Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

Simulation and verification of nonparametric psychoacoustical experiment methods

Chen Guobin, Xie Zhiwen

Acoustic Lab., Physics Dept., School of Science, South China University of Technology, Guangzhou, 510640, China

PACS: 43. 66

ABSTRACT

Adaptive methods are usually used in the current psychoacoustical experiments. There're roughly two types of adaptive methods--parametric and nonparametric method. The latter is widely applied in the simple estimations of threshold because of its intuition and convenient operation. The present work develops a simulation programme and makes experimental verifications for several nonparametric methods such as Transformed Up-Down, Parameter Estimation by Sequential Test (PEST), stochastic approximation & accelerated stochastic approximation (SA&ASA). Simulation programme establishes a type of psychometric function (PF) in advance, and makes use of Monte-Carlo method to simulate the processes of the different psychoacoustical methods. The statistical results of each 10000 simulations indicate that, with the same number of trials, starting level, target probability and PF, the standard deviation of ASA is the smallest of all, which means its convergence is the best. However, the results of psychoacoustical experiment shows that the influence of subjective factors can not be ignored, and standard deviation can not be a reliable standard for the comparison among different methods. In the comparison of several items, the overall behaviour of ASA is still the best. Although the stability of ASA would be affected by the fluctuation of the threshold, it can be improved by some procedures.

Psychoacoustical methods

The research on psychoacoustical methods has been developed for about 100 years. In this period, psychophysicists suggested many theoretically excellent and efficient methods. However, because of the particularity of the different psychophysics problems studied, different methods have their own application areas and the parameters of each method are not totally the same. Therefore, it is necessary to compare their performance from different aspects.

The topics psychophysics faced are various, one of them which attract researchers mostly and is widely applied in practice is the measurement of threshold. Based on the experimental threshold results of the predecessors', a concept of psychometric functions (PF) is put forward.

In psychoacoutics, questions with positive and negative responses are usually used to communicate with the subjects and the types of response can be roughly classified into two kinds: yes/no and nAFC. Literature [2] points out that the subjectivity of yes/no response is too strong, and the subjects' auditory standard could be different from time to time. However, the nAFC response is much more objective, since the responses are "correct" and "incorrect". The experiment results of nAFC response are more reliable too, since more details can be heard by the comparison between n items. For this reason, the 2AFC response is adopted in the following simulations and experiments.

The PFs of yes/no and nAFC responses are different, but both are based on an ideal PF $\varphi(x)$. The ideal PF has the properties of monotonicity and boundedness with the maximum 1 and the minimum 0. The expressions of PF can be various. Here, a logistic distribution is applied

$$\varphi(x) = \frac{1}{\exp[-k(x-m)]+1} \quad (1)$$

Where x is the physical variable of the signals in the experiment (here is the SPL of maskee), $\varphi(x)$ is the probability of positive response when the signal x is presented, k is the slope and m is symmetric point of PF. When the subjects are tested with nAFC method, the positive probability approaches 1 at high SPL, but $\frac{1}{n}$ at low SPL. The transformation from ideal PF $\varphi(x)$ to nAFC $\varphi_{nAFC}(x)$ is approximately as following equation:

$$\varphi_{nAFC}(x) = p_g + (1 - p_g)\varphi(x) = \frac{1}{n} + \left(1 - \frac{1}{n}\right)\varphi(x) \qquad (2)$$

Where $p_g = \frac{1}{n}$ is the guessing rate of nAFC.

The types of psychoacoustical methods can be roughly catalogued as classical and adaptive. Classical methods are intuitive but lack of mathematical rigor and generally have low efficiency. One of them is constant stimuli method. It is to choose a number of suitably located points of physical stimuli, and repeatedly present them to the subjects, the cumulative responses of whom are used to estimate their shape of PFs. In the following parts, the classical constant stimuli method is just applied as a standard to verify the other experimental results. However, adaptive methods are generally have high flexibility, efficiency and stability, and therefore are widely used in current psychophysical experiments. The adaptive methods can be further segmented into two kinds: parametric and nonparametric. The latter changes the SPL of the signal by certain stepsizes based on the responses of the subjects at early positions, and stabilize the SPL at target position (threshold). Nonparametric method is widely applied in the simple estimations of threshold because of its intuition and convenient operation.

