



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Journal of Sound and Vibration 277 (2004) 589–601

JOURNAL OF  
SOUND AND  
VIBRATION

[www.elsevier.com/locate/yjsvi](http://www.elsevier.com/locate/yjsvi)

# Evaluating the equivalence of verbal scales and question stems to be used in English and Japanese noise annoyance questions

Kirk Masden<sup>a,\*</sup>, Takashi Yano<sup>b</sup>

<sup>a</sup>*Faculty of Economics, Kumamoto Gakuen University, Oe 2-5-1, Kumamoto 862-8680, Japan*

<sup>b</sup>*Faculty of Engineering, Kumamoto University, Kurokami 2-39-1, Kumamoto 860-8555, Japan*

Accepted 25 March 2004  
Available online 31 July 2004

---

## Abstract

Two experiments were conducted to evaluate the equivalence of noise annoyance scales and questions recommended by Team 6 (Community Response to Noise) of IC BEN for English and Japanese, two strikingly different languages. The first experiment was conducted with bilingual subjects in order to test a key assumption of the method established by IC BEN Team 6 for the development of equivalent noise annoyance scales: subjects who speak different languages interpret the concept of “highest degree” of annoyance similarly. The results indicate that English- and Japanese-speaking subjects do interpret the “highest degree” similarly. The second experiment tested for effects of wording differences. English- and Japanese-speaking subjects were presented with noise annoyance questions of one of three format types. The first type was similar to the question format recommended by IC BEN Team 6. It focused on the degree to which a given noise would “bother, disturb, or annoy” the subject. The second asked subjects to evaluate the “bothersome, annoying, or disturbing” quality of the noise. The third asked how much the noise would “worry, irritate, or concern” the subject. No significant difference was found in responses to the three formats when subjects evaluated noise in laboratory conditions.

© 2004 Elsevier Ltd. All rights reserved.

---

\*Corresponding author. Tel.: +81-96-364-7964; fax: +81-96-372-7207.  
E-mail address: [masden@kumagaku.ac.jp](mailto:masden@kumagaku.ac.jp) (K. Masden).

## 1. Introduction

In 1993, the Community Response to Noise Team (Team 6) of the International Commission on the Biological Effects of Noise (ICBEN) initiated a project to develop standards for the construction of noise annoyance scales and questions in multiple languages. The project resulted in the development of a procedure for constructing comparable scales and a standard English-language question that may serve as a model for the preparation of questions in other languages [1]. If widely adopted, these standard scales and questions have the potential to facilitate a great leap forward in the advancement of international and intercultural research on community responses to noise. In order for the scales and questions to be widely adopted, however, researchers must first be convinced that they are indeed equivalent.

This paper reports on two experiments that were conducted to evaluate the equivalence of the English and Japanese scales and question stems recommended by ICBEN's Team 6. As English and Japanese are very different languages, it is particularly important to confirm the equivalence of the scales and questions endorsed by Team 6 for these two languages. The first experiment used 73 bilingual subjects to test a key premise of the ICBEN scale construction procedure: the upper extreme of the range of possible annoyance imagined by subjects does not differ widely between cultures and languages. The second experiment was conducted to evaluate the equivalence of English and Japanese question stems. In particular, the experiment was designed to test the equivalence of question stems employing base descriptors that point to the effect of the noise on the subject (e.g. the English term "annoyance") and wording that focuses attention on the quality of the noise (e.g. the Japanese term "urusasa"). Since a direct test of the difference between "annoyance" and "urusasa" is impossible, an indirect test was conducted. It was hypothesized that if the differences in wording and vocabulary used in Japanese and English question stems were to have a significant effect on subject responses to noise annoyance questions, similar effects should be observable within the languages when the wording and vocabulary are altered to approximate those differences. The experiment employed two separate groups of subjects, English- and Japanese-speaking subjects (not bilingual), to assess the effects of three question formats on annoyance responses.

