Jury studies of sound quality for low frequency noise

YAN Liang, CHEN Ke-an

(Department of Environmental Engineering, College of Marine, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Using 75 varied low frequency noise samples with adjusted-loudness as sound stimuli, we investigate sound quality for low frequency noise with paired comparison method. The jury testing with a total of 24 jurors has been carried out to explore the relationship between physical properties of low frequency noise and perceptual ratings. It is shown that the proposed 'unpleasantness' can be used as an reasonable and effective metric to indicate human's general reaction to low frequency noise, and some crucial factors such as low frequency components and its energy ratios, loudness, temporal and spectral characteristics, and sharpness could influence or even govern the sound quality of low frequency noise.

Key words: low frequency noise; sound quality; unpleasantness; paired comparison

低频噪声声品质评价实验研究

闫 靓, 陈克安

(西北工业大学航海学院环境工程系, 西安 710072)

摘要: 以等响处理后的 75 个稳态低频噪声样本为例, 采用成对比较法, 设计并完成了有 24 位评价者参与的大规模实验室主观评价实验。研究发现: 1)"不愉悦度"能够充分体现人对低频噪声的主观感受, 评价效果理想; 2) 存在决定低频噪声不愉悦感的频率或频带; 响度在决定稳态低频噪声的不愉悦感程度中占主导地位, 尖锐度的影响亦不容忽视; 与总声能相比, 低频段声能比与频谱形状在决定不愉悦程度时, 作用更大一些。

关键词: 低频噪声; 声品质; 不愉悦度; 成对比较法

中图分类号: O42　　文献标识码: A　　文章编号: 1000-3630(2006)-06-0540-07

1 INTRODUCTION

The character of sound that relates to acceptance is called sound quality, which may well become a large role in determining satisfaction[1]. As a very active and important research area, sound quality evaluations have been applied into automobile and electric appliances industry over last two decades. By psychophysical means, sound quality ev-aluation and prediction has advanced considerably, but studies of low frequency noise are rare.

Low frequency noise is a major component for many occupational and residential noise which has behaved as common background noise in urban environment but receives less attention. The primary perceived and most frequently reported quality characteristic of low frequency noise is not that of loudness or noisiness, but that of annoyance[2]. How-

Received Aug. 31, 2006; Revised Oct. 30, 2006

ever, the degree of annoyance or disturbance generated by a specific noise, regardless of frequency, is difficult to predict accurately by practical measurements[3].

Although A-weighted sound level (dBA) has proven itself useful as an approximated estimation of annoyance for medium and high frequency stationary noise, it severely underestimates annoyance as well as loudness when low frequency components dominate the noise spectrum. Bryan[4] found that the noise with low frequency components and high sound level gave rise to vigorous complaints even though the exposure level was only around 55 dBA. Tempest[5] found that the number of complaints were far larger than that predicted from A-weighted sound pressure. Persson and Björkman[6] compared four broadband fan noises centered at 80, 250, 500, and 1000 Hz, and found that the noise centered at 80 Hz was perceived to be significantly annoying than the other noises with the same A-weighted levels. Taking advantage of the psychophysical magnitude estimation technique, Broner and Levebthall[7] found that the B-weighted sound level is the most suitable measure in predicting the annoyance due to higher level (90~105 overall SPL) for low frequency noise. However, it remains a question whether a similar conclusion is valid for low-level and low frequency noise.

Moreover, insufficient measurement and control of the frequency range and harmonics may be identified as potential problems in both field recordings and experimental generation of low frequency noise[8]. The earliest studies employed exposure levels which would almost certainly not be allowed today. In Denmark a set of guidelines for measurement and assessment of environmental low frequency noise, infrasound and vibration were published in 1997[9].

Hence, it is very important to establish a mathematical model for predicting the sound quality for low frequency. This paper describes laboratory studies about the influence of the interference caused by several common low frequency noises on sound quality. To measure the quality of low frequency noise, the paired comparison method is used in jury testing with a total of 24 jurors to explore the relationship between physical properties of low frequency noise and perceptual ratings. A measurable perceived quantity named "unpleasantness" and the corresponding scales are proposed. Based on the jury tests, an empirical unpleasantness predictive model is obtained, which gives a quantitative appraisal of subjective response to low frequency noise initially.

