

inter.noise 2000

*The 29th International Congress and Exhibition on Noise Control Engineering
27-30 August 2000, Nice, FRANCE*

I-INCE Classification: 6.6

A TEST OF THE EFFECT OF QUESTION WORDING ON MEASURED NOISE REACTIONS: A COMPARISON OF JAPANESE AND ENGLISH QUESTIONS

J.M. Fields*, **T. Yano****, **K. Masden*****, **J. Hatfield******, **R.F.S. Job******

* Independent Researcher, 10407 Royal Road, 20903, Silver Spring, Maryland, United States Of America

** Department of Architecture and Civil Engineering, Kumamoto University, Kurokami 2-39-1, 860-8555, Kumamoto, Japan

*** Kumamoto Gakuen University, Oe 2-5-1, 862-0971, Kumamoto, Japan

**** Department of Psychology, University of Sydney, 2006, Sydney, Australia

Tel.: 301-439-4356 / Email: jfields@capaccess.org

Keywords:

NOISE, ANNOYANCE, RESPONSE, COMMUNITY RESPONSE

ABSTRACT

An experiment conducted in three countries has found that small, conventional variations in the wording of noise reaction questions do not affect respondents' answers. The respondents' response scores were similar regardless of whether respondents were asked for their "rating" of a noise environment, their personal annoyance with the environment or the extent they worry about the environment. Responses concerning similar noises were not systematically higher or lower in Japan than in two sites in English speaking countries.

1 - INTRODUCTION

The following question is one of two recommended for measurements of people's reactions to noise by an international group of community noise researchers [1]:

"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; Extremely, Very, Moderately, Slightly or Not at all?"

The answer scale words (Extremely,... Not at all) were selected for comparability across nine languages on the basis of empirical research. However, experts' judgments, rather than empirical data, were used to select phrases in the stem of the question for other languages that would be comparable to the general negative concept of "how much does noise from... bother, disturb, or annoy you" that is present in the English version. When a Japanese/English laboratory study was conducted and it was found that the Japanese subjects systematically expressed more annoyance, differences in the format of the stem of the question were a possible explanation. In this instance the Japanese format, unlike the English format above, followed the standard polite Japanese practice of asking for a characterization of the noise ("urusai") without directly mentioning "you" and the respondent's personal reaction [2]. One possible explanation for this finding is that Japanese are more sensitive to noise. A second, more general possibility that could have implications for all surveys is that differences in the formatting of the question stem wording may affect measured reactions. The present study is designed to test both possibilities with questions that vary the question format but use the same comparable annoyance response scales. The present study thus evaluates the intensity of the reaction and supplements the knowledge about conceptual framework for such reactions that has been addressed in previous work [3].

2 - DESCRIPTION OF STUDY PROCEDURES

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

- Part I: reactions to the 16 recorded sounds played during the test session (three formats),
- Part II: reactions to 10 hypothetical community noise situations (three formats),
- Part III: reactions to 22 additional noise and non-noise environmental problems (uniform format).

The subjects consisted of 202 Japanese students at the University of Kumamoto tested in three groups in standard classrooms with a single loudspeaker, 36 students at the University of Sydney (Australia) tested in eight groups in a small classroom, and 30 employees at NASA Langley Research Center in Hampton, Virginia (USA) tested in three test sessions in an acoustically treated, psychoacoustic test room in which sounds were presented via eight uniformly-distributed, high-fidelity loudspeakers mounted in the ceiling. Each subject completed one of the six versions of the questionnaire. Tests were conducted from October of 1999 to March of 2000. The carefully translated questionnaire and strictly prescribed instructions were developed by a team of Japanese and English-speaking researchers at Kumamoto University in close consultation with researchers at Sydney University.

Each subject was exposed to only one of the three following formats of the question stem in Parts I and II of the questionnaire (key Japanese phrases are given in square brackets):

- **Format A ("You"):** *How much would this (... noise...) bother, disturb, or annoy you?* [... urusaku, matawa hukai ni kanjiru ...] [Standard English question].
- **Format B ("Rate"):** *How bothersome, annoying or disturbing should this (... noise...) be rated as?* [... urusai, matawa kininaru ...] [Impersonal, rating question format].
- **Format C ("Worry"):** *How much would this (... noise...) worry, irritate, or concern you?* [... ni yotte dore kurai nayamasare masu ka ...] [Most extreme question format].

