved basic intonation patterns while making it mostly impossible to discern any semantic meanings. A separate group of 29 Japanese and 29 American undergraduates (both males and females) listened to each of the 360 words either in the original form or in the content-filtered form and rated the pleasantness of the vocal tone of each utterance (1 = *very unpleasant*, 7 = *very pleasant*). The American means and the Japanese means were highly correlated (rs = .83 and .74, ps $< .0001$). Moreover, the ratings for the original utterances and those for the filtered utterances were also highly correlated (rs = .95 and .82, ps $< .0001$).

The final step was to use these ratings to select the final set of 64 utterances: 8 utterances × 2 languages (English and Japanese) × 2 meanings (pleasant and unpleasant) × 2 vocal tones (pleasant and unpleasant) × 1 speaker (see Table 1). As can be seen in Table 2, in the final set of utterances, vocal tone was manipulated independently of culture and word meaning, and vocal tone and word meaning were equally extreme in the two languages.

Results and Discussion

Response time

Overall, responses were quite accurate, with a mean of 95% correct. We first report response times, followed by a discussion on accuracy data. Only correct responses were included in the analysis of

Table 1. *Words used in Study 1*

Word meaning			
Pleasant		Unpleasant	
Japanese	English	Japanese	English
Arigatai	Grateful	Fuman	Complaint
Atarashii	New	Itai	Sore
Atatakai	Warm	Kirai	Dislike
Kirei	Pretty	Mazui	Tasteless
Manzoku	Satisfaction	Shinpai	Anxiety
Ochitsuki	Calmness	Tsukare	Fatigue
Oyatsu	Refreshment	Tsurai	Bitter
Shizenna	Natural	Zurui	Sly

Table 2. *Mean pleasantness ratings for the unfiltered vocal tones and word meanings of Japanese and English utterances used in Study 1*

	Word meaning			
	Pleasant		Unpleasant	
Language	Pleasant vocal tone	Unpleasant vocal tone	Pleasant vocal tone	Unpleasant vocal tone
	Vocal-tone ratings[a]			
Japanese	5.55 (0.34)	2.12 (0.50)	5.77 (0.30)	2.29 (0.50)
English	5.80 (0.28)	2.23 (0.49)	5.51 (0.30)	2.25 (0.41)
	Word-meaning ratings[b]			
Japanese	5.59 (0.26)		2.37 (0.29)	
English	5.67 (0.34)		2.31 (0.27)	

Note. Standard deviations are given in parentheses Word meaning and vocal tone were each rated on a scale from 1 (*very unpleasant*) to 7 (*very pleasant*).

[a]An analysis of variance (ANOVA) performed on these means showed that only the main effect of vocal tone was significant, $F(1, 56) = 1{,}332.9$, $p < .0001$. The same ANOVA performed on the filtered counterparts similarly showed only a significant main effect of vocal tone, $F(1, 56) = 145.6$, $p < .0001$.

[b]An ANOVA performed on these means showed that only the main effect of word meaning was significant, $F(1, 60) = 1{,}962.0$, $p < .0001$.

response times. We first statistically controlled for utterance length. For this purpose, we regressed all the response times on utterance length. For each data point, we obtained a predicted response time, namely, the value predicted as a linear function of the length of the utterance. Deviations from the predicted values (i.e., residuals) were added to the overall mean response time to yield adjusted response times. Table 3 presents the pertinent means.

An analysis of variance (ANOVA) with two between-subjects variables (culture and judgment) and two within-subjects variables (word meaning and vocal tone) was performed on response times. As predicted, the Word Meaning × Vocal Tone interaction proved significant, $F(1, 210) = 43.54$, $p < .0001$. Further, this interaction was qualified by both judgment and culture. The four-way interaction proved significant, $F(1, 210) = 4.62$, $p < .05$. To facilitate further analyses, we computed an interference index by subtracting the mean response time for the congruous utterances from the mean response time for the incongruous ones. Positive scores suggest significant interference by information in the to-be-ignored channel. The mean interference scores are displayed in Figure 1.

