TUTORIAL

The very first step to start psychophysical experiments

Tsuyoshi Kuroda* and Emi Hasuo

*Graduate School of Medical Sciences, Kyushu University,
3-1-1, Maidashi, Higashi-ku, Fukuoka, 812-8582 Japan*

Abstract: This tutorial article aims to guide the reader to learn about three types of psychophysical methods: the method of adjustment, the method of limits and the method of constant stimuli. The article explains how to estimate the point of subjective equality and the just noticeable difference with each method. It also explains about pitfalls that one may encounter when designing experiments. Results could be distorted by some bias effects that are due to technical procedures adopted for experiments. One should know where these effects are likely to occur.

Keywords: Psychophysics, Method of adjustment, Method of limits, Method of constant stimuli, Point of subjective equality, Just noticeable difference

PACS number: 43.66.-x [doi:10.1250/ast.35.1]

1. INTRODUCTION

Let us suppose that you are an undergraduate student in a psychophysics laboratory. Psychophysical methods have been elaborated to assess human ability (or tendency) to perceive some properties of external objects (i.e., *stimuli*), and you are interested in conducting psychophysical studies to reveal mechanisms underlying some perceptual phenomena. To write a good dissertation, you have to start an experiment within a few months after the beginning of the first semester, but you do not know much about experimental designs. There are three possible solutions to learn about psychophysical methods. The first one is to buy or borrow text books explaining about psychophysics or experimental psychology. This is a standard way, but some books spend a lot of pages to explain theoretical and historical backgrounds of psychophysics, which masks the practical information required to prepare experiments right away. The second one is to read research articles and use the methods similar to the ones used in these articles. This is a favorable way for young researchers, but research articles, per se, do not teach several points of which you should be careful when designing experiments. You have to eliminate bias effects that could distort results, but these effects are sometimes due to technical procedures adopted for experiments. You may need some other sources to know where the bias effects are likely to occur. The last one, which seems the solidest way, is to ask guidance from your supervisor, tutor, or skillful friends. Being your good tutor is the main purpose of the present article. We will

explain about the fundamentals of designing psychophysical experiments, which will help beginners with starting experiments. We believe that some of the information, especially about where the bias effects are likely to occur, are helpful for a broad range of readers beyond beginners.

Three types of psychophysical methods are explained in the present article: the method of adjustment, of limits, and of constant stimuli (or the adjustment method, the limit method, and the constant method). Since the understanding of each method would be facilitated by taking specific examples of researches, we propose here taking the experiments of auditory time perception [1–4]. In the studies of time perception, listeners are often asked to compare the relative durations of two time intervals, each of which is delimited by two brief sounds presented successively (Fig. 1). Some physical properties of sounds, such as frequency or spatial location, can affect the perceived duration of the intervals delimited by these sounds [5–9]. For example, an interval between two sounds' onsets is perceived as longer when each sound lasts longer, even if this interval is physically fixed at a constant duration [10–12]¹. Such factors can also affect listeners' ability of discriminating two intervals; listeners

¹Strictly speaking, empty time intervals are classified into two categories: One, called the *inter-onset* interval (or stimulus onset asynchrony), is delimited by the onsets of two stimuli, while the other, called the *inter-stimulus* interval, is delimited by the termination of the preceding stimulus and the onset of the following stimulus. However, the experiments mentioned in the following text can be understood without distinguishing between inter-onset and inter-stimulus intervals, because in these experiments intervals are always delimited by sounds of equal length.

*e-mail: tkuroda@neurophy.med.kyushu-u.ac.jp

a. Method of adjustment

St = Standard interval
Co = Comparison interval

Ascending sequence trial

Descending sequence trial

b. Method of limits

Ascending sequence trial

Descending sequence trial

c. Method of constant stimuli

1st trial

2nd trial

3rd trial

Fig. 1 The scheme of stimulus series in the method of adjustment (a), of limits (b), and of constant stimuli (c). Participants make an adjustment or give a response after listening to each pair of time intervals (i.e., the standard and comparison intervals). For the method of adjustment (a) and of limits (b), each trial has the ascending or descending sequence. In the ascending sequence, the comparison interval is sufficiently shorter than the standard interval at the beginning of the trial. In the descending sequence, the comparison interval is sufficiently longer than the standard interval at the beginning of the trial.

are unable to detect the duration difference between two intervals if this difference is smaller than a certain point, which is modulated by changing some physical properties of the sounds [13–15]. The perceived duration of intervals and the ability (level) of discriminating intervals are often evaluated with the *point of subjective equality* (PSE) and the *just noticeable difference* (JND; or difference limen), respectively. How to estimate PSEs and JNDs with each psychophysical method is explained below.