## 2. Experiments

### 2.1. *Experiment I: the equivalence of English and Japanese verbal scales*

#### 2.1.1. *Outline of Experiment I*

In this experiment, bilingual subjects were asked to follow a procedure that was essentially the same as that devised by ICBEN Team 6 for the production of equivalent annoyance scales. The procedure employed in this experiment differed in that all subjects chose modifiers for use in both English and Japanese scales and evaluated the intensities of both English and Japanese modifiers. In addition, a paired comparison test of 12 English and Japanese modifiers was appended, though paired comparison was not part of the original ICBEN procedure.

2.1.2. Questionnaires

Each questionnaire contained the following tasks:

1. Construction of 5- and 4-point scales in English: Subjects constructed 5- and 4-point equidistant annoyance scales in English from the minimum to the maximum by selecting suitable modifiers from a pool of 21 English modifiers (Table 1).
2. Construction of 5- and 4-point scales in Japanese: Subjects constructed 5- and 4-point equidistant annoyance scales in Japanese from the minimum to the maximum by selecting suitable modifiers from a pool of 21 Japanese modifiers (Table 2).
3. Line-marking exercise for 42 modifiers in English and Japanese: Subjects evaluated the intensity of the 42 English and Japanese modifiers by placing a mark on a 10 cm line as shown in Fig. 1. The modifiers were presented sequentially in a random order.
4. Paired comparison test: Six English and six Japanese modifiers of high intensity were selected for evaluation in a paired comparison test on the basis of the results of the ICBEN study [1]. Details regarding the method of comparison have been published in another paper [2].

There were two types of questionnaires: “Annoyed” was used as the base descriptor throughout in one while “urusai” (the Japanese equivalent of “annoyed”) was used in the other. Both types were bilingual. In the questionnaires in which “annoyed” was used as the base descriptor, English text appeared in a column on the left side of each page and the corresponding Japanese appeared in a column on the right. This arrangement was reversed in the questionnaires in which “urusai” was the base descriptor.

2.1.3. Subjects

Seventy-three subjects between the ages of 20 and 71 who were fluent in Japanese and English participated in the study. The subjects were divided into the following four groups: (1) females,

Table 1  
Twenty-one English modifiers

---

Extremely, tremendously, severely, strongly, highly, very, significantly, substantially, considerably, importantly, rather, moderately, fairly, somewhat, partially, slightly, a little, hardly, barely, insignificantly, not at all

---

Table 2  
Twenty-one Japanese modifiers

---

Hijoni, kiwamete, hidoku, sugoku, taihen, soto, totemo, kanari, daibu, warini, hikakuteki, tasho, yaya, ikuraka, sukoshi, wazukani, sorehodo. . .nai, taishite. . .nai, amari. . .nai, hotondo. . .nai, mattaku. . .nai

---

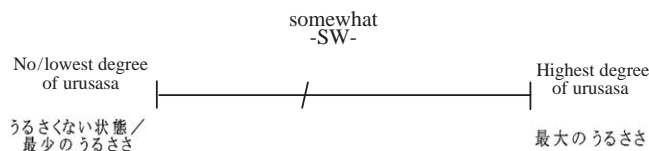


Fig. 1. Line-marking exercise.

Japanese is first language (20 subjects; mean age, 40); (2) males, Japanese is first language (17 subjects; mean age, 45); (3) females, English is first language (17 subjects; mean age, 37); and (4) males, English is first language (19 subjects; mean age, 38). The nationalities of the subjects who spoke English as their first language were as follows: USA, 21; Australia, 4; United Kingdom, 3; Ireland, 2; France, 1; New Zealand, 1; Japan, 1. The Japanese subjects had lived in English-speaking areas for an average of 5 years while those who spoke English as their first language had lived in Japan for an average of 11 years. Most of the subjects had experience working as professional translators, interpreters, or language teachers.

#### *2.1.4. Procedure*

The questionnaires were distributed evenly to each of the above four groups of the subjects. Most of the subjects who agreed to participate in the study received and returned the questionnaires via surface mail but questionnaires were hand-delivered to some subjects and filled out by them while a researcher waited. It took about an hour to complete each questionnaire. The data was collected between June of 2001 and October of 2001.