The least extreme, easiest question is hypothesized to be Format B, the format that is used in standard, polite Japanese in which ratings are requested but the subject's own feeling is not directly mentioned. Because the formats varied in the extent to which a subject must reveal personal feelings or the extremity of the reaction mentioned it was hypothesized that in each language the subjects would most often agree with the impersonal "rating" format (Format B) and least frequently agree with the more extreme "worry" format (Format C). It was therefore expected that answers chosen in response to the "worry" format would be lower on the response scale.

Every subject used both the previously described 5-point verbal answer scale and a 0-10 point numeric scale by answering half of Part I and half of Part II with the verbal answer scale and half with the numeric scale. The order in which the scales was presented was balanced within subjects so that the order of the question stems was the same and only the order of the answer scales was varied to yield six versions of the questionnaire. Thus half of the subjects for each Format (A, B, or C) received the verbal answer scale first in Part I (noises #1 to #12) and in Part II (common noise situations #1 to #5) while the other half received the verbal scales for noises #13 to #20 and common noise situations #6 to #10. The questions for Parts I and II were developed to test for question stem format effects within each language, but not to be strictly comparable between the two languages. The questions for Part III, however, were written to be as equivalent as questions in different countries and languages could be. This equivalence was sought by choosing situations that were present in all countries, using objective behaviors rather than subjective terms to describe each situation (e.g. "trucks that make your television hard to hear", not "noisy trucks"), checking for the comparability of terms within English (e.g. "junkyard" was rejected because it is an American term not used in Australia) and having additional translators make multiple checks on the comparability of the phrases used in Japanese and English. Data from the 22 hypothetical noise and non-noise situations rated in Part III of the questionnaire are used solely for comparing the reactions to the same situations by subjects from different cultures and study sites. For Part III all subjects received the following invariant format and answer scale:

<i>"If you had this problem, how annoying or unpleasant would this problem be for you?"</i>										
0	1	2	3	4	5	6	7	8	9	10
<i>Not at all</i>					<i>Extremely</i>					

Table 1.

The order of presentation of the 22 questions in Part III was balanced by reversing the order for half of the subjects, those who had earlier had the numeric scales first in Parts I and II of the questionnaire.

The 16 sounds rated in Part I of the questionnaire were 30-second recordings of road traffic exposures from a single location near an expressway. The sounds were prepared on a CD for playback at about 56, 64, 72 and 80 dB (LAeq) after being calibrated using a pink noise test sound. Levels were measured at from 9 to 23 subjects' positions in each room and used to estimate the levels at each subject's position for each of the 16 noise test exposures. The noise levels were very similar for the same test sound at all seats in the NASA test facility (within 2.5 dB at different seats) but varied by as much as 8 decibels between different positions in the classrooms at the university sites. All analyses were based on these actually measured test sounds. The six questionnaire versions were distributed around the room so as to not correlate question version with noise exposure. After the initial four test sounds, the 16 rated sounds were presented at the four noise levels in a Latin squares design.