2. METHOD OF ADJUSTMENT**2.1. Understanding of the Method**

In the method of adjustment, participants are given a device with which they can control the magnitude of a parameter of stimuli (i.e., the duration of intervals here). Figure 2 shows a computer display which was actually

Fig. 2 A computer display that was used by Hasuo, Nakajima and Ueda [16] for an experiment with the method of adjustment. The displayed text was originally in Japanese. Participants clicked on the “play” button to listen to the standard and comparison intervals. They used the sliding bar for rough adjustments of time intervals, and used the “<” and “>” buttons for fine adjustments of intervals. When they were satisfied with the adjustment, they clicked on the “next” button to move on to the next trial.

used by Hasuo, Nakajima and Ueda [16]. Participants dragged a slider with a computer mouse to adjust the duration of intervals; moving the slider to the left and the right made the intervals briefer and longer, respectively. When the mouse button was released, the computer shortened or lengthened the intervals according to the distance between the slider and the midpoint. Then, the slider automatically returned to the midpoint. The buttons labeled “<” and “>” were used for finer adjustments of intervals. This is an example of possible devices for the adjustment method. Indeed, various types of devices have been used in previous studies [12,17,18].

We propose here estimating the PSE and the JND of a 500-ms interval between two stimuli. The series of stimuli is illustrated in Fig. 1(a). Each trial begins by the presentation of two intervals. One interval is called the *standard* interval and the other is called the *comparison* interval. The experiment is usually divided into two sessions (or blocks); in one session, the standard is presented before the comparison (the standard-comparison order), and in the other session, the standard is presented after the comparison (the comparison-standard order). The order of these sessions is counterbalanced, i.e., half participants perform the standard-comparison session and then the comparison-standard session, while the other half perform these sessions in the opposite order.

If the aim of this experiment is to examine the effect of frequency difference between two sounds delimiting each interval (as examined in several studies [5–7]), the frequency of two stimuli of the standard interval is systematically changed. For example, the preceding stim-

ulus (S_1) of the standard is 500 Hz, 1,000 Hz, or 1,500 Hz while the following one (S_2) is always 1,000 Hz, resulting in three conditions (i.e., $S_1/S_2 = 500/1,000$, $1,000/1,000$ and $1,500/1,000$ Hz). We call the conditions where two stimuli of the standard differ in frequency (i.e., 500/1,000 and 1,500/1,000 Hz) the *experimental* conditions and call the condition where these stimuli have no frequency difference (1,000/1,000 Hz) the *control* condition. Two stimuli of the comparison are both fixed at 1,000 Hz (the same frequency as in the control condition).

Participants are asked to adjust the comparison duration and make it identical to the standard duration. They can repeatedly listen to the standard and comparison intervals and make adjustments until they perceive the comparison as equivalent to the standard duration. The standard interval, which cannot be changed by participants, is fixed at 500 ms. The comparison interval, which can be changed by participants, is *sufficiently* shorter (e.g., 300 ms) or longer (e.g., 700 ms) than 500 ms at the beginning of each trial (see Sect. 5.1 for understanding what we mean by “sufficiently”). Participants will lengthen the comparison if it is shorter than 500 ms at the beginning of the trial and will shorten the comparison if it is longer than 500 ms at the beginning of the trial. We thus call the former trial the *ascending-sequence* trial, and call the latter trial the *descending-sequence* trial. The initial duration of the comparison interval (which is presented at the beginning of the trial) is randomly changed within a certain range; for example, the comparison starts from 275 ms in a trial, 331 ms in a trial, 257 ms in a trial, 303 ms in a trial. . . of the ascending sequence. Would this initial duration be always the same value, participants might be able to learn or anticipate how much degree they should change the comparison by to make it identical to the standard duration, which would distort results.

The ascending and descending trials are each conducted many times (but equally often) in order to reduce the error of estimation of PSE and JND. If each trial is conducted ten times, each session consists of 60 trials ($= 3$ conditions $\times 2$ sequences $\times 10$ times). The order of these trials is randomized for each participant to cancel out the effects that might be caused by particular orders of trials (Table 1).

The standard and the comparison intervals are separated by a silence which is long enough to prevent perceptual interactions (e.g., rhythmic groping) between these intervals. How long this silence should be depends on the purpose of experiments. In the present case, we will use the silence of 2 s because several studies [19,20] suggested that successive sounds are not perceptually grouped if they are separated by intervals of 2 s or more. If the intervals are presented after a mouse clicking or key pressing, the sound that could be produced by the clicking or pressing should

Table 1 Examples of trial series and estimated parameters (in ms) in the method of adjustment.