### *2.2. Experiment II: the equivalence of English and Japanese question stems*

#### *2.2.1. Outline of Experiment II*

English- and Japanese-speaking subjects were asked to do the following three tasks: (1) to evaluate sounds in a laboratory experiment; (2) to respond to hypothetical questions about noise annoyance; and (3) to respond to hypothetical questions about noise annoyance and annoyance not caused by noise. Three different types of question stems were used in the first two tasks to test for the effects of wording differences. The third task was designed to test for cultural differences in sensitivity to noise among the English- and Japanese-speaking subjects.

#### *2.2.2. Questionnaires*

The questionnaire for the study was divided into three independent parts:

Part I: Reactions to the 16 recorded sounds played during the test session.

Part II: Reactions to 10 hypothetical community noise situations: (1) hearing a distant aircraft about once a week; (2) hearing big trucks (when you are in your home) every time the traffic signal changes at a nearby intersection; (3) hearing dogs barking at night; (4) a door squeaking; (5) neighbors' TV; (6) background music from a business; (7) backup warning signals on trucks; (8) being awoken by motorcycles; (9) aircraft interfering with TV; and (10) speech interference from traffic.

Part III: Reactions to 3 additional types of hypothetical problems: (1) Environmental, transportation noise problems at home: trucks interfering with hearing television; awakening by aircraft; hearing a distant expressway; and aircraft interfering with conversation. (2) Household noise problems: hearing neighbors shouting at their children; hearing the refrigerator; and plumbing noise. (3) Non-noise problems: industrial odors; air pollution; mosquitoes in the home; a burnt-out streetlight; neighbor's trash; forgetting people's names; eyesight problems; automobile not starting; cockroaches; a neighbor's light; street dangerous for children; factory dirt; hearing loss; sticky doors; and a junk car business.

In Parts I and II, each subject was asked only one of the following formats of the question stem:

- Format A:* How much would this (noise) bother, disturb, or annoy you?  
(urusaku, matawa fukai ni kanjiru)
- Format B:* How bothersome, annoying or disturbing should this (noise) be rated as?  
(urusai, matawa kininaru)
- Format C:* How much would this (noise) worry, irritate, or concern you?  
(nayamasareru)

Format A is closest to the standard IC BEN question: “Thinking about the last (12 months or so), when you are here at home, how much does noise from (noise source) bother, disturb, or annoy you?” The wording of Format B is intended to simulate typical Japanese questions about noise annoyance which ask subjects to rate the degree to which a noise source is “urusai” (“annoying”) rather than the extent to which they personally are “annoyed.” Format C is similar to Format A but uses base descriptors that indicate deeper psychological disturbance. Of the three, it was hypothesized that Format B would be most likely to elicit a strong response because its focus on the quality of the noise source does not require the subject to admit to any personal loss of psychological equilibrium. Similarly, it was hypothesized that Format A would be somewhat less likely to elicit a strong response because subjects must admit that they would be annoyed or disturbed by the noise source. By the same logic, it was hypothesized that Format C would elicit the weakest response because it requires subjects to admit to more profound levels of personal disturbance. The Japanese versions of the three formats are not exact translations of the English; rather, care was taken to approximate the different nuances of the three English formats. The questions were constructed such that the order of the anticipated strengths of response (Format B, strongest; Format C, weakest) under the authors’ hypothesis was the same as the English questions.

In Parts I and II, subjects were asked to use one of two types of scales when responding to each question: a 5-point verbal scale or an 11-point numeric scale. The labels used on the 5-point verbal scales were “extremely,” “very,” “moderately,” “slightly” and “not at all” in English and “hijoni,” “kanari,” “tasho,” “amari...nai,” and “mattaku...nai” in Japanese. The 11-point scale extended from 0 (labeled “not at all” or “mattaku...nai”) to 10 (labeled “extremely” or “hijoni”). In Part III, all subjects received the following invariant format and answer scale:

If you had this problem, how annoying or unpleasant would this problem be for you?

