Comparison of Absolute Thresholds Derived from an Adaptive Forced-Choice Procedure and from Reaction Probabilities and Reaction Times in a Simple Reaction Time Paradigm

PETER HEIL,¹ HEINRICH NEUBAUER,¹ ANDREAS TIEFENAU,^{1,2} AND HELLMUT VON SPECHT²

¹*Leibniz Institute for Neurobiology, Brenneckestr. 6, D-39118 Magdeburg, Germany*

²*Institute of Experimental Audiology and Medical Physics, Otto-von-Guericke University, D-39118 Magdeburg, Germany*

Received: 3 March 2006; Accepted: 4 May 2006; Online publication: 6 July 2006

ABSTRACT

An understanding of the auditory system's operation requires knowledge of the mechanisms underlying thresholds. In this work we compare detection thresholds obtained with a three-interval-three-alternative forced-choice paradigm with reaction thresholds extracted from both reaction probabilities (RP) and reaction times (RT) in a simple RT paradigm from the same listeners under otherwise nearly identical experimental conditions. Detection thresholds, RP, and RT to auditory stimuli exhibited substantial variation from session to session. Most of the intersession variation in RP and RT could be accounted for by intersession variation in a listener's absolute sensitivity. The reaction thresholds extracted from RP were very similar, if not identical, to those extracted from RT. On the other hand, reaction thresholds were always higher than detection thresholds. The difference between the two thresholds can be considered as the additional amount of evidence required by each listener to react to a stimulus in an unforced design on top of that necessary for detection in the forced-choice design. This difference is inversely related to the listener's probability of producing false alarms. We found that RT, once corrected for some irreducible minimum RT, reflects the time at which a given stimulus reaches the listener's reaction threshold. This suggests that the relationships between simple RT and loudness

(reported in the literature) are probably caused by a tight relationship between temporal summation at threshold and temporal summation of loudness.

Keywords: absolute threshold, simple reaction time, loudness, temporal integration, psychometric function, Pieron's Law, day-to-day variability

INTRODUCTION

The detection of sounds constitutes a prerequisite for their further analysis by the auditory system. A thorough understanding of the system's operation requires knowledge of the mechanisms underlying detection thresholds, a fact which might explain the continued interest in this topic (e.g., Viemeister and Wakefield 1991; Viemeister et al. 1992; Clock Eddins et al. 1993, 1998; Eddins and Green 1995; Dau et al. 1996; Moore and Oxenham 1998; Dai and Wright 1999; Krishna 2002; Heil and Neubauer 2003, 2004; Neubauer and Heil 2004; Meddis 2006). A wide variety of psychoacoustic techniques are available to measure perceptual thresholds, and the theoretical foundations of many of these techniques have been worked out in considerable detail (see Treutwein 1995 for review). A possible technique, which may provide additional information about the mechanisms underlying thresholds, involves simple reaction time (RT) experiments. Here, a human subject or an animal is asked or trained to promptly perform some specified response, the reaction, upon detection of the reaction stimulus. Therefore, a simple RT experiment can also be viewed as a kind of stimulus detection experiment

Correspondence to: Peter Heil · Leibniz Institute for Neurobiology · Brenneckestr. 6, D-39118 Magdeburg, Germany. Telephone: +49-391-6263332; fax: +49-391-6263328; email: peter.heil@ifn-magdeburg.de

(John 1967; Snodgrass 1975). Commonly, the proportion of (correct) reactions are plotted as a function of the stimulus parameters of interest and estimates of thresholds are derived from these psychometric functions (e.g., Pfingst et al. 1975b; Gerken and Sandlin 1977; Su and Recanzone 2001; Recanzone and Beckerman 2004). However, the relationships between “reaction” thresholds derived in this way from RT experiments and detection thresholds derived with other psychoacoustic techniques have not been comprehensively studied. To our knowledge, only Pfingst et al. (1975b) obtained data directly bearing upon this issue. They compared reaction thresholds obtained from reaction probabilities to tones of different amplitudes and frequencies with detection thresholds measured under similar conditions by using a forced-choice procedure. They found the reaction thresholds to be 5–6 dB higher than the detection thresholds. However, their small sample (a single human subject) and the way they dealt with false alarms warrant a reexamination of the topic. The major goal of the present study was to examine the relationship between detection thresholds on one hand and reaction thresholds on the other.