In all the four conditions defined by culture and judgment, a reliable interference effect was observed (all ps $< .05$). As predicted, however, the relative size of the effect depended on both judgment and culture. For Japanese, the interference was greater for word-meaning judgments than for vocal-tone judgments. A separate ANOVA performed on the Japanese data showed that the difference was reliable, $t(117) = 1.85$, $p < .05$, one-tailed. This provides evidence that Japanese spontaneously pay more attention to vocal tone than to word meaning. For Americans, however, the interference was significantly stronger in vocal-tone judgments than in word-meaning judgments, $t(210) = 1.78$, $p < .05$, one-tailed. This supports the hypothesis that Americans are attentionally attuned more to verbal content than to vocal tone. To look at the data from a different angle, the interference in vocal-tone judgments was significantly greater for Americans than Japanese, $t(210) = 2.49$, $p < .02$. Although the interference in word-meaning judgments tended to be greater for Japanese than Americans, the difference was statistically trivial, $t < 1$.

Table 3. *Mean response time (in milliseconds) and accuracy in the two judgment conditions of Study 1*

	Word meaning			
	Pleasant		Unpleasant	
Judgment and language	Pleasant vocal tone	Unpleasant vocal tone	Pleasant vocal tone	Unpleasant vocal tone
	Response time			
Word meaning				
Japanese	1,004 (184)	1,094 (222)	1,060 (199)	1,027 (143)
English	1,324 (282)	1,429 (310)	1,395 (272)	1,406 (298)
Vocal tone				
Japanese	916 (315)	949 (319)	976 (255)	944 (224)
English	1,328 (341)	1,427 (428)	1,465 (367)	1,375 (394)
	Accuracy			
Word meaning				
Japanese	.99 (.05)	.97 (.08)	.95 (.08)	.99 (.04)
English	.99 (.04)	.92 (.09)	.95 (.07)	.96 (.07)
Vocal tone				
Japanese	.98 (.06)	.98 (.04)	.94 (.12)	.94 (.08)
English	.95 (.10)	.93 (.10)	.93 (.12)	.95 (.07)

Note. Standard deviations are given in parentheses.

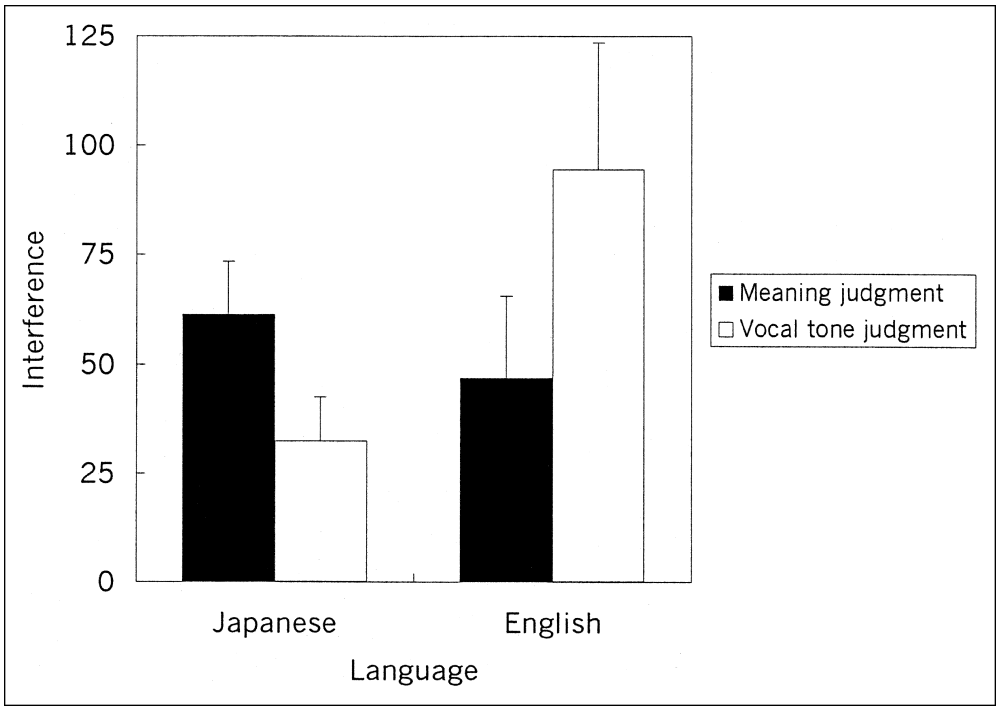

Fig. 1. The Stroop-type interference effect in response time in Study 1. The index of the interference was computed by subtracting mean response time for congruous stimuli from mean response time for incongruous stimuli. Results are shown separately for word-meaning and vocal-tone judgments of Japanese- and English-speaking respondents.