3 - RESULTS FROM QUESTION FORMAT COMPARISONS

The responses to the 16 recorded sounds are contrasted for the three questionnaire formats for English in Fig. 1 and Japanese in Fig. 2. In these figures the five points of the verbal scale have been scored 0, 2.5, 5, 7.5 and 10 so that the range matches the 0-10 numeric scale. Figures 1 and 2 show that there is not a systematic tendency for one format to elicit any stronger annoyance reactions than does any other format. Multiple regression analyses of responses on noise level and type of format confirmed these findings. In these analyses the "worry" format was estimated to be scored about 0.33 points higher with a 95 percent confidence interval of ± 0.51 that meant that the difference was not statistically significant ($p > 0.05$). With the regression coefficient for noise level being about 0.27, the 0.33 difference is the equivalent of the effect of only about 1 decibel ($1.2 = 0.33/0.27$). This difference is therefore not statistically significant. The findings were also not consistent with the direction predicted by our hypothesis (Format B, "impersonal rating", generated the lowest frequency of responses in both languages) and not consistent across languages (Format C, "Worry", generated slightly, but not significantly, higher reports in Japanese than Format B while the reverse was true in English). The 95 percent confidence intervals for the estimates from the Japanese language with 202 subjects were precise enough to reject any large effects (the equivalent of less than ± 2 dB(A)) while those for the English language (only 66 subjects) are so wide as to not be able to reject some important effects (the equivalent of about ± 4 dB(A)). The same pattern of no significant relationships was found when dummy variables were introduced into the regression to control for the study site (Sydney University or NASA) and the type of response scale (verbal or numeric). Similarly no relationship was present when the answers were scored at their midpoints following a procedure used by Miedema [4] or when the verbal scales were given the scores assigned through word scoring studies conducted in English and Japanese [2].

The three questionnaire formats were also examined in Part II when subjects were asked to "... imagine what it would be like if you lived in a home with this noise" and then, using the same format as previously, to answer questions about 10 noises that ranged in severity from the least annoying of "Hearing a distant aircraft about once a week" to the most annoying of "Hearing big trucks (when you are in your home) every time the traffic signal changes at a nearby intersection". The remaining eight noises, in order of ascending severity, concern dogs barking at night, a door squeaking, neighbors' TV sound, background music from a business, backup warning signals on trucks, being awoken by motorcycles, aircraft interfering with TV, and speech interference from traffic. The average of the subjects' reactions to these ten noise situations is given for each study site and format in a bar in Fig. 3. In Fig. 3 differences between formats are small and not consistent across study sites. Analyses of variance show that the differences between formats are not statistically significant for a simple univariate analysis of variance, separate univariate analyses for each study site, or for a multivariate analysis that includes study site as well as format type. Analyses showed no differences when only numeric scales or only verbal scales were examined. When the 10 questions were examined individually for each of the three study sites for the two scales (verbal or numeric), 59 of the 60 tests were not statistically significant ($p < 0.05$). The sole statistically significant effect was from Japanese ratings of motorcycles using the numeric scale, an effect that was not significant when using the verbal scale.

A principal components factor analysis of the 10 reaction questions extracted a factor that accounted for 33 percent of the variance and loaded positively on reactions to hearing music from businesses and neighbors and negatively on the low noise-annoyance items. The finding for this measure was consistent with the other findings; individuals' scores on the factor were not related to the question format.

4 - RESULTS FROM STUDY SITE AND LANGUAGE COMPARISONS

Part III of the study provides the evidence on site differences in sensitivity to noise. Four of the 22 questions in Part III described hypothetical environmental, transportation noise problem situations while