2 LOW FREQUENCY NOISE QUALITY AND UNPLEASANTNESS

Numerous researches have indicated that the noise quality is not described sufficiently by a one-dimensional measure. The multi-dimensional character of both acoustic and human perceptual response can be studied by questionnaire about sound quality descriptors, factor analysis and cluster analysis. Then, two fundamental jury tests namely paired comparison and semantic differentials are often adopted to deal with the multi-sensory dimensions in sound quality[1].

On the analogy with the existed sound quality descriptor "annoyance", the authors have presented a new metric, named unpleasantness (expressed as 'up'), which can indicate the actual feelings of human beings for low frequency noise environments, and is more suitable for laboratorial research. The authors have pointed out that the 'unpleasantness' can indicate the low frequency noise quality more veritably[10], which is divided into 5 degree ranks[11], namely 'not at all', 'slight', 'moderately', 'very' and 'extremely'. We think it seems a good choice to use the 'up' as an objective estimator for low frequency noise quality because its applicable conditions and scopes are more explicit, and it can be more suitable to address the emotional response to low frequency's interference.

It is assumed that there are m subjects randomly selected to take part in a set of jury tests, n of them give ranks upwards of 'moderately' (including 'moderately'), and then the up can be calculated as follow,

$$up=\frac{n}{m}\times 100\% \quad (m\geq 20) \tag{1}$$

Theoretically speaking, the numbers of subjects in the jury test can be decided by the accuracy and distribution of the test results, however, only the empirical approach can be used to ensure how many subjects should take part in the jury test. In 1999, by comparative analysis, Hempel and Chouard found that no distinct diversity existed between the 54-subjects and the 20-subjects jury test[12] and both the average values and the standard deviations of them were nearly consistent. Then they proposed that for most psychoacoustic studies, 20-subjects were enough. Following this, in the present study is m≥20 used.

3 PAIRED COMPARISON BASED JURY TESTING

Jury testing is an essential step in sound quality evaluation[13]. At present, the commonly used subjective assessment methods mainly include making scores, ranking orders, paired comparison and semantic differentials[14]. In general, for different testing tasks, corresponding methods should be applied to achieve good performances. In this study, paired comparison based on jury test is selected to characterize the low frequency noise quality.

3.1 Recording of noise samples

Low frequency noise samples from car interior, ventilation systems and transformers in community were respectively recorded by a BSWA VS302USB Spectra Plus dual channel spectrum analyzer equipped with a B&K Type 2230 integrating sound level meter. After real-time spectral analysis, recording, playback and post-processing are accomplished, 9 effective appraisal samples were obtained. Then 100 stable samples were downloaded from Internet, which have obvious low frequency characteristics and oriented from various vehicles, machines and equipments under different operation situations as well as natural sounds.

Theoretically speaking, large numbers of samples can ensure the results' reliability and accuracy, but lead to enormous workload and overtime experiments, even cause hearing fatigue and affect the appraisal results ultimately. Thus, 60 optimal samples are selected from various types of low frequency noise, from which we could find out the subjective response rules and establish a preliminary sensation model.

Before the formal jury test, all samples are pretreated as follows:

(1) Remove the unstable samples and length interception is done to preserve typical low frequency characters. All samples have the same temporal dimension, 10 seconds for playing.

(2) Equalize the sound level of every sample, and make sure that the sound level differences between every evaluating spot are smaller than 3 dBA.

(3) Match the defined frequency range (lower limits 5, 10, 16, 20, 50Hz, and upper limits 100, 150, 200, 250, 500, 600Hz) in pairs randomly, and then calculate the corresponding low frequency energy and low frequency energy ratio, respectively.

3.2 Subjects

Twenty four subjects, 12 males and 12 females who have an average age of 23 years from 18 to 27, took part in the jury test. They were either university staff or postgraduate whose majors are acoustics or environmental science, all of whom had normal hearing.

Facing terminal loudspeaker, the subjects were seated at evaluating spots distributing at a semicircle with a radius of 3m in the laboratory. Prior to the formal tests, three pre-training phases were carried out for the subjects: getting knowledge about noise samples, explaining meaning of sound quality metrics, and describing judgment tactics and conversation. Finally, every subject was asked to recollect their feelings and to give advices for the tests. Depending on these, we could aware the subjects' comprehensive extent of the Chinese adjective describing low frequency noise quality metric-unpleasantness.

3.3 Method

An important aspect of paired comparison with forced choice is the decision of how to form pairs in advance. In our test, the total of 75 samples was divided into two sets, 15 recording samples in one and 60 optimal samples in the other, which adopt two different pairing designs, respectively. Each coupled set is presented to the subjects who were instructed to decide his preference each time, by answering questions such as: 'choose which one is more unpleasant'.