The main task of the paper is to find out the advantages and shortcomings of three nonparametric methods, Transformed up-down (3d1u), PEST, and ASA with MATLAB simulations and psychoacoustic simultaneous masking experiments.

The principles of Transformed up-down, PEST and ASA

The principle of transformed up-down is to change the SPL based on the responses of one or more previous trials and determine the SPL where the probabilities of increasing and decreasing SPL are almost the same (both equal to 0.5). Different rules lead to different target positions. Here we use 3d1u rule, whose target probability is fixed at 0.794[3].

Since a 2AFC response is used in the simulation and experiment. The subjects' PF has been transformed. To obtain the probability of ideal PF, an inverse transformation of equation (2) is needed. From the inverse transformation, we know the probability of ideal PF is 0.588. To compare the performance of different methods under the same condition, 0.588 is also chosen to be the target probabilities of PEST and ASA.

The principle of PEST (Parameter Estimation by Sequential Test) is to offer the subject a group of trials to listen continuously at the same SPL, require him to count the number of positive response, and compare it with the expectation value at the target position to estimate whether the current SPL is higher or lower than the target. If the trials are independent from each other, then each group is an independent multi-experiment and the expectation of positive response number is (under 2AFC)

$$E[N(C)] = \emptyset_{2AFC} * T \qquad (3)$$

Where $N(C)$ indicates the expectation of positive response number, and T is the current number of trials in one group. A floating range w is needed to be set in the experiment. When the subject is hearing the signals in one of the group, if the number is in the range

$$N_b(C) \in [E[N(C)] - w, E[N(C)] + w] \qquad (4)$$

Then he will be asked to continue to hear the group, otherwise, the current group should be finished, and SPL will be changed. The rules of the stepsizes change are introduced by Taylor[4], the inventor of PEST.

ASA (Accelerated Stochastic Approximation) is an improvement version of Stochastic Approximation (SA). In 1958, Kesten proved that ASA is 100% convergent to the target as SA, and converge even faster than SA[5] does. In ASA, there is an expression that:

$$\begin{cases} x_{N+1} = x_N - \frac{c}{N}(Z_N - \emptyset), \quad N \leq 2 \\ x_{N+1} = x_N - \frac{c}{2+m_{shift}}(Z_N - \emptyset), \quad N > 2 \end{cases} \quad (5)$$

Where x_N and Z_N are the SPL and the response (1 for positive and 0 for negative) of the Nth trial respectively, m_{shift} is the number of reversals in the first N trials, and c is a constant to control the scale of stepsizes. The first two steps of ASA are identical to those of SA, but it decreases the stepsizes just when reversals appear in later stage.

MATLAB simulation results

The computational simulation of psychoacoustical methods is necessary, because it can preliminarily verify the performances of different procedures with different parameters and PFs.

The results of simulation offer us a useful reference but can not be accept as an absolute standard. The reason is that the simulation supposes time invariable PFs and an ideal psychoacoustical model but neglects some subjective factors. However, we all know that all these factors may greatly affect the results of real experiments.

The simulation of this paper includes two parts. The first part is about the comparison among three nonparametric methods under the following same conditions:

Experiment numbers: 10000;

Slope:$0.4 \leq k \leq 1$, step 0.1;

Symmetric point m: 65dB; Starting SPL: 10dB above m but fluctuate 5dB around the value;

Number of trials: 40 for 3d1u and ASA.

For PEST, the trials can not be fixed in advance, so the float range and start stepsize and atop stepsize are adjusted to make the average trials be about 40.

Before simulation, the parameters of the three methods are optimized as shown in Table 1.

Table 1. Optimized Parameters in Different Methods

Method	Parameter	Optimal Value
3d1u	Stepsize	2dB
PEST	w	1
	Starting stepsize	4dB
	Stoping stepsize	2dB
ASA	Constant c	8dB

Because of the particularities of the three psychoacoustical methods, the threshold estimation ways are slightly different from each other too.