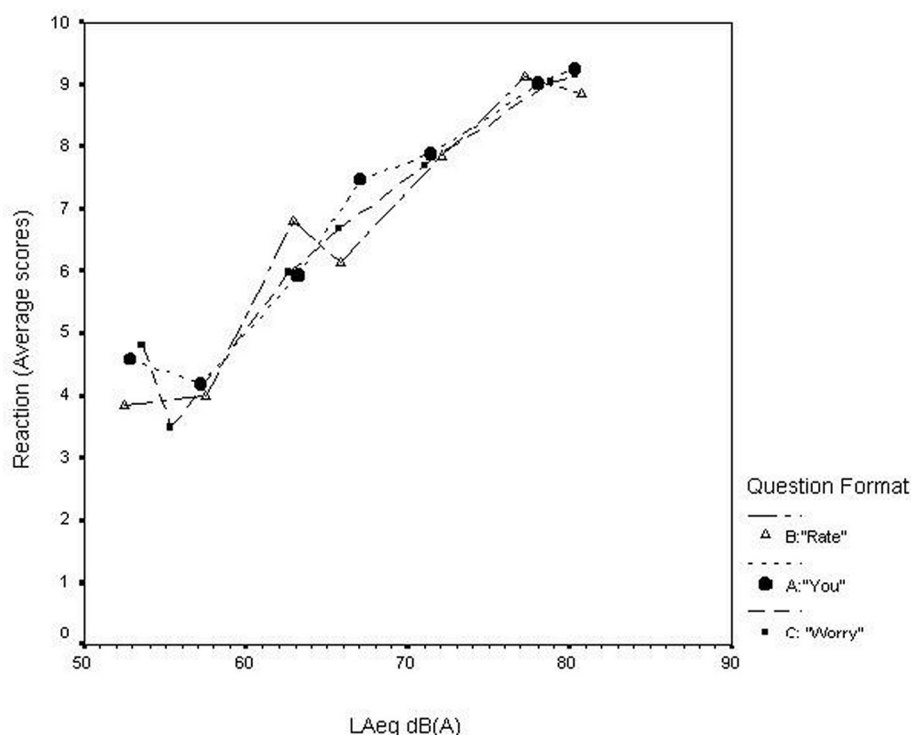


Figure 1: Reactions to 16 noise events for three English question formats.

at home concerning; 1) trucks that interfere with hearing television, 2) awakening by aircraft, 3) hearing a distant expressway, and 4) aircraft speech interference. In Fig. 4, each site's left-most bar represents the average of the reactions to those four environmental noise situations. The next bar represents the average of the reactions to the following three household noise problems: 1) hearing neighbors calling children, 2) hearing the refrigerator, and 3) plumbing noise. Site differences in average reactions to these noises are very small, and not statistically significant (less than 0.2 scale points).

The remaining bar for each site in Fig. 4 represents the average of the reactions to the following fifteen non-noise problems: industrial odors, air pollution, mosquitoes in the home, a burnt-out streetlight, neighbors' trash, forgetting people's names, eyesight problems, automobile not starting, cockroaches, a neighbor's light, factory dirt, hearing loss, sticky doors, and a junk car business. The Sydney University reactions to these non-noise problems are statistically significantly lower than those at the other two sites. As a result the relative sensitivity to noise and non-noise problems is greatest for the Sydney University subjects, least for the NASA subjects and at an intermediate level for the Kumamoto University subjects. There is, again, no evidence that Japanese and English speakers' reactions systematically differ.

Differences between English study sites can be examined with the same data that were used to examine question format effects. In Fig. 3, the comparison of the bars representing the 10 hypothetical noise environments shows that reactions are greater at the NASA site than the Sydney University site. An examination of the scores averaged over the three types of formats from Fig. 3 as well as from a multiple regression analysis of reactions to the recorded sounds finds that the NASA site reactions are somewhat higher than the Sydney University reactions for the 10 hypothetical noise situations (0.6 scale points, not statistically significant) and for the 16 recorded sounds (1.4 scale points, statistically significant). As previously noted, however, there are not significant differences in Part III between the NASA and Sydney University ratings of the four environmental and three home noise situations.

5 - CONCLUSION AND DISCUSSION

These test sessions suggest that differences in the tested alternative common words and sentence constructions for the stem of a noise reaction question do not have large, if any, effects on respondents' answers to questionnaires. Any effects that would as large as those from a two-decibel difference in noise exposure should have been detectable in the Japanese study. About a four-decibel difference should have been detectable with the English data. The tests did not find systematic differences or statistically significant differences between the Japanese reactions and the reactions at the two English sites. There is not, therefore, any support for the assumption that Japanese are systematically more or less sensitive