3.3.1 15 samples pairing design

According to traditional complete designs, 15 effective independent samples are paired randomly with no restrictions on pairings. Then, a sequence of 255 couples, including 15 i-i comparisons and 105 j-i comparisons, is formed. In order to minimize order effects, these couples were presented in a totally random order for each subject. Each couple was presented for 25s (20s playback duration, 5s interval for evaluating). Considering adverse effects and fatigue caused by low frequency noise, we divided all sample pairs into two groups. The mean testing duration for each group was 30 min. with 1h break.

3.3.2 60 optimal samples pairing design

The complete pairing design would result in an overall test session with 25 hours for each subject individually. Nevertheless, as an empirical rule, every session in the jury test should be performed within 20~30min, not more than 45min. And the interval time between every two sessions should be not less than 30min. Therefore, the incomplete pairing design which the same sample pairs (i-i) and the invert order pairs (j-i) were not demanded was employed to fulfill the above requirements. Concerning validity, 20 j-i pairs and 10 i-i pairs were selected and added to the sequence randomly, resulting in a total of 1800 pairs which is separated into 30 groups. The subjects carried out this task on 3 separated days.

3.4 Analysis

The testing data were analyzed with the SPSS software as follows: 1) according to the obtained reference scale described in section 2, measurable attributes of low frequency noise quality are translated into the corresponding numerical values; 2) for the subjective data processing, i-i and ij-ji error rate analysis, normal distribution tests, rank scale analysis are successfully adopted; 3) for all samples, the correlation of the mean opinion score data and 3 calculated psychoacoustic parameters are analyzed; and 4) with the aid of principal components analysis and multiple linear regression analysis, an empirical unpleasantness predictive model for low frequency noise is obtained as follows.

$$UP=0.834N+0.156R+0.041S \tag{2}$$

where N, R and S are respectively the loudness, roughness and sharpness and can be calculated by Zwicker's model[14], and UP is the mean value of the unpleasantness opinion scores from each subject. This model shows a nearly perfect performance (the goodness of fit is about 96%, the standard error is about 95%) in the prediction of the mean opinion score data of various low frequency noise samples.

4 RESULTS AND DISCUSSIONS

Combining the unpleasantness does-response curves and cause analysis for low frequency pure tones[8], we have following results:

(1) Considerable variations of unpleasantness as well as roughness and sharpness can be observed because of the differences of the amplitude of low frequency components before and after loudness enhancement.

(2) A direct proportional relationship between sharpness and roughness can be given. It is con-

cluded that loudness is a dominant feature for any sound quality evaluation, can also give sound intensity information about the low frequency content of noise samples. Loudness enhancement can induce smaller masking by high frequency content in the sample itself or in the background noise, and then roughness tends to increase due to the corresponding enhancement of the low frequency components. Oppositely, for the sharpness of noise samples the tendency is to decrease.

(3) For most samples, unpleasantness decreases with the increase of roughness and sharpness. This result demonstrates that unpleasantness from low frequency noise is much higher than that from noise without dominating low frequency components at the same dBA level. Whereas there is an opposite relationship for the recording samples. This discrepancy is probably due to the affected low frequency components under the actual settings, the low frequency response and fidelity of the recording and replaying equipments, and so on.

(4) Taking $M_1=-0.835N+0.681R+0.550S$ and $M_2=0.024N-0.597R+0.776S$ as principal-components, a model consisted of $F_1=0.571N-0.465R-0.376S$ and $F_2=0.025N-0.623R+0.809S$ can be implemented. This outcome shows that the descriptor 'unpleasantness' is a two dimensional quantity, and there are two different kind of judgment tendency of subjects. The first dimension of unpleasantness has a positive relation with loudness as well as a negative relation with both of roughness and sharpness. But, the second dimension of unpleasantness has a positive relation with sharpness which is stronger than its negative relation with roughness, moreover, its positive relation with loudness becomes inconspicuous. In spite of an attribute of significant spectral components at high frequencies, sharpness can affect the low frequency noise quality due to ambiguous mechanism.