<i>Trials</i>				
Trial No.	Standard-comparison session		Comparison-standard session	
	Condition	Sequence	Condition	Sequence
1st	Condition 2	Ascending	Condition 1	Descending
2nd	Condition 1	Ascending	Condition 2	Ascending
3rd	Condition 3	Descending	Condition 1	Ascending
4th	Condition 3	Ascending	Condition 2	Descending
5th	Condition 2	Descending	Condition 3	Descending
—	—	—	—	—
—	—	—	—	—
<i>Results for a condition</i>				
	Standard-comparison session		Comparison-standard session	
	Ascending	Descending	Ascending	Descending
The final duration of the comparison in each trial	501	414	520	488
	423	572	489	512
	520	468	496	448
	502	432	612	580
	512	418	439	562
	543	478	615	569
	423	478	426	426
	447	390	498	444
	587	436	557	498
	524	511	511	530
Mean	498	460	516	506
σ	50	51	60	52
JND	34	34	41	35
Mean between four types of trials PSE = 495, JND = 36				

Note. JND is given by multiplying the standard deviation (σ) by a z score of 0.75 probability. The standard deviation is calculated from ten PSEs for each sequence for each session.

also be separated from the first interval by a silence long enough.

2.2. PSE and JND

A stimulus parameter that is manipulated by the experimenter is called an *independent variable*, while an index of participant’s response that is recorded in the experiment and might be changed with independent variables is called a *dependent variable*. In the present case, the frequency difference between the two stimuli of the standard interval corresponds to an independent variable, while PSE and JND correspond to dependent variables.

When the trial is finished, the final duration of the comparison is recorded as PSE. As the standard interval is perceived as longer, the comparison will be made longer. Thus, a higher PSE means that the standard interval is perceived as longer.

One may expect that the comparison interval is adjusted at 500 ms ($=$ the standard interval) at the end of

the trial if the two sounds of the standard have exactly the same physical parameters as those of the comparison (i.e., in the control condition). However, PSE is indeed modulated according to two types of errors: random errors and systematic errors. Random errors are the variation of results which is due to chance or individual participants. This could be reduced by increasing the number of participants (samples). Systematic errors are caused by some particular experimental settings. For example, PSE is affected by whether the trial consists of ascending or descending sequence (error of habituation and of expectation [21,22])², and also by whether the standard is presented before or after the comparison (time order error [21–23]). These types of systematic errors could be canceled out by averaging PSEs obtained from these trials, and the resulting PSE is used as a dependent variable for an individual participant and for a condition (Table 1). This PSE nevertheless may not be exactly 500 ms for some unexpected systematic errors. Thus, the control condition should always be included in experimental designs and compared with the experimental conditions; otherwise, it would not be possible to rule out a possibility that the results found in an experimental condition is attributed to some unexpected systematic errors. Systematic errors are also called constant errors and are sometimes expressed by PSE minus the standard duration [21].

JND is rarely estimated with the adjustment method and indeed we could not find any references where JND was estimated in the time-perception field. However, several studies in different fields [21,24,25] suggested that JND is given by multiplying the standard deviation of PSEs within each individual by a z score of probability of 0.75 (about 0.6745). The standard deviation is also used solely as JND [18,21,22,26] or multiplied by different coefficients [17]. JND is estimated separately for the ascending and the descending trials and separately for the standard-comparison and the comparison-standard trials (Table 1). The average of the JNDs obtained from these trials is used as a dependent variable for an individual and for a condition. The variability of PSEs is decreased if participants match the comparison and standard durations

more sensitively. Thus, a lower JND means better discrimination.

3. METHOD OF LIMITS

3.1. Understanding of the Method

Few studies have yet adopted the method of limits in the time-perception field, whereas we see no reason to refuse this method to be used in this field. Indeed, the adaptive method, which is an extended version of the method of limits, has been used in the field [14].