A second goal of the present study was to examine whether reaction thresholds can not only be obtained from reaction probabilities but also from reaction times, more generally than hitherto appreciated. Simple RT to auditory stimuli is commonly thought to be a measure of the loudness of the reaction stimuli, and is consequently viewed as a useful tool to study loudness perception (e.g., Chocholle 1940; Stebbins 1966; Moody 1973, 1979; Pfingst et al. 1975a; Dooling et al. 1975; Marshall and Brandt 1980; Buus et al. 1982; Humes and Ahlstrom 1984; Seitz and Rakerd 1997; Leibold and Werner 2002; Ariei and Marks 2003; Wagner et al. 2004; Florentine et al. 2005; Little and May 2005; see also Scharf 1978; Luce 1986). In this context, it has been observed that plots of the sound levels of tones of different frequencies that evoked identical and very long mean RTs yielded curves in close proximity to, and of similar shape as, the animal’s audiogram (e.g., Pfingst et al. 1975a; Dooling et al. 1975), suggesting that iso-RT contours for very long mean RTs might reflect the audibility curve. Other investigators, however, have concluded that there is no relationship between threshold and RT, and that detection threshold and RT should not be traded in assessing perception (Abel et al. 1990). In line with the idea that a simple RT experiment constitutes a kind of stimulus detection experiment, we will provide evidence here that simple RTs, and not only long ones, do indeed yield measures of threshold.

Finally, a third goal of our study was to explore and to relate day-to-day variability in detection

thresholds and in performance in RT tasks. To obtain reliable estimates of the location and other parameters of RT distributions, of reaction probabilities, and of detection thresholds, and to study the relationships of these parameters with stimulus parameters of interest, many RT trials or threshold measurements are required. This may be impossible to achieve on a single day, so data are often accumulated over several days spread over some period of time. As pointed out by Luce (1986, p. 132), there is considerable danger that some of the subject’s parameters may vary considerably and that such variation, if ignored or unknown, may lead to wrong conclusions about the mechanisms underlying the RT distributions. Here, we show the existence of pronounced day-to-day variability of detection thresholds, of simple RTs, and of reaction probabilities to auditory stimuli, and exploit this variability to show the tight relationships between all three measures.

METHODS

Listeners

Two types of experiments were conducted with 11 listeners (6 female, 5 male; 20–36 years of age), measurements of detection thresholds, and measurements of simple RTs. In addition, data from another 11 listeners who participated in measurements of detection thresholds only are used here to document day-to-day variation of such thresholds (see below). All listeners were informed about the purpose of the experiments, gave their written consent, and were paid for their participation. Procedures were approved by the ethics committee of the Otto-von-Guericke University Magdeburg (142/03). All listeners had normal or near-normal audiograms, as established with standard audiometry.

Experimental set-up and apparatus

Listeners sat comfortably on a chair within an illuminated and ventilated double-walled soundproof chamber (IAC). They wore circumaural headphones (Sennheiser HDA 200) through which auditory stimuli were presented, and they could view a computer screen that provided some written instructions. For measurements of detection thresholds, the screen also marked the observation intervals and provided feedback. Listeners were equipped with a response box to indicate their choice. For measurements of RTs, listeners had access to a response key, which was a piano key with its lever arms modified such that it moved easily and silently. Its movement was detected by an optocoupler. All other equipment was outside the chamber. Stimuli were generated

with a sampling rate of 96 kHz and a resolution of 24 bit by a PC with a soundcard (Terratec EWX 96/24). The output of the soundcard was attenuated (Tucker Davis Technologies PA5) to obtain the desired stimulus amplitudes and then fed directly to the headphone. The system was calibrated with an artificial ear (Brüel & Kjaer 4153), a 1/2-in. condenser microphone (G.R.A.S. 26 AM) and a piston-phone (Brüel & Kjaer 4228). The soundcard also registered the listeners' RTs, at a rate of 96 kHz equivalent to a resolution of 10.4 μ s.

Procedure

Detection thresholds and RTs were measured alternately within individual sessions. A session lasted approximately 2½–3 h including two breaks of about 15 min each. Typically, detection thresholds for two or three stimuli were obtained first, followed by a period of RT measurements, by measurements of detection thresholds for another two stimuli, by another period of RT measurements, and so on. The first measurement after a break was always one of detection threshold to allow listeners to adapt to soft sounds again. Stimuli were presented monaurally, to the listeners' right ears, and all listeners performed the RT tasks with their right hands, even though one listener (L5) was left-handed.