(5) Figure 1 gives the power spectra of the sample 13, 14, and 4 used in the jury tests. Table 1 shows 30 low frequency energy ratios. The ranking number based on the mean unpleasantness score of each measured noises indicates that the sample 13 is the most unpleasant, then is the sample 14, followed with the sample 4. It seems that low frequency

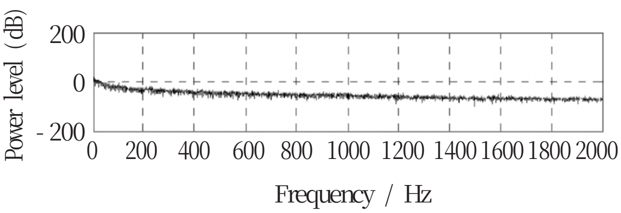
Table 1 30 low frequency energy ratios of the sample13, 14, 3

Low frequency region (Hz)		5~100	5~150	5~200	5~250	5~500	5~600
Energy Ratio%	13	7.2295	7.2569	7.2629	7.2688	7.2727	7.2731
	14	6.8549	6.8645	6.8670	6.8681	6.8703	6.8705
	4	7.6298	7.6523	7.6565	7.6592	7.6655	7.6667
Low frequency region (Hz)		10~100	10~150	10~200	10~250	10~500	10~600
Energy Ratio%	13	4.0250	14.0526	4.0586	4.0644	4.0684	4.0687
	14	3.4768	3.4864	3.4889	3.4900	3.4922	3.4924
	4	2.9651	2.9877	2.9918	2.9946	3.0009	3.0021
Low frequency region (Hz)		16~100	16~150	16~200	16~250	16~500	16~600
Energy Ratio%	13	2.2261	2.2536	2.2596	2.2654	2.2694	2.2697
	14	1.4849	1.4946	1.4971	1.4981	1.5003	1.5005
	4	1.2626	1.2852	1.2894	1.2921	1.2984	1.2996
Low frequency region (Hz)		20~100	20~150	20~200	20~250	20~500	20~600
Energy Ratio%	13	1.6279	1.6554	1.6614	1.6672	1.6712	1.6715
	14	0.82339	0.8330	0.8355	0.83655	0.83875	0.83895
	4	0.81218	0.8347	0.83889	0.84159	0.84798	0.84913
Low frequency region (Hz)		50~100	50~150	50~200	50~250	50~500	50~600
Energy Ratio%	13	1.6279	1.6554	1.6614	1.6672	1.6712	1.6715
	14	0.09863	0.10824	0.11073	0.11178	0.11398	0.11419
	4	0.052826	0.075347	0.079532	0.082236	0.088622	0.089772

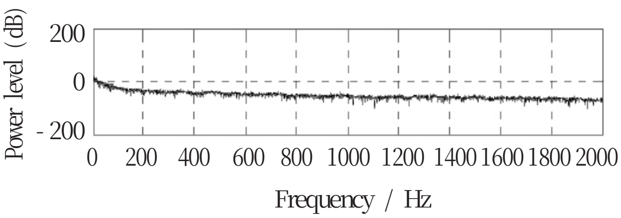
noise quality would be influenced by not only the entire sound energy but also the temporal envelope, spectral shape and low frequency energy ratio.

(6) Unpleasantness rating based on the jury test results differs from that on the entire noise energy, which may be due to the fact that dBA is not the best estimator for low frequency noise quality. Because of masking by high frequency components in actual and experimental settings, replaying levels are much larger than recording levels especially for the recording samples. Therefore loudness equalization should be preferred. Some uncertain disturbing factors such as equipments performance and experiment conditions must be considered.

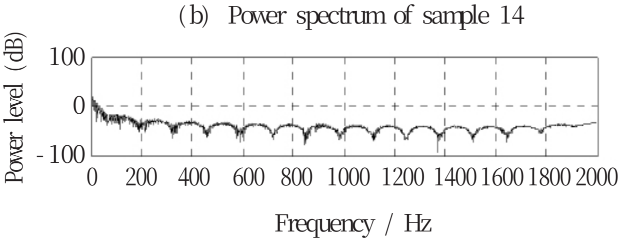
(7) At very low frequency human's auditory system has not response in the normal sense, and the common acoustical metrics are useless in describing perceptible. However, the subjective reactions such as unpleasantness arise and unpleasantness may contribute in complicated ways. The present jury studies indicate that, generally, the observable unpleasant noises possess prominent low frequency components in the range of 5~20Hz. It is obvious that the lower limit of low frequency noise at 20Hz is not suitable and further studies need to be performed.