Almost the same procedure as in the adjustment method is used in the limit method. However, participants are asked to judge whether the comparison is “shorter” or “longer” than the standard duration, instead of being asked to adjust the comparison duration. They respond, for example, by clicking on a computer-display pane labeled “shorter” or “longer.” The series of stimuli is illustrated in Fig. 1(b). The standard interval is fixed at a constant value (e.g., 500 ms) while the comparison interval is changed by the experimenter or automatically changed by the computer, as shown in Table 2. At the beginning of the ascending trial, the comparison interval is sufficiently shorter than the standard, and then the comparison interval is lengthened by a constant value (e.g., 20 ms). Participants will respond that the comparison is “shorter” than the standard at the beginning of the trial, and will change the response into “longer” when the comparison exceeds a certain point of duration. The trial is typically finished when participants respond “longer,” but we recommend increasing the comparison duration a few more times to ensure that participants do not return their response to “shorter” as suggested by Baird and Noma [26]. At the beginning of the descending trial, the comparison interval is sufficiently longer than the standard, and then the comparison is repeatedly shortened until participants respond “shorter.” As in the adjustment method, the initial duration of the comparison interval is randomly changed within a certain range (Table 2).

Instead of the two-alternative choice (i.e., “shorter” or “longer”), the three-alternative choice can be used for the limit method, adding a response that participants perceive the comparison as “equal” to the standard duration. Participants will respond “shorter” at the beginning of the ascending trial, “equal” when the comparison exceeds a point of duration, and then “longer” when the comparison exceeds a further higher point. These responses will appear reversely in the descending trial.

3.2. PSE and JND

As in the adjustment method, PSE is estimated for each trial. Dependent variables for each individual and for each condition are given by averaging the results of the ascending and descending trials and of the standard-

²If participants are habituated to the experimental procedure, they may be biased to continue lengthening the comparison interval in the ascending sequence and to continue shortening the comparison interval in the descending sequence. This would result in a higher PSE in the ascending than the descending sequence (error of habituation). Participants may also expect (or hope) after each adjustment that the comparison will be perceived as equal to the standard duration. This may bias participants to perceive the two intervals as equal even when they do not fully lengthen the comparison in the ascending sequence or when they do not fully shorten the comparison in the descending sequence. This would result in a lower PSE in the ascending than the descending sequence (error of expectation).

Table 2 Examples of participants' responses and estimated parameters (in ms) in the ascending (a) and the descending (d) sequence in the method of limits.

<i>Two-alternative choice</i>								
	a	d	d	a	d	a	a	d
Comparison interval (ms)	390			–				
	410	–		–		–	–	
	430	–		–		–	–	
	450	–	–	–		–	–	
	470	–	–	–	–	–	–	–
	490	+	–	+	–	–	+	–
	510	+	–	+	+	–	+	–
	530		+	+	+	+		+
	550		+	+	+	+		+
	570		+	+	+	+		+
	590		+	+	+			+
	610		+		+			
PSE	480	520	480	500	500	520	480	520
<i>Three-alternative choice</i>								
	d	a	a	d	a	d	d	a
Comparison interval (ms)	390		–		–			
	410		–	–	–			–
	430		–	–	–			–
	450		–	–	–	–	–	–
	470	–	–	–	–	–	–	–
	490	–	–	–	–	–	–	–
	510	–	–	–	–	–	–	–
	530	+	+	–	+	+	–	+
	550	+	+	+	+	+	–	+
	570	+		+	+	+	–	
	590	+		+		+	+	
	610	+					+	
Lower limit	500	480	480	480	460	480	500	480
Upper limit	520	520	540	500	500	520	580	520
PSE	510	500	510	490	480	500	540	500
JND	10	20	30	10	20	20	40	20

Note. “+” represents that participants respond “longer”; “=” represents “equal”; and “–” represents “shorter.”

comparison and comparison-standard trials. In the two-alternative choice, PSE is a median of two consecutive comparison durations where the participant's response changes from “shorter” to “longer” (or from “longer” to “shorter”). For example, if the response changes from “shorter” to “longer” when the comparison is increased from 530 to 550 ms (by 20 ms), PSE is estimated as 540 ms.

In the three-alternative choice, a lower limit and an upper limit are estimated. The former is a median of two consecutive comparisons where the response changes from “shorter” to “equal,” and the latter is a median of two consecutive comparisons where the response changes from “equal” to “longer.” PSE is a median of these limits. JND is also estimated with the three-alternative choice and

defined as a half difference between the two limits. If a lower and an upper limit are 480 and 560 ms, respectively, JND is estimated as 40 ms and PSE is estimated as 520 ms. As in the adjustment method, a higher PSE means a longer perceived duration of the standard interval, and a lower JND means that participants can detect a smaller duration difference between intervals.