Finally, there was one unexpected finding: Mean response time was longer for Americans than for Japanese (*M*s = 1,394 vs. 994), $F(1, 210) = 117.32$, $p < .0001$. We have no obvious interpretation of this result, especially in view of the fact that no such difference was found in our earlier study (Kitayama & Ishii, 2002). For the present purposes, however, it is important that despite the cultural difference in mean response time, the pattern of interference effects (see Fig. 1) was in full conformity with our predictions.

Accuracy

We submitted data on the percentage of correct responses to an ANOVA. Overall, regardless of culture, there was stronger interference in word-meaning judgments than in vocal-tone judgments. The Word Meaning × Vocal Tone interaction proved significant, $F(1, 210) = 27.19$, $p < .0001$. Judgmental accuracy was considerably lower for incongruous utterances than for congruous utterances. Further, this interference effect was more pronounced in word-meaning judgments than in vocal-tone judgments, $F(1, 210) = 9.68$, $p < .005$. No interaction involving culture was found. In addition, accuracy was slightly lower for Americans than for Japanese (*M*s = .94 vs. .96), $F(1, 210) = 7.64$, $p < .01$. Further, in both judgment conditions, accuracy was lower for utterances with negative word meaning than for those with positive word meaning (*M*s = .95 vs. .96), $F(1, 210) = 7.21$, $p < .01$.

The accuracy measure suggests that Americans found it as hard to ignore vocal tone as Japanese did. This clearly demonstrates that both Japanese and Americans do pay attention to both channels of information. Although we did not find the expected interaction with culture in the accuracy measure, a ceiling effect might have made it difficult to detect it. Indeed, we did find such an interaction in the response time measure. In conjunction with our earlier finding (Kitayama & Ishii, 2002), the current data can be taken to suggest that whereas Americans are attentionally attuned more to verbal content, Japanese are attuned more to vocal tone.

STUDY 2: TAGALOG-ENGLISH BILINGUALS IN THE PHILIPPINES

Study 2 was designed to address two prominent issues that were left unexplored in previous work. First, it was important to examine whether the Japanese pattern could be generalized to other high-context cultures and languages. Second, it was also important to get some insight into the relative role played by culture and language in mediating the attentional biases. To address these issues, in Study 2 we tested Tagalog-English bilinguals in the Philippines.

There is good reason to assume that the Filipino culture is interdependent or collectivist in its central ethos (Church, 1987) and, furthermore, that its indigenous language of Tagalog is high-context in its pragmatic usage. Several Filipino linguists we consulted consensually endorsed this characterization of Tagalog. Hence, we expected that Filipinos should show a high-context pattern of interference, especially when tested in Tagalog.

Furthermore, an examination of Tagalog-English bilinguals in the Philippines provides an ideal setting for testing the relative merits of linguistic relativity and cultural relativity. Tagalog is an indigenous language in the Philippines, spoken by virtually everyone in the coun-

try. Yet, in 1901, during the period of the American occupation, English was adopted by the Department of Education of the Filipino government as the official language of instruction in schools at all levels (Gonzales, 1997). English therefore is currently spoken by a vast majority of the Filipino population, especially in its well-educated segments (see Gonzales, 1996, for an analysis of available census data). Moreover, English has been so heavily inculcated into daily life that many Filipinos regard both Tagalog and English as their native languages (Bautista, 2000).

Given this state of affairs in the contemporary Philippines, the linguistic relativity hypothesis and the cultural relativity hypothesis suggest two contrasting predictions. If attentional biases are guided by certain properties intrinsic to languages used (as suggested by the linguistic relativity hypothesis), Filipinos should show a high-context pattern of interference when tested in Tagalog, but a low-context pattern of interference when tested in English. If, however, the attentional biases are fostered primarily by cultural practices associated with daily communications and conversations (as suggested by the cultural relativity hypothesis), then Filipinos should show a high-context pattern of interference regardless of the language used.