	0	1	2	3	4	5	6	7	8	9	10
Not at all											Extremely

The following two versions of the questionnaire were prepared for each of the three question stem formats: (1) a version in which the verbal scale of Parts I and II appeared first and the numeric scale followed in each of the two Parts; and (2) a version with the opposite order of verbal and numeric scales in each Part. In Part III, the order of presentation was reversed for those who received the numeric scales first in Parts I and II. Thus, a total of six versions (two ordering schemes for each of three question stem formats) were prepared in both English and Japanese.

### 2.2.3. Subjects

The Japanese subjects consisted of 157 male and 41 female students tested at the University of Kumamoto. The English-speaking subjects consisted of 13 male and 23 female students tested at the University of Sydney, 47 male and 16 female students tested at the University of Melbourne, Australia, and 6 male and 24 female employees tested at the NASA Langley Research Center in Hampton, Virginia, USA. The age ranges and the mean ages at the four sites were as follows: Kumamoto, 19–30, mean 21.0; Sydney, 19–36, mean 21.1; Melbourne, 18–27, mean 19.6; NASA, 26–62, mean 45.6.

### 2.2.4. Procedure

The 16 sounds rated in Part I of the questionnaire were 30 s recordings of road traffic noise exposures from a single location near an expressway. They were prepared on a CD for playback at about 56, 64, 72 and 80 dB ( $L_{Aeq}$ ) after being calibrated using a pink noise test sound. At the University of Kumamoto (Japan), the University of Sydney (Australia), and the University of Melbourne (Australia) subjects were tested in ordinary classrooms. At the NASA Langley Research Center in Hampton, Virginia (USA), the experiment was conducted in an acoustically treated, psychoacoustic test room. Each subject completed one of the six versions of the questionnaire. After the first four sounds were presented so that the subjects could practice the marking procedure, the 16 rated sounds were presented at the four noise levels in a Latin squares design. After completing Part I, the subjects then completed Parts II and III in silence in the same venue. The tests were conducted from October of 1999 to March of 2000.

## 3. Results

### 3.1. Experiment I

#### 3.1.1. Scale construction

In accordance with the method devised by ICBEN Team 6 [1], the following criteria were used to determine the scale-point labels:

1. Intensity difference score (I-C Delta): the difference between the modifier's mean and the scale point's ideal intensity score on a scale of 0–100 (viz., 0, 25, 50, 75 or 100).
2. Net preference score (P%): the net number of selections of the modifier for a particular scale point (the number of selections for the scale point minus the number of selections for other scale points) divided by the total number of subjects.
3. Standard deviation of intensity scores (StD): the standard deviation of the intensity scores for each modifier.

Table 3 shows the 5-point scales constructed using the data produced by all subjects of the present study (bilingual) and the 5-point scales produced by the ICBEN study [1]. The English scale is the same as ICBEN's English scale except that "a little" was selected as the second lowest category. However, the Japanese scale is completely different from that of the ICBEN study as "mattaku...nai" was fixed as the lowest category.

Table 3  
Modifiers for 5-point scales in English and Japanese

English, bilingual:	“extremely,” “very,” “moderately,” “a little” and “not at all”
English, ICBEN:	“extremely,” “very,” “moderately,” “slightly” and “not at all”
Japanese, bilingual:	“kiwamete,” “totemo,” “hikakuteki,” “sukoshi” and “mattaku...nai”
Japanese, ICBEN:	“hijoni,” “daibu,” “tasho,” “sorehodo...nai” and “mattaku...nai”

This result stems in part from differences between Japanese and English. In English, one of the 21 modifiers is clearly dominant in each of the five intensity ranges whereas the group of 21 Japanese modifiers contains two or more modifiers of similar suitability in each intensity range [3]. Moreover, in Japanese, impressions about various modifiers are more affected by differences between subject groups than is the case in English. For example, when regression analysis was applied to the data from the ICBEN study (the intensity score was a dependent variable and the age of the subjects was an independent one) the age effect on the intensity was more dominant in Japanese than English [4]. The regression coefficients were significant at the 1% level for eight of 21 modifiers and at the 5% level for three modifiers in Japanese, whereas they were significant at 1% for three modifiers and at 5% for three modifiers in English.