One problem that could occur in the estimation of PSE and JND is that the participant's response sometimes returns within one trial; for example, the response may change from “shorter” to “equal” when the comparison is increased from 550 to 570 ms but return to “shorter” when the comparison is increased to 590 ms. One may employ two consecutive comparison durations where the response *first* changes (550 and 570 ms in the present case) [24], although there is no established method for solving this problem. In addition, participants may give no “equal” response within a trial with the three-alternative choice. In this case, PSE is a median of two consecutive comparison durations where the participant's response changes between “shorter” and “longer,” while JND is estimated as 0 ms [24]. If many trials have no “equal” responses, the experiment should be reconstructed, for example, by shortening the steps of change of the comparison interval.

4. METHOD OF CONSTANT STIMULI

4.1. Understanding of the Method

The method of constant stimuli generally aims to construct *psychometric functions* from which PSE and JND are estimated. If this method is adopted for estimating the PSE and JND of a 500-ms interval, the standard interval is fixed at 500 ms while the comparison is discretely changed within a range; for example, the comparison is 400, 440, 480, 520, 560, or 600 ms. The standard and the comparison intervals are presented only once in each trial (i.e., no ascending or descending sequence; see Fig. 1(c)). Participants are asked to judge whether the comparison is “shorter” or “longer” than the standard duration. The experiment could be divided into two sessions according to whether the standard is presented before or after the comparison interval, but there is no need to divide the experiment if participants are instructed to judge whether the following interval is shorter or longer than the preceding one in each trial; the same holds for the limit method. In each session, the six comparison intervals (400–600 ms) are presented many times each in a random order³. If there are three conditions (2 experimental + 1 control),

³It is difficult for us to say how many trials should be used for each comparison interval. Indeed, the number of trials seemed inconsistent between the studies where the constant method was used. However, we feel, based on our experience, that at least twenty trials are needed for each comparison interval to make a better fitting of curves to psychometric functions.

and if participants respond 30 times for each comparison, the experiment consists of 1080 trials ($= 2 \text{ sessions} \times 3 \text{ conditions} \times 6 \text{ comparisons} \times 30 \text{ responses}$). Given the number of “1,080,” it would be easy to understand that this method is likely to be time-consuming. However, the results derived from the constant method are not distorted by the effects that would be caused by the ascending and descending sequence in the adjustment and limit methods. Thus, it has been posited that the constant method leads to a better estimation of PSE and JND than the other methods [21].

The three-alternative choice instead of the two-alternative one can be used for the constant method, by adding a response that participants are “unsure” which interval is longer (or that they perceive the comparison as equal to the standard duration)⁴. Note that how many “unsure” responses appear during the experiment depends on individuals; some participants may prefer to choose this response while some participants may not prefer [21]. This individual difference may be reduced by instructing participants not to respond “equal” except when definitely necessary [27,28], but the validity of such an instruction depends on the purpose of experiments.

4.2. PSE and JND with Two-alternative Choice

Psychometric functions are constructed from the probability of responding that the comparison is “longer” than the standard interval [(the number of responses)/(the number of trials)] as a function of the comparison interval. The standard-comparison and comparison-standard trials are usually combined, and thus one function is constructed for each participant and for each condition. As seen in Fig. 3(a), psychometric functions are likely to be S-shaped curves. The cumulative normal (Gaussian) distribution or the logistic curve can be fit to each function in order to estimate PSE and JND; the fitted curves (cumulative normal distribution) are also shown in Fig. 3. A straight line can also be fit if the y axis (“longer” probability) is expressed in a z-score scale⁵. Regardless of which type of curve is adopted, it is important to check how well the curve is fit to each psychometric function. Indeed, an index of the goodness of fit (e.g., the R^2 value⁶) is sometimes reported in the literature.

⁴However, Guilford [21] suggested that “unsure” is more proper to be used than “equal.” Indeed, “unsure” includes the meaning of “equal” because participants cannot decide which interval is longer if they perceive the two intervals as equal duration.

⁵If there is no software to conduct the fitting procedure, the linear interpolation method can be used [21,27]. In this method, two data points between which the probability exceeds just 0.50 are connected by a straight line segment. However, this method wastes the data points that are not used for interpolation. Note that today some free software (e.g., gnuplot, <http://www.gnuplot.info/>) are available to fit various types of curves.

PSE is defined as an x -axis value (comparison interval) at which the curve crosses the probability of 0.50. If the standard interval is perceived as longer, the probability of perceiving the comparison as “longer” than the standard is decreased, resulting in a rightward shift of the psychometric function (Fig. 3(a)). Thus, a higher PSE means a longer perceived duration of the standard interval.