Method

Respondents and procedure

One hundred twenty-two Filipino undergraduates (61 females and 61 males) at a Filipino University, all Tagalog-English bilinguals, participated in the experiment. They were randomly assigned to one of four conditions defined by judgment (word meaning vs. vocal tone) and language (Tagalog vs. English). All instructions were given in the language in which the stimuli were presented. The procedure was identical to the one in Study 1 except for the following two points. First, there were 60 trials, preceded by 10 practice trials (see Materials). Second, we included a condition in which the to-be-ignored channel contained relatively neutral information.

Materials

The same steps detailed in the discussion of Study 1 were followed to select the final set of 180 utterances (10 words × 2 languages × 3 meanings × 3 vocal tones; see Table 4). Unlike in Study 1, we also included utterances neutral in word meaning and neutral in vocal tone in the final set. The same group of four bilingual speakers, both males and females, read the words in different emotional tones. A separate group of bilinguals (total $N = 108$) provided ratings for word meaning and vocal tone. The vocal-tone ratings were obtained for both the original utterances and their filtered counterparts. These ratings were used in the stimulus selection. As can be seen in Table 5, in the final set of utterances, vocal tone was manipulated independently of language and word meaning, and, moreover, vocal tone and word meaning were equally extreme (see the Table 5 footnotes for details). Only those words that had either pleasant or unpleasant meaning were used in the word-meaning judgment condition, resulting in a set of 60 utterances (10 utterances × 2 meanings × 3 vocal tones). Likewise, in the vocal-tone judgment condition, only those utterances that had either pleasant or unpleasant vocal tone were used, resulting in a set of 60 utterances (10 utterances × 3 meanings × 2 vocal tones).

Results and Discussion

Table 6 shows all pertinent means. Although neutral utterances were included in this study, they are not directly relevant to our primary hypothesis. In the main body of analyses, therefore, we examined only utterances that had emotional verbal content and emotional vocal tone (see Table 6 for the means for neutral utterances). As in Study 1, we controlled for effects of utterance length.

Response time

A 2 (language) × 2 (judgment) × 2 (word meaning) × 2 (vocal tone) ANOVA showed a significant interaction between word meaning and vocal tone, $F(1, 118) = 12.35$, $p < .001$. Further, a second-order interaction involving word meaning, vocal tone, and judgment also proved significant, $F(1, 118) = 5.20$, $p < .03$. There was significant interference by vocal tone in word-meaning judgments, $t(118) = 2.05$, $p < .05$, but no interference by word meaning in vocal-tone judgments was observed ($t < 1$). This pattern was found regardless of the language used. The third-order interaction involving word meaning, vocal tone, judgment, and language was negligible, $F < 1$. The size of the interference effect in each condition is displayed in Figure 2.

Table 4. *Words used in Study 2*

Word meaning					
Pleasant		Neutral		Unpleasant	
Tagalog	English	Tagalog	English	Tagalog	English
Bago	New	Amoy	Smell	Asiwa	Clumsy
Hangarin	Purpose	Elektron	Electron	Galit	Anger
Indibidwalidad	Individuality	Impluwensiya	Influence	Kasinungalingan	Lie
Kislap	Sparkle	Inaantok	Sleepy	Lagnat	Fever
Kulay	Color	Konkreto	Concrete	Maingay	Noisy
Natural	Natural	Lugar	Location	Marumi	Dirty
Pagkakataon	Chance	Ordinaryo	Ordinary	Mayabang	Vain
Pagsisikap	Effort	Pagitan	Midway	Nakakatakot	Scary
Posible	Possible	Reyalidad	Reality	Pinsala	Injury
Sigurado	Certain	Uri	Type	Reklamo	Complaint

Table 5. *Mean pleasantness ratings for the unfiltered vocal tones and word meanings of Tagalog and English utterances used in Study 2*