For 3d1u, the stepsize is constant. So, the values of SPL are discrete and equal-spaced. If the last trial is accepted as the estimation of threshold, decreasing the stepsize can improve the precision but does not favour to approach the target. On the other hand, large stepsize is harmful to the precision but is helpful to approach the target. Under the condition that the target can be reached and the SPL is fluctuating steadily around the target, the average value of all the reversal points is taken as an estimation of threshold. The reason is that in ideal situation, the probabilities of going up and going down are both 0.5 when the target is about to be reached and all the reversal points should distribute uniformly in both side of the target.

For PEST and ASA, the final trial is accepted as the estimation of the threshold. Because when the number of trials is small, the stepsize of both methods will decrease rapidly and it is unnecessary to make average of all trials.

The second part is about the simulation of constant stimuli method which is not so efficient but much more reliable. Because it is regarded as a relative standard of real threshold, the constant stimuli method should have higher precision than other three adaptive methods. So, the conditions of the simulation are set as follows:

Experiment numbers: 10000.

The parameters of PF and target probability are the same as in the first part.

Range of SPL: 60-74dB, stepsize 2dB.

Number of sample for each SPL: 30.

To estimate k and m of PF of the subjects, the data is fitted by using type 2 nonlinear regression equation with $p_g = 0.5$. The simulation results are shown in Figure 1. From the figure, it is known that performance of ASA is better than those of 3d1u and PEST. Especially, when the slope of PF is larger than 0.8, the standard deviation of ASA method is lower than 1dB, which is the JND of SPL for human's auditory. The performance of 3d1u is the next best one and the performance of PEST is the poorest for its largest error. On the other hand, the result of constant stimuli is even better than ASA, thus it can be considered as a relative standard of threshold measurement.

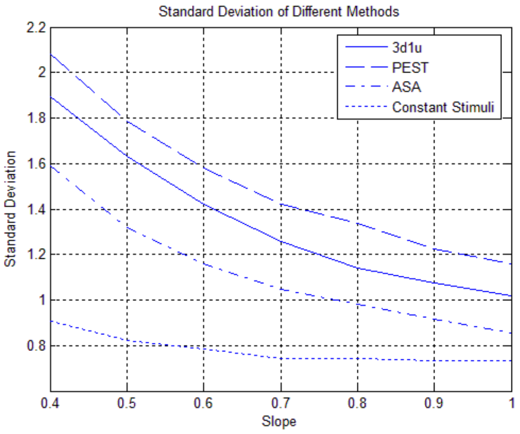

Figure 1. The Standard Deviations of different Methods under different slopes of PF

Experimental verification and result analysis

To verify the performance of the three methods, simultaneous masking experiments are implemented. The masker is narrow-band pink noise of the 9th critical band, its SPL is 70dB; the maskee is a 1017Hz pure tone, just located at the center of the 9th critical band, its starting SPL is 75dB. The test signals are produced by computer, and replayed with GUM96 sound card and a DT770 headphone which has been calibrated.

Three subjects, 2 males and 1 female, aged 24-26, no experience in psychoacoustic experiment, were tested after preliminary auditory training. Similar to the simulation above, the experiment includes constant stimuli method and three adaptive methods.

The data of constant stimuli method must pass Pearson X^2 test[6], the purpose of the test is to assess whether the data come from a PF which is similar to the empirical one. Being lack of fit would lead to a bad estimation of the threshold. However, this assessment is not always reliable. The reasons are, first, the number of sample in the experiment is still small, second, the subjects' responses are not completely independent from each other, and third, the subjects' PFs are time-variable in the most time.

Pearson X^2 test is one of the most widely used assessments. The calculation equation is as follows:

$$\sum_{i=1}^{K} \frac{n_i (y_i - p_i)^2}{p_i (1-p_i)} < X^2_{1-\alpha}(K) \qquad (6)$$

Where K=8 is the number of sample, $n_i = 30$ is multiplicity of each sample, p_i and y_i are the frequency and probability (calculated by the result of fitting) of positive response at the ith sampling respectively. Let α be 0.05, and then it can be calculated out that $X^2_{0.95}(8) = 15.5$.