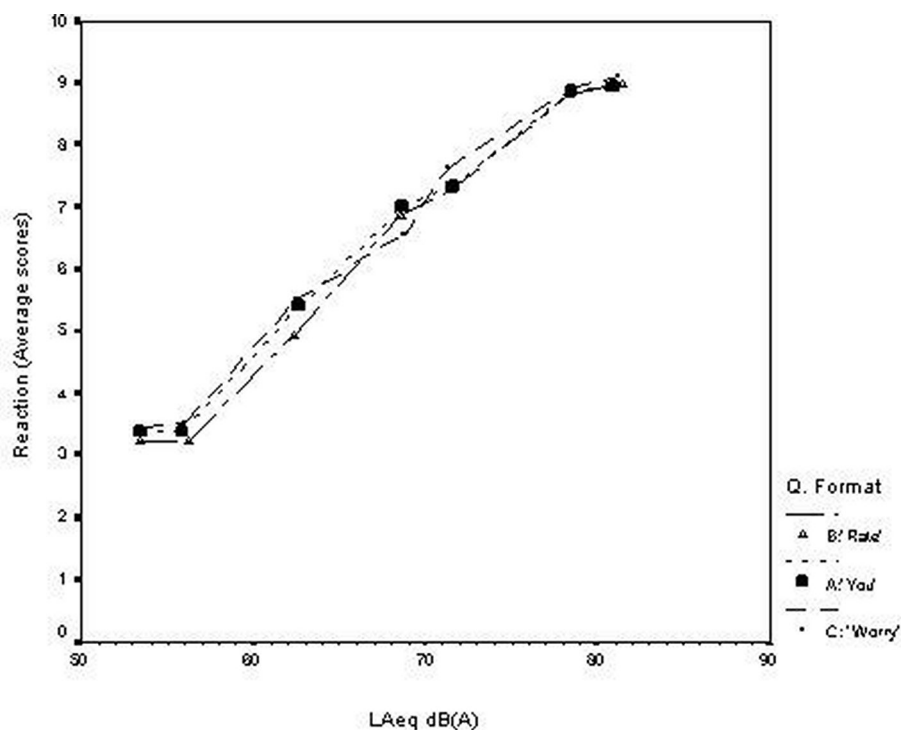


Figure 2: Reactions to 16 noise events for three Japanese question formats.

to noise than are English speakers. Some statistically significant differences between the English sites (e.g. NASA subjects were statistically significantly more annoyed for the 16 recorded noises) were not consistent across the three parts of the study.

Additional studies on test subjects and general populations are needed to provide definitive evidence on both the question format and cultural difference issues. Larger numbers of test subjects under more carefully controlled acoustical conditions could help to determine if there are small question format effects. The significant, but not consistent, differences between the English language sites suggest that other variables, that could best be assessed with general population surveys, may be important. Although the ratings of hypothetical situations economically provide insight into different reactions, definitive tests would require standard survey measurements of reactions to real rather than hypothetical situations.

The evidence presented in this paper suggests that the earlier reported difference between Japanese and English-speaking subjects in a laboratory experiment was not due to wording differences or general Japanese/English cultural differences. It is possible that the earlier finding might be explained by the characteristics of English-speaking subjects who live in Japan, other subtle differences between the test questions or by random variation. A replication of the previous study would be useful, especially if it could be conducted at laboratories in several countries.

REFERENCES

1. **J.M. Fields, R.G. de Jong, I.H. Flindell, T. Gjestland, R.F.S. Job, S. Kurra, P. Lercher, A. Schuemer-Kohrs, M. Vallet, and T. Yano**, Recommendation for shared noise annoyance questions in noise annoyance surveys, In *Noise Effects '98*, pp. 481-486, 1998
2. **T. Yano, K. Masden, J.M. Fields, K. Kanda, and K. Kawai**, An experiment on the equivalence of noise annoyance scales in English and Japanese, In *Internoise 99*, pp. 1331-1336, 1999
3. **M. Florentine, S. Namba, and S. Kuwano**, Concepts of loudness, noisiness, noise and annoyance in the U.S.A., Japan, and England, In *Internoise 86*, pp. 831-834, 1986
4. **H.M.E. Miedema and H. Vos**, Exposure-response relationships for transportation noise, *Journal of the Acoustical Society of America*, Vol. 104, pp. 3432-3445, 1998