	Word meaning								
	Pleasant			Neutral			Unpleasant		
Language	Pleasant vocal tone	Neutral vocal tone	Unpleasant vocal tone	Pleasant vocal tone	Neutral vocal tone	Unpleasant vocal tone	Pleasant vocal tone	Neutral vocal tone	Unpleasant vocal tone
				Vocal-tone ratings[a]					
Tagalog	5.67 (0.54)	4.13 (0.42)	2.32 (0.59)	5.78 (0.42)	4.00 (0.21)	2.31 (0.51)	5.53 (0.36)	3.93 (0.44)	2.23 (0.59)
English	5.85 (0.34)	4.16 (0.33)	2.32 (0.83)	5.44 (0.44)	4.24 (0.29)	2.10 (0.44)	5.63 (0.34)	4.07 (0.37)	2.32 (0.56)
				Word-meaning ratings[b]					
Tagalog		5.53 (0.22)			4.26 (0.29)			2.66 (0.24)	
English		5.56 (0.22)			4.15 (0.39)			2.33 (0.24)	

Note. Standard deviations are given in parentheses. Word meaning and vocal tone were each rated on a scale from 1 (*very unpleasant*) to 7 (*very pleasant*).
[a]An analysis of variance (ANOVA) performed on these means showed that only the main effect of vocal tone was significant, $F(2, 162) = 776.0$, $p < .0001$. The same ANOVA performed on the filtered counterparts similarly showed only a significant main effect of vocal tone, $F(2, 162) = 143.5$, $p < .0001$.
[b]An ANOVA performed on these means showed that only the main effect of word meaning was significant, $F(2, 174) = 1{,}811.3$, $p < .0001$.

Accuracy

The pattern for accuracy paralleled the pattern for response time. Thus, the interference by vocal tone in word-meaning judgments was much stronger than the interference by word meaning in vocal-tone judgments. This pattern is underscored by both a significant interaction between word meaning and vocal tone and an interaction involving word meaning, vocal tone, and judgment, $F(1, 118) = 26.00$, $p < .0001$, and $F(1, 118) = 6.92$, $p < .01$, respectively. Finally, the interference effects in both judgments were stronger in English than in Tagalog, $F(1, 118) = 6.11$, $p < .02$. The pattern was found regardless of language, however.

GENERAL DISCUSSION

Drawing on our earlier studies (Kitayama & Ishii, 2002), the current work supports the hypothesis that people in different cultures are differentially attuned to verbal content vis-à-vis vocal tone in comprehending emotionally spoken emotional words. Specifically, Americans were attentionally biased toward verbal content, whereas both Japanese and Filipinos were attentionally biased toward vocal tone. Moreover, the divergent pattern of attentional bias appears largely cultural rather than linguistic.

The hypothesis that Americans are especially attuned to verbal content has numerous implications. The vast majority of studies in

Table 6. *Mean response time (in milliseconds) and accuracy in the two judgment conditions of Study 2*

	Word meaning								
	Pleasant			Neutral			Unpleasant		
Judgment and language	Pleasant vocal tone	Neutral vocal tone	Unpleasant vocal tone	Pleasant vocal tone	Neutral vocal tone	Unpleasant vocal tone	Pleasant vocal tone	Neutral vocal tone	Unpleasant vocal tone
				Response time					
Word meaning									
Tagalog	1,406 (327)	1,369 (402)	1,574 (451)	—	—	—	1,538 (538)	1,508 (518)	1,464 (375)
English	1,365 (416)	1,299 (287)	1,505 (565)	—	—	—	1,457 (534)	1,400 (361)	1,300 (234)
Vocal tone									
Tagalog	1,462 (361)	—	1,594 (479)	1,572 (473)	—	1,675 (633)	1,590 (455)	—	1,669 (490)
English	1,385 (377)	—	1,396 (406)	1,432 (486)	—	1,356 (339)	1,457 (431)	—	1,406 (357)
				Accuracy					
Word meaning									
Tagalog	.97 (.08)	.98 (.07)	.86 (.27)	—	—	—	.84 (.28)	.82 (.28)	.89 (.23)
English	.93 (.12)	.91 (.16)	.72 (.29)	—	—	—	.76 (.30)	.75 (.32)	.92 (.17)
Vocal tone									
Tagalog	.82 (.24)	—	.75 (.25)	.79 (.25)	—	.70 (.26)	.72 (.31)	—	.67 (.30)
English	.89 (.17)	—	.77 (.20)	.73 (.25)	—	.86 (.22)	.80 (.23)	—	.83 (.13)

Note. Standard deviations are given in parentheses.