Table 2. The Fitting Results of Constant Stimuli Method

Statistical Variables \ Subjects	A	B	C
k	0.5	0.9	0.5
m	64.8	69.6	64.6
Pearson X^2	1.9	3.3	3.9
Threshold Estimation	65.4	70.0	65.3

The experiment result of constant stimuli method and its Pearson X^2 test for three subjects are shown in Table 2. From the table, it is known that the experiment results of subjects A, B and C all have passed Pearson X^2 test. Because Pearson X^2 test is just a judgement of the quality of fitting but is powerless to estimate the bias of the data, the result of constant stimuli method should be just regarded as a relative but not an absolute standard.

Following the constant stimuli method test, the tests of the other three nonparametric methods have been done then. Table 3-5 show the test results of different methods for subject A, B and C.

Table 3. The Testing Results of Subject A

Time	Method	Average (dB)	Sample Standard Deviation
Day 1	ASA	64.6	0.7
Day 2	UDTR	65.8	1.1
Day 3	ASA	65.4	1.1
Day 4	UDTR	65.6	0.9
Day 5	PEST	63.5	1.8
		Average Trials: 38	

Table 4. The Testing Results of Subject B

Time	Method	Average (dB)	Sample Standard Deviation
Day 1	ASA	72.1	0.9
Day 2	UDTR	70.3	0.6
Day 3	ASA	70.6	0.7
Day 4	UDTR	69.7	0.7
Day 5	PEST	68.3	0.7
		Average Trials: 27	

Table 5. The Testing Results of Subject C

Time	Method	Average (dB)	Sample Standard Deviation
Day 1	UDTR ASA	65.6	1.3
Day 2	UDTR	65.8	1.3
Day 3	UDTR ASA	64.8	1.4
Day 4	ASA	62.0	2.5
Day 5	ASA UDTR	63.1	2.4
Day 6	PEST	61.5	0.9
		Average Trials: 48	

From these 3 tables, it is known that the results are similar to those of constant stimuli method test. However, compared to the results of simulation, the differences can be observed obviously. Such facts indicate that subjective factors in the experiments can not be ignored. The other reason of the existence of the difference is that the simulation use a time-invariable PF, which means the target position will not change along with time, and assume that the processes of hearing and subjects' judgements are independent from each other. Nevertheless, these assumptions are not completely true in real experiment.

The main reason of the time-variance of PEs is that the subjects' physiological and psychological conditions are different in different time. The other reason is that there is mutual dependence between processes of hearing and judgements. Moreover, the repeat listening of the same signal (the spectrums of masker and maskee forming the signal are constant) makes the subjects are familiar with the signal. As a demonstration, the sample standard deviation of the subject C is relatively large in each day, but the average value of the threshold becomes lower and lower day by day.

From the 3 tables, it can also be confirmed that the largest deviation brought by subjective factors are as large as several dB and the results are not randomly distributed. On the contrary, the simulated standard deviations of different methods, which are shown in Figure1, are just as little as 0.1-0.7 and the thresholds are randomly distributed in a small range. That is to say, the deviation brought by subjective factors almost has covered the theory differences of different methods and standard deviation is not a reliable indicator of the quality of different methods too.

Table 6 shows a comparison among several items. The italics are considered to be the advantages of the methods. From the table, we know that compared with other two methods, ASA posses 4 superior items in the 5 items discussed. If PF is strictly constant, ASA is mathematically rigorous. But variance of PFs and the fluctuation of threshold can easily damage its stability. And the results of subject B and C are just evidences. Thus, the verification of stability is necessary when the listening procedure is finished.