Measurements of detection thresholds. Detection thresholds were obtained with an adaptive three-interval-three-alternative forced choice (3I-3AFC) procedure. Each trial contained three consecutive 2-s long observation intervals, separated from each other by 0.5 s. The test tone was presented during one of the three intervals with equal *a priori* probability and at random. The listener had to indicate the interval that contained the signal by pressing the corresponding number on the response box. Immediate feedback was provided on the screen and the next trial started 2 s thereafter. After two correct responses in a row, the signal level decreased and after one incorrect response increased ("2-down 1-up" rule). The starting level of the signal was always clearly above the expected threshold. The step size was 5 dB until the fourth reversal, after which it was decreased to 1 dB. After another eight reversals, the track was terminated. The mean of the signal levels (in dB SPL) at the last four reversals was taken as the detection threshold level of one track.

Measurements of simple reaction times. Stimuli for which simple RTs were obtained were presented in blocks by using a "high signal rate vigilance" procedure (Luce 1986; p. 178ff), i.e., no warning signal and foreperiod were given. The listener started a block by pressing a mouse button and was informed, by a message on the computer screen, when the

block was over. A block lasted approximately 10 min and was partitioned into 170 trials where each trial had a duration that varied randomly and uniformly between 2.5 and 4 s. One hundred sixty of these trials were stimulus trials, with stimulus onset coinciding with trial onset, and 10 were catch trials (i.e., trials without an experimenter-controlled stimulus). During a block, 16 different stimuli were presented 10 times each and in pseudorandom order. There were six blocks within a session.

The listeners had been instructed to press the response key as soon as they detected an auditory stimulus, but they received no feedback on their RTs. If listeners pressed the response key in the interval from 0 to 1.9 s with respect to trial onset, an RT was scored. If they pressed the key more than once during that interval, only the timing of the first key press was scored. If no response occurred within that interval, a miss was scored.

Stimuli

All stimuli presented to the listeners, whether used to obtain detection thresholds or RTs, were tones of 3.125 kHz. Tones differed in amplitude, in duration, and in the shape of the temporal envelope. Amplitude transitions followed a sine-squared function.

Detection stimuli. Detection thresholds were obtained to a wide range of stimuli, also for the purpose of another study. In this study, these measurements serve two purposes: (1) to obtain a robust estimate of a listener's sensitivity and its variance at the selected frequency on each experimental day and (2) to derive a so-called "temporal integration function," which relates detection threshold (in dB SPL) to stimulus duration, so that detection and reaction thresholds can be compared for different stimulus durations. Therefore, detection stimuli also varied in duration. These stimuli can be grouped into single-burst, multiple-burst, and onset–offset duration series (for illustration, see Fig. 1 in Gerken et al. 1990 or Fig. 2a in Heil and Neubauer 2003). The shortest stimulus (stimulus 1) was composed only of onset and offset portions of 4.16 ms without plateau. A single-burst duration series was created from stimulus 1 by inserting between these portions a constant-amplitude plateau, to yield stimuli with total durations ranging from 8.32 to 1056.64 ms (step size, factor 2). In addition, a single-burst stimulus with a duration of 536.64 ms was created. This particular stimulus from hereon will be termed the *reference* stimulus. Its envelope and duration are identical to those of stimuli A-1 to A-11 and B-1 to B-6 from the RT experiments (see below). A multiple-burst duration series was created by repeating stimulus 1, after a silent interburst interval of 4.16 ms, to form stimuli

that differed in the number of bursts (viz. 2 to 128; step size, factor 2). An onset–offset duration series was created by varying onset and offset times (from 4.16 to 532.48 ms each; step size, factor 2).

Reaction stimuli. All but two of the reaction stimuli were single-burst stimuli with onset and offset times of 4.16 ms. They differed in level and in duration and were assigned to two sets of stimuli which were employed in two separate RT experiments (A and B). Table 1 provides all details. The purpose of both experiments was to explore the effects of stimulus level and stimulus duration on RT, but with different relative emphasis. In experiment A, where the emphasis was on stimulus level, the *reference* stimulus (536.64 ms duration) was presented at 11 different levels (reaction stimuli A-1 to A-11) covering a 30-dB range in 3-dB steps. The lowest level (A-1) was below some initial estimate of the detection threshold of each listener. Three stimuli (A-12 to A-14) had an amplitude equal to that of A-3, but were of shorter durations (viz., 404.56 ms, 272.48 ms, and 140.40 ms), and thus together with A-3 formed a duration series. The remaining two stimuli consisted of multiple bursts (see Table 1 for details). In experiment B, where the emphasis was on stimulus duration, the *reference* stimulus was presented at six

different levels (reaction stimuli B-1 to B-6) covering a 30-dB range in 6-dB steps. These stimuli were identical to A-1, A-3, A-5, A-7, A-9, and A-11 of experiment A, respectively. Reaction stimuli B-7 to B-9 had amplitudes identical to that of B-1 but were of longer durations (viz., 800.80, 1064.96, and 1593.28 ms), and together with B-1 formed a first duration series of experiment B. Stimuli B-10 to B-16 had amplitudes identical to that of B-2 and together with B-2 formed a second duration series of experiment B. The durations were individually selected for each listener and varied from 16.64 to 1064.96 ms (see Table 1 for details).