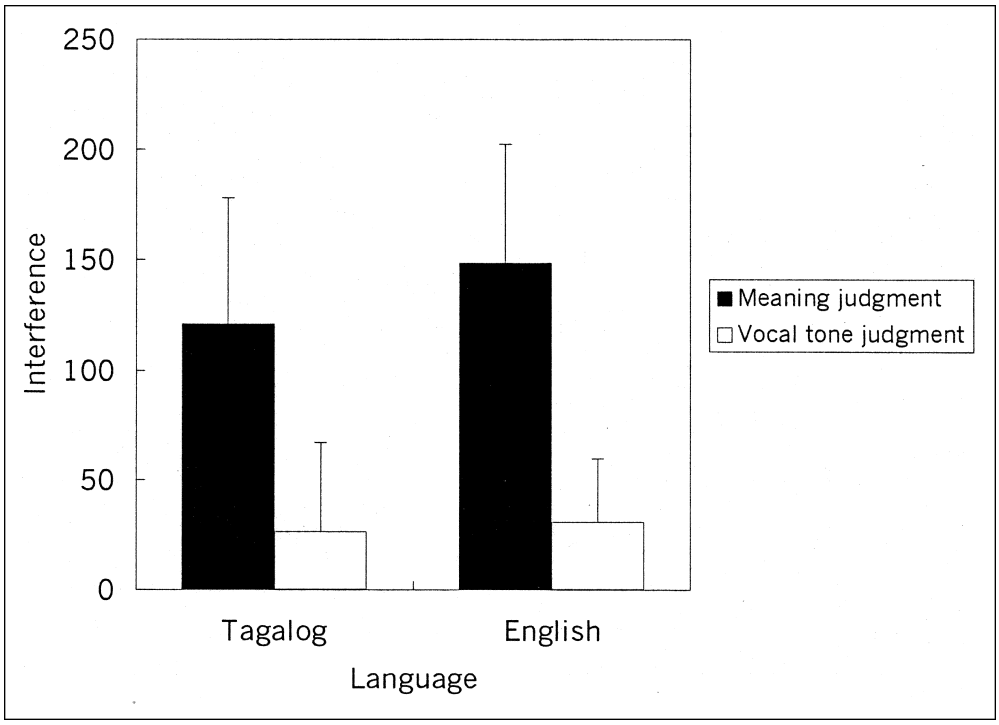

Fig. 2. The Stroop-type interference effect in response time in Study 2. The index of the interference was computed by subtracting mean response time for congruous stimuli from mean response time for incongruous stimuli. Results are shown separately for word-meaning and vocal-tone judgments of Tagalog and English words.

many areas of social cognition—including person perception, priming, and attribution—have used verbal materials. On the basis of the current findings, we suspect that some of the phenomena identified in this literature may depend on the attentional bias that favors verbal content; if so, they may be more difficult to obtain in cultures outside of North America, especially in high-context cultures such as those of Japan and the Philippines. For example, Americans often show a persistent bias to infer speech intent from what is being said, while failing to appreciate the impact of existing social constraint on the speaker (Gilbert & Malone, 1995). This bias, called the correspondence bias, may be less pronounced in high-context cultures (Choi & Nisbett, 1998; Miyamoto & Kitayama, 2002).

The present work has provided a significant insight into the debate on linguistic relativity (see Lucy, 1992, for a review) by examining Tagalog-English bilinguals in the Philippines. Our data indicate that the effect of language is minimal as long as different languages are integrated into a single system of cultural practices. This conclusion is quite consistent with a recent study by Ji, Nisbett, and Zhang (2001). These researchers employed a categorization task and showed that whereas Americans tend to be analytic (i.e., to use taxonomical rules in categorization—e.g., *Mother* and *Father* are both adults, but *Child* is not), Hong Kong Chinese tend to be holistic (i.e., to use event schemas for the same purpose—e.g., *Mother*, but not *Father*, takes care of a *Child*). The Chinese manifested the holistic tendency regardless of whether they were tested in Chinese or English.