### 3.1.2. Comparison of intensity scores

Table 4 shows the mean intensity scores for the 21 English and Japanese modifiers on a scale of 100 for this bilingual study and the ICBEN study. The difference between the intensity score of the highest English modifier (extremely, 96.9) and that of the highest Japanese modifier (kiwamete, 93.3) was 3.6 in this study while the difference between the same modifiers was 1.1 (extremely, 94.9; hijoni, 93.8) in the ICBEN study. Fig. 2 compares the average intensity scores produced by subjects for whom English is the first language (L1 = English) and those who speak Japanese as their first language (L1 = Japanese) when evaluating English modifiers. Similarly, Fig. 3 compares the average intensity scores produced by the two groups of subjects when evaluating Japanese modifiers. Significant differences in intensity can be observed in certain individual modifiers (e.g. “fairly” in English and “soto” in Japanese) but a general pattern of difference is not apparent. Cluster analysis was applied to the intensity scores of all 42 modifiers and they were classified into five clusters. Table 5 shows the average intensity scores for the six modifiers of the highest cluster by the first language of the bilingual subjects and in comparison with the ICBEN scores. Although the native speakers of English and Japanese differed by as much as nearly 5 points in their interpretations of individual modifiers, the average difference in their intensity scores in this cluster is less than 1 point.

### 3.1.3. Paired comparison test

The 12 modifiers were placed on a distance scale according to the method of successive categories. The order produced by the paired comparison test was “extremely,” “tremendously,” “hijoni,” “severely,” “sugoku” and “taihen” for the higher intensity modifiers. That for the lower intensity modifiers was “strongly,” “highly,” “very,” “soto,” “kanari” and “daibu.” These orders were consistent with the intensity scores shown in Table 4 except that the positions of “soto” and “kanari” were reversed in the paired comparison test.

Table 4  
Intensity scores of 42 modifiers

English	Bilingual	ICBEN	Japanese	Bilingual	ICBEN
Extremely	96.9	94.9	kiwamete	93.3	91.8
Tremendously	95.6	92.3	hijoni	92.2	93.8
Severely	91.8	90.7	hidoku	90.6	91.0
Strongly	80.3	79.7	sugoku	86.5	89.5
Highly	80.1	78.7	taihen	84.1	86.3
Very	78.4	75.6	totemo	79.9	83.9
Significantly	73.9	67.2	kanari	73.6	83.9
Considerably	71.3	62.2	soto	72.2	84.9
Importantly	71.3	65.1	daibu	71.2	75.2
Substantially	70.7	64.5	hikakuteki	50.9	55.9
Rather	56.0	47.9	warini	49.2	57.4
Fairly	55.2	40.5	ikuraka	36.4	39.2
Moderately	48.1	43.7	tasho	35.6	44.5
Somewhat	35.3	35.7	yaya	34.2	43.5
Partially	31.9	29.6	sukoshi	20.3	34.8
A little	17.2	13.2	sorehodo . . . nai	17.6	21.0
Slightly	16.3	15.4	wazukani	15.0	26.0
Insignificantly	12.7	7.6	taishite . . . nai	14.5	19.6
Hardly	9.0	10.3	amari . . . nai	10.8	18.6
Barely	7.5	8.1	hotondo . . . nai	6.0	6.9
Not at all	0.6	0.8	mattaku . . . nai	0.8	1.0

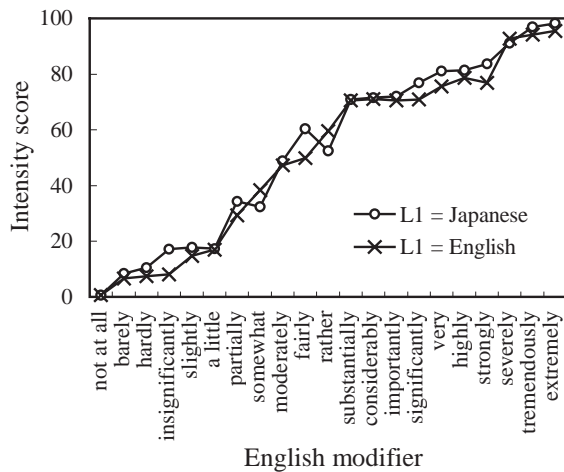


Fig. 2. Comparison of English intensity scores by first language.