Table 6. The Function Comparison of the Three Nonparametric Methods

UDTR (3d1u)	
Item	Advantange/Disadvantange
Parameters	Rule of going up and down, Stepsize. The former is easy to confirm, but the later is fixed and is lack of flexibility.
Starting SPL	To decide the stepsize and starting SPL. PFs should be known. If starting SPL is far from the threshold, lots of trials will be wasted.
Trial Number Controlling	*Controllable.*
Stability	*Averaging reversals leads to good stability*. But it may not be a true value.
Asymmetr y	Asymmetry and can just reach some discrete positions. But it *can be solved by averaging reversals.*

PEST	
Parameters	Fluctuating range w, start and stop stepsize, which are too complex to choose optimal values.
Starting SPL	The same to UDTR.
Trial Number Controlling	Just the average number of trials can be controlled.
Stability	*Stable*
Asymmetry	The estimated threshold is the SPL of the final trial, so the results distribute asymmetrically`.

ASA	
Parameters	*Only one parameter, constant c which controls the scale of stepsize.*
Starting SPL	*Not important to the precision. And the procedure can still detect the threshold well even the information of PFs is not known enough.*
Trial Number Controlling	*Can be controlled absolutely.*
Stability	When the real threshold is fluctuated, the stability will go bad.
Asymmetry	*Symmetry.*

ASA is in fact an improved version of weighted up-down method[7]. The different ratio of upward step to downward step makes the process converge to different target position. It can be known from equation (6) that the ratio of steps is

$$\frac{S_{\downarrow}}{S_{\uparrow}} = \frac{1-\emptyset}{\emptyset} \tag{7}$$

It is also known that weighted up-down method can choose target probability arbitrarily. Because the probabilities of going up and going down are $1-\emptyset$ and $\emptyset$ respectively at the position near the target, after N trials, the expectation numbers of going up and going down are $N(1-\emptyset)$ and $N\emptyset$. So

$$N(1-\emptyset)S_{\uparrow} - N\emptyset S_{\downarrow} \approx 0 \tag{8}$$

This equation means that the weighted up-down method can go stable near the target. And such way of decreasing the stepsize, which is used in ASA method, is beneficial to convergence. Thus, the stability of an ASA procedure can be verified by calculating the ratio of upward step to downward step in the second half of the trials and comparing this ratio with $\frac{1-\emptyset}{\emptyset}$. We believe such calculation and comparison might cover the flaw of ASA method to a certain extent in the practical application.

Conclusion

The comparison between simulation and psychoacoustical experiments of the three nonparametric methods, 3d1u, PEST and ASA, shows that they’re remarkably different from each other because of the influence of subjective factors. Although the standard deviation of ASA is the smallest in the simulation, the deviation brought by subjective factors make it unreliable to do the comparison of the three methods. In the comparison of several items, the overall performance of ASA is still the best. The stability of ASA could be affected by the fluctuation of the threshold, but this can be improved with some corresponding procedures.

ACKNOWLEDGMENTS: The authors acknowledge the support of the National Nature Science Fund of China Grant No. 10874048

Reference

[1] Bernhard Treutwein, "Adaptive Psychophysical Procedures", *Vision Research*, Vol. 35, No. 17, 2503-2522 (1995)

[2] Walter H. Ehrenstein and Addie Ehrenstein, "Psychophysical Methods", in *Modern techniques in neuroscience research.* Uwe Windhorst and Håkan Johansson (Springer-Verlag Berlin Heidelberg 1999 Print in German) pp. 1211-1242

[3] H. Levitt, "Transformed Up-Down Methods in Psychoacoustics", *J. Acoust. Soc. Am*, Vol. 49, No. 2 (Part 2), 467-477 (1970)

[4] M. M. Taylor and C. Douglas Creelman, "PEST: Efficient Estimates on Probability Functions", *J. Acoust. Soc. Am*, Vol. 41, No. 4, 782-787 (1967)

[5] Harry Kesten, "Accelerated Stochastic Approximation", *Ann. Math. Statist*, Vol. 29, No. 1, 41-59 (1958)

[6] Felix A. Wichmann and N. Jeremy Hill, "The Psychometric Function: Ⅰ. Fitting, Sample, and Goodness of Fit", *Perception & Psychophysics*, Vol. 63, No. 8, 1293-1313 (2001)

[7] Christian Kaernbach, "Simple Adaptive Testing With the Weighted Up-Down Method", *Perception & Psychophysics*, 49 W, 227-229 (1991)