JND is defined as a distance between two x -axis values at which the function crosses 0.50 and 0.75 (or 0.25 and 0.50) [21,22,26]. Technically, this JND is given by multiplying the standard deviation of the fitted cumulative normal distribution or logistic curve by a z score of 0.75 (about 0.6745) [24]; note that the standard deviation here means a parameter deciding the fitted curve, instead of expressing the dispersion of results between participants. Several studies also proposed using the standard deviation solely as JND [8,21,29,30]. A lower JND means a steeper slope of the psychometric function (Fig. 3(a)) and thus means a better discrimination of intervals.

4.3. PSE and JND with Three-alternative Choice

If participants do not frequently respond “unsure,” psychometric functions can simply be constructed from the probability of “longer” responses to which half “unsure” responses are assigned for each data point [21,27,28]. PSE and JND are then estimated in the same way as in the two-alternative choice.

Another procedure is to construct two psychometric functions [21,24]: One function is constructed from *one minus* the “shorter” probability and the other from the “longer” probability as a function of the comparison interval (Note that in this procedure any “equal” responses are not assigned to “shorter” or “longer” responses). These functions are plotted as shown in Fig. 3(b). A curve is fit to each function in the same way as in the two-alternative choice, and an x -axis value at which the curve crosses the probability of 0.50 is estimated from each curve. We call this value the bisection point. PSE is a midpoint between the bisection points of the two curves. JND is a half difference between these points. This procedure is indeed similar to the one used in the three-alternative choice of the limit method, if “two bisection points” are read as “two limits.”

5. A FEW ADDITIONAL COMMENTS RELEVANT TO THE THREE METHODS

5.1. Ensure That the Trial Really Works as the Ascending or the Descending Sequence

The methods of adjustment and limits require the

⁶Note that various types of equation have been proposed for R^2 values and these sometimes provide different results [31]. Some statistics software automatically calculate R^2 values when conducting the curve fitting, whereas it seems important to check what type of equation is adopted for the software.

a. two-alternative choice

b. three-alternative choice

Fig. 3 Psychometric functions in the two-alternative choice (a) and the three-alternative choice (b). The cumulative normal distribution is fit to each function. PSE and JND (in ms) that are estimated from each function are also shown.

ascending- and the descending-sequence trials. The results of these trials are averaged in order to cancel out the bias effects that might be caused by each sequence. This canceling procedure is justified only when participants respond "shorter" and "longer" at the beginning of the ascending and descending trial, respectively. However, participants may sometimes respond, for example, "longer" at the beginning of trials that are to be used as the ascending sequence. In this case, these trials cannot be used to cancel out the bias effects (and may have to be re-conducted or removed from data analysis). Therefore, the

comparison interval must be sufficiently different from the standard interval at the beginning of each trial, so that the participants respond "shorter" and "longer," or lengthen and shorten the comparison interval, at the beginning of ascending and descending sequence, respectively. Given this, it would be favorable to record what participants respond at the beginning of each trial, or how they adjust the comparison interval during each trial.

There is another point to be noticed in the adjustment method. For example, when the comparison interval is adjusted by rolling a volume, the experimenter should not

completely turn up or down the volume at the beginning of each trial. As mentioned earlier, participants may perceive the comparison interval as longer than the standard even at the beginning of the ascending trial. If the volume is turned down completely in this case, participants cannot turn down the volume any more to make the comparison interval shorter. This would then distort the participant's response and make it impossible to check whether the initial duration of the comparison is appropriate to realize the ascending sequence. Thus, the experimenter has to allow participants both to shorten and to lengthen the comparison interval at the beginning of each trial.

5.2. Cancel out Trial-order Effects Correctly

The order of trials is randomized for each participant to cancel out the bias effects that might be caused by particular orders of trials. The experimenter should not use one (the same) order of trials for all participants, even if this order is derived after randomization. If all participants listen to stimuli in the same order, results would be distorted by this order. If the number of trials is large enough and if the order of these trials is randomized for each participant correctly, the same order would rarely be given to more than one participant.

However, if the number of trials is small, the randomization may not work well to cancel out the order effects; it may be likely to give the same order to more than one participant. An alternative method may be to balance or counterbalance the order of trials. If there are two conditions (the conditions 1 and 2), participants can be divided into two groups; one group first performs the trials consisting of the condition 1 and then the trials consisting of the condition 2 while the other group performs these sets of trials in the opposite order. If there are two ascending (A) and two descending (D) trials for each condition, these trials can be arranged like ADDA and DAAD. In each order, the effect that could be caused by the order AD is canceled out by the effect that could be caused by the order DA. The effects that might be caused according to whether AD precedes or follows DA are cancelled out by splitting participants equally between ADDA and DAAD. This is just an example of possible arrangement of trials.