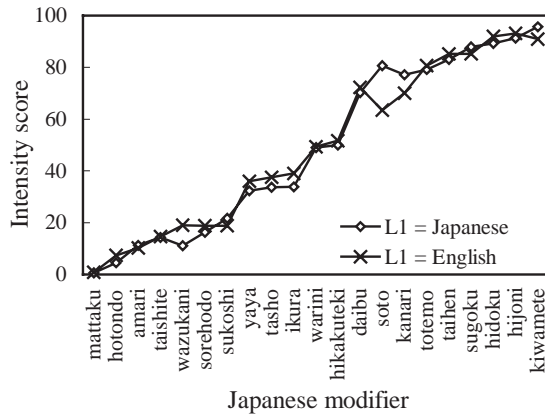


Fig. 3. Comparison of Japanese intensity scores by first language.

Table 5  
Average intensity scores for six modifiers of highest cluster

Subjects	Hidoku	Kiwamete	Hijoni	Severely	Tremendously	Extremely	Japanese average	English average	Average
English L1	92.0	90.8	93.1	92.7	94.2	95.5	92.0	94.1	93.1
Japanese L1	89.2	95.7	91.3	91.1	96.9	98.2	92.1	95.4	93.7
English ICBEN				90.7	92.3	94.9		92.6	
Japanese ICBEN	91.0	91.8	93.8				92.2		

### 3.2. Experiment II

#### 3.2.1. Comparison of annoyance responses to laboratory noise exposure as measured by different question stems

Multiple regression analysis of the Japanese and English data from Part I of Experiment II did not reveal a statistically significant effect of question format on subject responses. In the analysis of the Japanese data, the decibel level of the stimulus, the format (A, B, or C) of the question, and the type of scale used (verbal or numeric) constituted the independent variables while the response score constituted the dependent variable. Responses on the 5-point verbal scale were scored 0, 2.5, 5, 7.5, and 10 to facilitate comparison of the data from the verbal and numeric scales. The English-language data was analyzed in the same manner as the Japanese data with the addition of the test site (Sydney, Melbourne or NASA) as a fourth independent variable. Fig. 4 shows the relationships between noise level and annoyance response for the English-speaking subjects. The abscissa is the LAeq of the stimuli and the ordinate is the average response (for English- and Japanese-speaking subjects respectively) for each of the three question formats. The figure shows that there is not a systematic tendency for any one format to elicit more negative responses. The analogous results for the Japanese subjects were almost the same except that the differences between the three formats were even smaller. Although the multiple regression analysis of the data

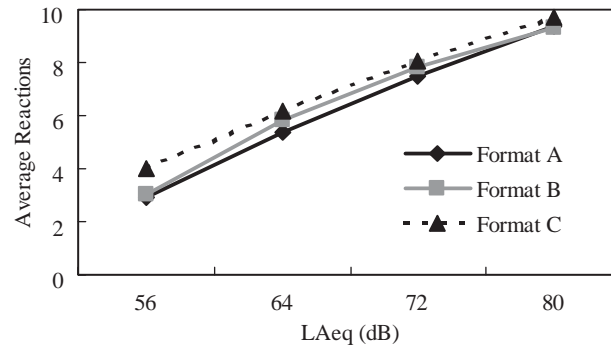


Fig. 4. Comparison of English language subjects' annoyance responses to laboratory noise exposure by question type.

from the three English-language sites did not reveal a significant effect of the different formats, "test site" was found to be statistically significant at the 1% level. This may be due to the lower levels of background noise at the NASA facility.