Combination of reaction and detection stimuli. RT experiment A was combined with measurements of detection thresholds for six stimuli of the multiple-burst and six of the onset–offset duration series. The same 12 detection stimuli were tested in every session. RT experiment B was combined with stimuli of the multiple-burst and the single-burst duration series. Here, only partially overlapping sets of detection stimuli were tested on different days. On most days, one or several threshold estimates were also obtained for the *reference* stimulus. The sequence in which the different detection stimuli were tested in a given session was random. On different days they were tested in different random orders. All 11 listeners participated in RT experiment A and on average completed 11 sessions each. Four of them also participated in experiment B and completed 12 or 13 sessions each.

Stimulus no.	Exp. A		Exp. B	
	Duration (ms)	Level (dB re stimulus 1)	Duration (ms)	Level (dB re stimulus 1)
1	536.64	0	536.64	0
2	536.64	3	536.64	6
3	536.64	6	536.64	12
4	536.64	9	536.64	18
5	536.64	12	536.64	24
6	536.64	15	536.64	30
7	536.64	18	800.80	0
8	536.64	21	1064.96	0
9	536.64	24	1593.28	0
10	536.64	27	33.15	6
11	536.64	30	41.34	6
12	404.56	6	74.36	6
13	272.48	6	107.38	6
14	140.40	6	206.44	6
15	33.28 (MB)	12	338.52	6
16	133.12 (MB)	12	1064.96	6

All stimuli had onset and offset times of 4.16 ms each and were single-burst stimuli, except A15 and A16, which were multiple-burst (MB) stimuli, consisting of 4 (A-15) or 16 (A-16) bursts. Each burst had a 4.16-ms onset and a 4.16-ms offset without plateau. Bursts were separated by silent intervals of 66.56 ms duration. The total stimulus durations of B10–B16 were individually selected. The ones listed are for listener L1. For L2, B10 was replaced with one of 800.80 ms duration; for L4 and L5, B15 was replaced with one of 16.64 ms duration. All 11 listeners participated in experiment A; 4 of them (L1, L2, L4, L5) also in B. The SPLs of stimulus 1 (in both experiments) were as follows: –3.6 dB (L1), –4.2 dB (L2), 2.5 dB (L3), –0.9 dB (L4), –1.2 dB (L5), –5.4 dB (L9), 2.3 dB (L11), 3.2 dB (L12), 9.7 dB (L14), –3.5 dB (L26), and –6.5 dB (L31).

Data analysis

Calculation of the detection threshold for the *reference* stimulus. As will become evident from the results, it was important to obtain a reliable estimate of a listener's sensitivity at the test frequency (3.125 kHz) on each day on which RTs had been recorded. Because the stimuli for which detection thresholds were obtained were not always the same on different days, a simple average of the threshold SPLs for these stimuli on a given day would not have provided an estimate that would have allowed us to compare a listener's sensitivity on different days. We therefore calculated a listener's detection threshold (in dB SPL) for a single common stimulus, the *reference* stimulus, on a given day from the measured detection thresholds for all of the stimuli tested on that day. The calculation relied on the differences (in dB) between the detection thresholds for these stimuli and that for the *reference* stimulus from days when the latter had been measured. These differences were established with great precision, because (1) the differences in detection threshold levels with differences in stimulus duration, which define the shape of the temporal integration function, were obtained during each session, and (2)

the distance (in dB) between the (average) position of the temporal integration function and the *reference* stimulus was measured several times on different days. Adding these differences, which were specific to each listener, to the measured detection threshold levels for the tested stimuli thus resulted in as many (most of them surrogate) measures of the detection threshold for the *reference* stimulus as there were stimuli for which detection thresholds had actually been measured on a given day. The average of these measures thus provided an estimate of the detection threshold for the *reference* stimulus on a given day that was nearly as reliable as if we had measured the threshold for it as many times as the number of different stimuli tested. The associated confidence interval was somewhat larger because the confidence interval of the differences between the detection threshold levels for the tested stimuli and for the *reference* stimulus used for the calculations had to be taken into account as well.