Having argued for the primacy of culture over language, however, we should hasten to add that it is often through language socialization that cultural practices and meanings are inculcated into new members of a cultural group (Heath, 1990; Lucy, 1992; Ochs, 1996). Hence, in all likelihood, language is deeply implicated, and perhaps even indispensable, in producing cultural differences in mental processes. Yet the current evidence suggests that the language's hold on mental processes is possible only in conjunction with the associated cultural practices of communication and social interaction.

Acknowledgments—We thank Mayumi Karasawa, Takahiko Masuda, and members of the Kyoto University cultural psychology lab for their help in data collection.

REFERENCES

Ambady, N., Koo, J., Lee, F., & Rosenthal, R. (1996). More than words: Linguistic and nonlinguistic politeness in two cultures. *Journal of Personality and Social Psychology, 70*, 996–1011.

Barnlund, D.C. (1989). *Communicative styles of Japanese and Americans: Images and realities*. Belmont, CA: Wadsworth.

Bautista, M.L.S. (2000). *Defining standard Philippine English: Its status and grammatical features*. Manila, the Philippines: De La Salle University Press.

Brown, R. (1976). Reference: In memorial tribute to Eric Lenneberg. *Cognition, 4*, 125–153.

Choi, I., & Nisbett, R.E. (1998). Situational salience and cultural differences in the correspondence bias and actor-observer bias. *Personality and Social Psychology Bulletin, 24*, 949–960.

Church, A.T. (1987). Personality research in a non-Western culture: The Philippines. *Psychological Bulletin, 102*, 272–292.

Gilbert, D.T., & Malone, P.S. (1995). The correspondence bias. *Psychological Bulletin, 117*, 21–38.

Gonzales, A. (1996). The Philippine experience with the English language: The limits of science in language teaching. In E.S. Castillo (Ed.), *Alay sa Wika: Festschrift in honor of Fe T. Otanes on her 67th birthday* (pp. 139–151). Manila, the Philippines: Linguistic Society of the Philippines.

Gonzales, A. (1997). The history of English in the Philippines. In M.L. Bautista (Ed.), *English is an Asian language: The Philippine context* (pp. 25–40). Manila, the Philippines: De La Salle University Press.

Hall, E.T. (1976). *Beyond culture*. New York: Doubleday.

Heath, S.B. (1990). The children of Trackton's children: Spoken and written language in social change. In J.W. Stigler., R.A. Shweder., & G. Herdt (Eds.), *Cultural psychology: Essays on comparative human development* (pp. 496–519). New York: Cambridge University Press.

Ji, L.J., Nisbett, R.E., & Zhang, Z. (2001). *Culture, language and categorization.* Unpublished manuscript, University of Michigan, Ann Arbor.

Kitayama, S. (1996). Remembrance of emotional speech: Improvement and impairment of incidental verbal memory by emotional voice. *Journal of Experimental Social Psychology*, *32*, 289–308.

Kitayama, S. (2000). Cultural variations in cognition: Implications for aging research. In P.C. Stern & L.L. Carstensen (Eds.), *The aging mind: Opportunities in cognitive research* (pp. 218–237). Washington, DC: National Academy Press.

Kitayama, S. (2002). Culture and basic psychological processes: Toward a system view of culture. *Psychological Bulletin*, *128*, 89–96.

Kitayama, S., Duffy, S., Kawamura, T., & Larsen, J.T. (in press). Perceiving an object and its context in different cultures: A cultural look at New Look. *Psychological Science*.

Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional speech in two cultures. *Cognition and Emotion*, *16*, 29–59.

Lucy, J.A. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge, England: Cambridge University Press.

Markus, H.R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*, 224–253.

Masuda, T., & Nisbett, R.E. (2001). Attending holistically vs. analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology*, *81*, 922–934.

Miyamoto, Y., & Kitayama, S. (2002). Cultural variation in correspondence bias: The critical role of attitude diagnosticity of socially constrained behavior. *Journal of Personality and Social Psychology*, *83*, 1239–1248.

Nisbett, R.E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic vs. analytic cognition. *Psychological Review*, *108*, 291–310.

Ochs, E. (1996). Linguistic resources for socializing humanity. In J.J. Gumperz & S.C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 407–437). New York: Cambridge University Press.

Scherer, K.R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*, 143–165.

Whorf, B.L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. New York: Wiley.

(RECEIVED 9/24/01; REVISION ACCEPTED 1/27/02)