5.3. A Higher PSE *also* Means a Shorter Perceived Duration of the Comparison Interval

We said so far that a higher PSE means a longer perceived duration of the standard interval. This is true if PSE is used to express how long the standard interval is perceived as. However, in some studies with the method of constant stimuli, PSE was used to express how long the comparison interval, instead of the standard one, was perceived as, and indeed a physical parameter of two stimuli of the comparison interval was manipulated. In this

case, a higher PSE means a shorter perceived duration of the comparison interval, because the psychometric function is shifted rightward if the probability of perceiving the comparison as "longer" is decreased. It may be convenient to remember that a rightward shift of psychometric functions (i.e., a higher PSE) means a longer perceived duration of the standard but a shorter perceived duration of the comparison interval.

5.4. Practice Session is Recommended

Participants may make mistakes when responding at the beginning of the experiment. These responses may impair the estimation of means and deviations, but could be reduced by introducing a practice session at the beginning of the experiment. This session will not be included in data analysis but should be conducted with the same procedure as the subsequent (experimental) sessions. The number of presentation of each comparison interval may be reduced in the practice session but this session should include all stimuli (intervals) that are to be used in the experiment, or should include an enough amount of representative stimuli if it is practically difficult to include all stimuli.

In addition, at the beginning of each block (or after a break), a few trials can be added as warm-ups for participants. Stimuli for these trials can be selected randomly from all stimuli that are used in the experiment. Alternatively, the stimuli that are to be presented at the last few trials of the block can be used for warm-up trials.

6. CONCLUDING REMARKS

We believe that you have enough information to start designing experiments if you read this article until this section. Although we took time-perception experiments as examples, you can apply the information provided in this article to conduct the experiments of different fields, if you read "intervals" as "sounds" or "stimuli." Note however that you need to conduct statistical tests after obtaining data. We could not mention this essential part of psychophysical experiments, but you will find a lot of sources to learn about statistical tests in library or bookshops. In addition, we could not spend pages to explain the theoretical backgrounds of each specific procedure. You may wonder, for example, why JND can be estimated from the variability of PSEs within each individual, or how results would differ between the three methods. For understanding these issues, please see references cited in this article [21,22,24,26]. Finally, you will also have to notice the problems that could occur when generating stimulus sounds. These problems are argued in several references [32,33].

Authors' Notes

The authors are the Research Fellow of Japan Society for the

Promotion of Science. This work was made possible by research grants awarded to TK by JSPS KAKENHI Grant Number 25-3275 and to HE by JSPS KAKENHI Grant Number 25-6091. We would like to extend special thanks to Kazuo Ueda, Kaoru Ashihara, and Yoshitaka Nakajima for valuable suggestions to an earlier version of the manuscript.