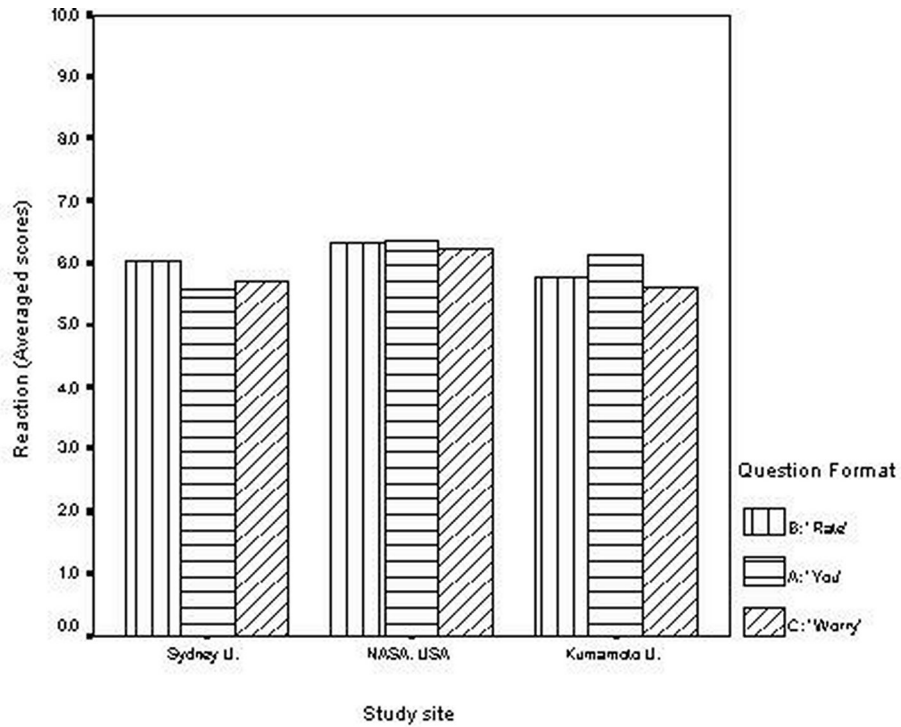


Figure 3: Reactions to 10 hypothetical noise situations for three question formats.

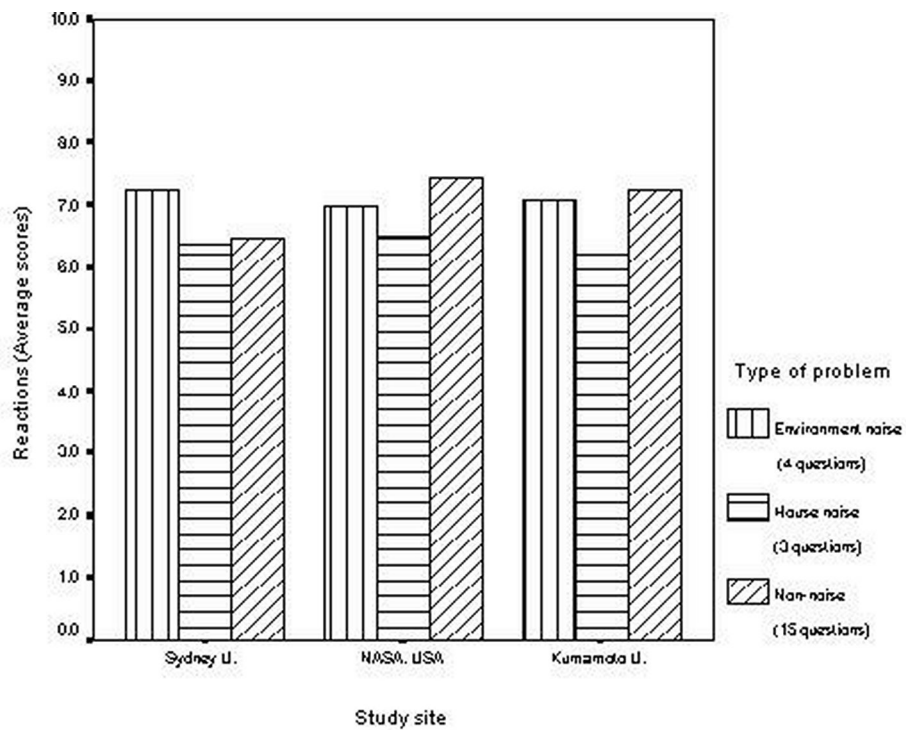


Figure 4: Averaged ratings of the same 7 noise and 15 non-noise problems by each site's subjects.