(a) Power spectrum of sample 13



(b) Power spectrum of sample 14



(c) Power spectrum of sample 14

Fig.1 Power spectra of the sample 13, 14, and 4

5 CONCLUSIONS

It is shown from this study that unpleasantness is an alternative descriptor of low frequency noise quality and is simple and feasible for the laboratory investigation. As a two dimensional perceived quantity, unpleasantness can give reliable and representative results to represent typical low frequency noise experienced by the occupants. The reported unpleasantness model for low frequency noise quality is a reasonably predictor of the respective jury test results (Goodness of Fit =0.96). Loudness plays a dominating role in shaping perceptions of low frequency noise quality, furthermore, some crucial low frequency components and its energy ratios, temporal and spectral characteristics, sharpness could influence or even govern the sound quality of low frequency noise. In spite of the fact that the proposed estimator is initially developed for very few kinds of low frequency noise, the results obtained for other stable noises could also be meaningful, especially for low frequency noises similar to the testing samples.

ACKNOWLEDGMENTS

This work is a part of a project No. 10574104 supported by NSFC (Natural Science Foundation of China). The authors are grateful to the experimental fellows.

REFERENCES

[1] U Jekosch. Meaning in the context of sound quality assessment[J]. Acta Austica, 1999, 85(5): 681-684.

[2] N Broner. The effects of low frequency noise on people-a review[J]. Journal of Sound and Vibration, 1978, 58: 483-500.

[3] C M Haslegrave. Auditory Environment and Noise Assessment. Evaluation of Human Work[M]. edited by J.

R. Wilson and N. Corlett., London: Taylro & Francis. 1990. 406-439.
[4] M E Bryan. Low Frequency Noise Annoyance. Infrasound and Low Frequency Vibration[M]. edited by W. Tempest, London. Academic. 1976. 65-69.
[5] W Tempest. Loudness and Annoyance due to Low Frequency Sound[J]. Acustica, 1973, 29: 200-205
[6] K.Persson, M. Bjӧrkman. Loudness, Annoyance and dBA in Evaluating Low Frequency Noise[J]. Journal of Low Frequency Noise and Vibration. 1990, 9: 32-45.
[7] N Broner, H G Levebthall. A Criterion for Predicting the Annoyance due to Higher Level, Low Frequency Noise[J]. Journal of Sound and Vibration. 1982. 84(3): 443-448.
[8] B Berglund, P Hassmén. Sources and Effects of Low frequency Noise[J]. Journal of the Acoustical Society of America. 1996. 99(5): 2985-3002.
[9] J Jakobsen. Danish Guidelines on Environmental Low Frequency Noise, Infrasound and Vibration[J]. Journal of Low Frequency Noise, Vibration and Active Control, 2001, 2(3): 141-148.
[10] YAN Liang, CHEN Kean. Subjective Evaluate on Unpleasantness of Low frequency Pure Tone[J]. Chinese Applied Acoustics, 2006, 25(5): 319-324.
[11] D Botteldooren, A Verkeyn, A Fuzzy rule based framework for noise annoyance modeling[J]. Journal of the Acoustical Society of America, 2003, 14(3): 1487-1498.
[12] YAN Liang. Objective Evaluation of Subjective Reaction to Low frequency Noise[D]. Master's Degree Thesis, Northwestern Polytechnical University, 2006.
[13] B. R. Giovanni, F. Crenna, M. Codda. Measurement of Quantities Depending upon Perception by Jury-Test Methods[J]. Measurement, 2003, 34: 57-66.
[14] H. Fastl. The psychoacoustics of sound-quality evaluation[J]. Acta Acustica, 1997, 83(3): 754-764.

简 讯

中国声学学会检测声学分会 2006 年活动信息

2006 年 10 月 20 日中国声学学会第六届全国会员代表大会期间，检测声学分会在李明轩主任召集下召开分会委员座谈会，就检测分会近来的工作及下一年的活动进行讨论。大家一致认为有必要加强有关领域的热点问题的交流和讨论，包括与其他分会如医学超声分会的交流和讨论，以进一步促进检测声学各领域学术和技术的进步、发展和融合。座谈会达成共识，将在 2007 年夏召开一次以近年来热点问题为专题的讨论会，并拟邀请医学超声分会有关专家就相关专题作会议报告。参加座谈会的委员有检测声学分会副主任钱梦騄教授、程建春教授、王小民教授以及医学超声分会主任张海澜教授等。

中国船舶工业第九设计院　吕玉恒