REFERENCES

- [1] S. Grondin, "From physical time to the first and second moments of psychological time," *Psychol. Bull.*, **127**, 22–44 (2001).
- [2] S. Grondin, "Sensory modalities and temporal processing," in *Time and Mind 2: Information Processing Perspectives*, H. Helfrich, Ed. (Hogrefe & Huber, Cambridge, MA, 2003), pp. 75–92.
- [3] S. Grondin, "Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions," *Atten. Percept. Psychophys.*, **72**, 561–582 (2010).
- [4] G. ten Hoopen, R. Miyauchi and Y. Nakajima, "Time-based illusions in the auditory mode," in *Psychology of Time*, S. Grondin, Ed. (Emerald, Bingley, 2008), pp. 139–188.
- [5] J. Cohen, C. E. M. Hansel and J. D. Sylvester, "Interdependence of temporal and auditory judgments," *Nature*, **174**, 642–644 (1954).
- [6] S. Shigeno, "The auditory tau and kappa effects for speech and nonspeech stimuli," *Percept. Psychophys.*, **40**, 9–19 (1986).
- [7] M. J. Henry and J. D. McAuley, "Evaluation of an imputed pitch velocity model of the auditory kappa effect," *J. Exp. Psychol. Hum. Percept. Perform.*, **35**, 551–564 (2009).
- [8] S. Grondin and M. Plourde, "Discrimination of time intervals presented in sequences: Spatial effects with multiple auditory sources," *Hum. Move. Sci.*, **26**, 702–716 (2007).
- [9] M. Roy, T. Kuroda and S. Grondin, "Effect of space on auditory temporal processing with a single-stimulus method," in *Advances in Sound Localization*, P. Strumillo, Ed. (In Tech., Rijeka, Croatia, 2011), pp. 95–104.
- [10] H. Woodrow, "Behavior with respect to short temporal stimulus forms," *J. Exp. Psychol.*, **11**, 167–193 (1928).
- [11] E. Hasuo, Y. Nakajima and Y. Hirose, "Effects of sound-marker durations on rhythm perception," *Perception*, **40**, 220–242 (2011).
- [12] E. Hasuo, Y. Nakajima, S. Osawa and H. Fujishima, "Effects of temporal shapes of sound markers on the perception of interonset time intervals," *Atten. Percept. Psychophys.*, **74**, 430–445 (2012).
- [13] M. J. Penner, "The effect of marker variability on the discrimination of temporal intervals," *Percept. Psychophys.*, **19**, 466–469 (1976).
- [14] T. H. Rammsayer and D. Leutner, "Temporal discrimination as a function of marker duration," *Percept. Psychophys.*, **58**, 1213–1223 (1996).
- [15] T. Kuroda, E. Hasuo and S. Grondin, "Discrimination of brief gaps marked by two stimuli: Effects of sound length, repetition, and rhythmic grouping," *Perception*, **42**, 82–94 (2013).
- [16] E. Hasuo, Y. Nakajima and K. Ueda, "Does filled duration illusion occur for very short time intervals?" *Acoust. Sci. & Tech.*, **32**, 82–85 (2011).
- [17] B. L. Cardozo, "Adjusting the method of adjustment: SD vs DL," *J. Acoust. Soc. Am.*, **37**, 786–792 (1965).
- [18] C. C. Wier, W. Jesteadt and D. M. Green, "A comparison of method-of-adjustment and forced-choice procedures in frequency discrimination," *Percept. Psychophys.*, **19**, 75–79 (1976).
- [19] P. Fraisse, *Les structures rythmiques* [Rhythmic structures, in French] (Publications Universitaires de Louvain, Louvain, 1956).
- [20] Y. Nakajima, S. Shimojo and Y. Sugita, "On the perception of two successive sound bursts," *Psychol. Res.*, **41**, 335–344 (1980).
- [21] J. P. Guilford, *Psychometric Methods* (McGraw-Hill, New York, 1954).
- [22] G. A. Gescheider, *Psychophysics: The Fundamentals*, 3rd ed. (Lawrence Erlbaum, Mahwah, N.J., 1997).
- [23] H. Eisler, A. D. Eisler and Å. Hellström, "Psychophysical issues in the study of time perception," in *Psychology of Time*, S. Grondin, Ed. (Emerald, Bingley, 2008), pp. 75–109.
- [24] T. Ohyama, "Methods of measurements in sensation and perception I," *Jpn. J. Ergon.*, **4**, 37–47 (1968) (in Japanese).
- [25] H. Ujike and H. Ono, "Depth thresholds of motion parallax as a function of head movement velocity," *Vision Res.*, **41**, 2835–2843 (2001).
- [26] J. C. Baird and E. Noma, *Fundamentals of Scaling and Psychophysics* (Wiley, New York, 1978).
- [27] T. Kuroda, Y. Nakajima and S. Eguchi, "Illusory continuity without sufficient sound energy to fill a temporal gap: Examples of crossing glide tones," *J. Exp. Psychol. Hum. Percept. Perform.*, **38**, 1254–1267 (2012).
- [28] T. Kuroda and S. Grondin, "No time stretching illusion when a tone is followed by a noise," *Atten. Percept. Psychophys.*, **75**, 1811–1816 (2013).
- [29] D. J. Getty, "Discrimination of short temporal intervals: A comparison of two models," *Percept. Psychophys.*, **18**, 1–8 (1975).
- [30] S. Grondin, "Methods for studying psychological time," in *Psychology of Time*, S. Grondin, Ed. (Emerald, Bingley, 2008), pp. 51–74.
- [31] T. O. Kvalseth, "Cautionary note about R^2 ," *Am. Stat.*, **39**, 279–285 (1985).
- [32] T. Kuroda, R. Miyauchi, E. Hasuo and Y. Nakajima, "What one should care when conducting auditory psychological experiments," *J. Des.*, **9**, 63–66 (2008) (in Japanese).
- [33] H. Riquimaroux, "Pitfalls for auditory experiments," *J. Acoust. Soc. Jpn. (J)*, **60**, 614–619 (2004) (in Japanese).