### 3.2.2. Comparison of annoyance responses to hypothetical noise situations as measured by different question stems

In Part II, subjects were presented with ten different noise environments and asked to imagine what it would be like to live in a home with each noise. Of the 10 noise environments "Hearing a distant aircraft about once a week" elicited the lowest rating while "Hearing big trucks (when you are in your home) every time the traffic signal changes at a nearby intersection" elicited the highest. An analysis of variance in which response was the dependent variable and format, question, site, and scale type were independent variables showed the effect of question format on subject responses to be significant at the 1% level in the Japanese-language data and at the 5% level in the English-language data. The mean responses for all Japanese questions in Part II by format type were as follows: Format A, 6.1; Format B, 5.8; Format C, 5.4. The corresponding means for the English-language data were as follows: Format A, 5.7; Format B, 6.1; Format C, 6.0. These values are not consistent with the hypothesis that Format B should elicit the highest response and Format C should elicit the lowest.

### 3.2.3. Comparison of annoyance responses to hypothetical noise and non-noise situations

An analysis of variance in which response was the dependent variable and the language of the respondent (English or Japanese) and the type of problem suggested (environmental noise, household noise, or non-noise) were independent variables indicated an effect of language on the response to the three types of problems that was significant at the 1% level. When environmental noise and household noise were combined into one category, the same analysis showed the effect of language on response to the noise vs. non-noise problems to be significant at the 5% level. The mean responses for English speakers were 6.8 (noise) and 6.9 (non-noise) whereas the mean responses for Japanese speakers were 6.7 and 7.2, respectively. Though Part III produced results that were determined to be statistically significant, they should not be interpreted as evidence that Japanese speakers are less sensitive to noise than English speakers because the differences between the mean responses of the two are very small.

## 4. Discussion

### 4.1. *Equivalence of highest degree of annoyance in English and Japanese*

Under the IC BEN procedure, before subjects begin to evaluate the intensity of individual modifiers in the line-marking exercise, they are instructed that the “highest degree” point on the line-marking exercise is the “highest degree of annoyance imaginable.” This imaginary “highest degree” then becomes the standard against which the intensity of each modifier is measured. The cross-cultural comparability of the resulting intensity scores is predicated on the hypothesis that subjects of differing linguistic and cultural backgrounds interpret this “highest degree” level similarly. Testing this hypothesis is difficult, however, because there is no obvious standard against which subjects can be asked to measure their interpretations directly. The use of bilingual subjects in this study, however, allows us look for indirect indications of different interpretations.

In this study, all subjects evaluated all of the English and Japanese modifiers using the line-marking exercise. On each questionnaire, the base descriptor and the bilingual format of the exercise were consistent throughout; that is, the format of the line-marking exercise was the same for both English and Japanese modifiers. If there were a significant difference in the “highest degree” imagined by those bilingual subjects who spoke English as their first language and those for whom Japanese was the first language, that difference should lead to a significant numerical difference in intensity scores between the two groups. Moreover, the difference should be most apparent in the modifiers of high intensity because they are closest to the “highest degree” standard.

Accordingly, the average intensity scores of the modifiers found in the highest cluster were calculated as shown in Table 5. As pointed out earlier, the average difference in intensity between subjects who speak English as the first language (English L1) and those who speak Japanese as their first language (Japanese L1) was less than 1 point. This indicates that the two groups did not interpret the “highest degree” standard in significantly different ways.

For half of the subjects, the base descriptor on the line-marking exercise was the English “annoyance” while the remaining subjects received questionnaires in which the Japanese “urusasa” was used. It is also conceivable that these English and Japanese base descriptors might elicit different responses from the subjects based on differing cultural and linguistic norms, but such a difference was found for only one modifier in a two-factor analysis of variance test. Thus, the analysis of variance test did not produce strong evidence of a cultural difference that might affect the interpretation of “highest degree” on the line-marking exercise.

Finally, the agreement between the intensity scores and the order determined by the paired comparison test is further evidence that a difference in the interpretations of “highest degree” did not corrupt the intensity data.

### 4.2. *Equivalence of annoyance responses to different question stems*

The results of Part I of Experiment II indicate that differences in the degree to which these question stems focus on the character of the noise environment or, conversely, the subjective experience of the respondent does not have a significant effect on responses in laboratory situations. Moreover, Sato et al. [5] conducted a Japanese social survey in which no systematic

effect was found for four variations in question type, three of which were analogous to questions employed in this experiment. In Part II of Experiment II, significant effects were found in both the English- and Japanese-language data but the effects did not confirm the author's hypothesis regarding the relative strength of the responses that the three formats should elicit. While the authors hypothesized that Format B should elicit the strongest response and C the weakest, the response to C was stronger than the response to A in the English data; in the Japanese data, the response to A was stronger than that to B. Though the exact reasons for the observed responses are unclear, the hypothetical nature of Part II may have made subjects more likely to focus on the phrasing of the questions.

In sum, neither the psychoacoustic experiment conducted in this study nor a separate social survey study indicated that shifting the focus of question wording between the quality of the noise and the impact of the noise of the psychological state of the subject had a significant effect on subject response. Question wording was found to be significant in responses to hypothetical questions but not in accord with the researchers' hypothesis. On the whole, therefore, we may conclude that the types of question wording examined here do not produce systematic differences in subject responses.

## **5. Conclusions**

This paper reports the results of two experiments. In the first, bilingual subjects constructed annoyance scales in English and Japanese according to the ICBEN protocol. The results of this experiment clearly indicate that English- and Japanese-speaking subjects do not differ significantly in their interpretations of the "highest degree" of annoyance. Thus, a key premise of the equivalence of the ICBEN scales was confirmed for English and Japanese. In the second experiment, English- and Japanese-speaking subjects were asked to evaluate noise presented in a laboratory situation and hypothetical noise problems through a question worded in one of three ways. No significant effect of the differences in the wording was found in the laboratory situation. Significant effects were observed when subjects were asked about hypothetical noise problems but the effects did not conform to the researchers' hypothesis. Thus, the results of the laboratory experiment (Part I) in the second experiment provide support for the equivalence of questions stems constructed according to the ICBEN method, while the results of the hypothetical experiment (Part II) are inconclusive in that a systematic difference between the format types was not found.

## **Acknowledgements**

The studies reported in this paper were supported financially by two Grants-in-Aid for Scientific Research (Nos. 09650653 and 12650599) from the Japanese Ministry of Education, Culture, Sports, Science and Technology. The authors gratefully express their appreciation to Dr. James M. Fields for inspiring this research and to Dr. R.F.S. Job and Dr. J. Hatfield of the University of Sydney, Dr. B. Rohrmann of the University of Melbourne and Dr. K. Shepherd of the NASA Langley Research Center for their cooperation in the administration of Experiment II.

## References

- [1] J.M. Fields, R.G. d Jong, T. Gjestland, I.H. Flindell, R.F.S. Job, S. Kurra, P. Lercher, M. Vallet, R. Guski, U. Felscher-Suhr, R. Schumer, Standardized general-purpose noise reaction questions for community noise surveys: research and a recommendation, *Journal of Sound and Vibration* 242 (2001) 641–697.
- [2] K. Masden, T. Yano, Equivalence of noise annoyance scales in Japanese and English: an experiment using bilingual subjects, *Acoustical Science and Technology* 25 (2004) 144–152.
- [3] T. Yano, J. Igarashi, J. Kaku, K. Kanda, T. Kaneko, S. Kuwano, Y. Nii, T. Sato, M. So, I. Yamada, Y. Yoshino, International joint study on the measurement of community response to noise: the validity of noise annoyance modifiers and question wording in Japanese, *Journal of the Acoustical Society of Japan* 58 (2002) 165–172 (in Japanese).
- [4] K. Kanda, J. Igarashi, J. Kaku, T. Kaneko, S. Kuwano, Y. Nii, M. So, T. Sato, I. Yamada, T. Yano, Y. Yoshino, International joint study on the measurement of community response to noise: comparison of Japanese noise annoyance modifiers between age brackets and areas, *Journal of the Acoustical Society of Japan* 58 (2002) 93–100 (in Japanese).
- [5] T. Sato, T. Yano, T. Morihara, Relationships between rating scales and community responses to railway noise, *Journal of Sound and Vibration* 277 (2004) 611–618.