Analysis of intra- and intersession variance of detection thresholds. As stated above, in combination with RT experiment A, we measured from each listener detection threshold levels for the same 12 stimuli in each session. Also, another 11 listeners, otherwise not included in this study, participated in measurements of detection thresholds but not in the RT experiments. The thresholds of these 22 listeners for this identical set of $n = 12$ stimuli were subjected to an analysis of variance to disentangle the total variance, V_{total} , into an intersession, V_{inter} , and an intrasession component, V_{intra} , where:

$$V_{\text{total}} = V_{\text{inter}} + V_{\text{intra}} \quad (1)$$

and V_{inter} and V_{intra} result from normally distributed deviations. For each of the n stimuli, we calculated the variance of the detection threshold levels across the different sessions, and define the mean of these variances as V_{total} . We also calculated for each session the mean of the detection threshold levels for the n stimuli, and define the variance of these mean thresholds across sessions as V_{mean} . In the absence of intersession variation, V_{total} should be identical to V_{intra} , so that:

$$V_{\text{mean}} = V_{\text{total}}/(n - 1) \quad (2)$$

In the presence of intersession variance, however:

$$V_{\text{mean}} = V_{\text{inter}} + V_{\text{intra}}/(n - 1) \quad (3)$$

because V_{inter} is not reduced by averaging detection threshold levels measured in a given session. Equations (1) and (3) can be used to calculate V_{inter} and V_{intra} from V_{total} and V_{mean} , where

$$V_{\text{inter}} = [(n - 1) \cdot V_{\text{mean}} - V_{\text{total}}]/(n - 2) \quad (4)$$

and

$$V_{\text{intra}} = (n - 1)/(n - 2) \cdot (V_{\text{total}} - V_{\text{mean}}) \quad (5)$$

Considerations of comparable measures for detection and reaction thresholds. To examine the relationship between detection thresholds on one hand and reaction thresholds on the other, with reaction probabilities and reaction times as the key measures, requires measures of detection and of reaction thresholds that are as comparable as possible. Detection threshold levels were obtained with an adaptive 3I-3AFC procedure combined with a “2-down 1-up” rule (see above). This rule converges on the stimulus level, L , yielding a probability of correct responses, $F_{\text{corr}}(L)$, of 0.707 (Levitt 1971; Zwillocki and Relkin 2001), and this level was taken as the detection threshold level, $L_{\text{T(D)}}$, for the stimulus. However, in an n -interval- n -alternative forced-choice procedure the subject may also correctly identify the stimulus interval by guessing. We used the “correction for guessing transformation” (e.g., Treutwein 1995; Gescheider 1997; Klein 2001) to estimate the proportion of trials, $F_{\text{ST}}(L)$, on which a subject has identified the stimulus interval by means other than correct guessing:

$$F_{\text{ST}}(L) = \frac{F_{\text{corr}}(L) - 1/n}{1 - 1/n} \quad (6)$$

The formula is widely used in psychophysics and is valid if a subject is unbiased, i.e., it has no preference for a particular interval, and does not deliberately produce errors. Because, in our case, $n = 3$ and $F_{\text{corr}}(L)$ at the defined threshold is 0.707, it follows that $F_{\text{ST}}(L_{\text{T(D)}}) = 0.561$. This value thus marks the quantile of correct responses to stimuli at the defined threshold in the adaptive forced-choice procedure that cannot be attributed to lucky guessing.

According to this reasoning, we derived two measures from the listener’s performance in the simple RT task, which are based on the same quantile and which thus enable a fair comparison with $L_{\text{T(D)}}$ measured with the forced-choice procedure. The first, and more obvious, measure is the stimulus level that evokes reactions with a probability of 0.561. This measure, $L_{\text{T(RP)}}$, was derived from fits to the functions relating the reaction probability (RP) to stimulus level, as described in detail in Results. The other measure was the 0.561 quantile of the empirical RT distribution, i.e., the time after stimulus onset by which reactions have occurred on a fraction of 0.561 of all trials. As will be elaborated in Results, this quantile may yield an estimate of the initial portion of a stimulus, of given level and envelope, that is necessary and sufficient in order to be detected on a

fraction of 0.561 of all trials. The 0.561 quantile RT for a given stimulus was derived by linear interpolation between the two adjacent RTs of the empirical RT distribution function.

Much more conventional measures of the location of RT distributions are the mean RT and the median RT. Both of these measures are usually computed only from trials on which reactions occurred. If all trials were considered, or if there are no misses, the median RT is only slightly shorter than, and closely related to, the 0.561 quantile RT. However, because our goal requires measures of detection and of reaction thresholds that are as comparable as possible, we used the 0.561 quantile rather than the median. The mean RT may also be a good measure of the location of an RT distribution, although it has the disadvantage of being very sensitive to “outliers,” particularly to late reactions (see Ulrich and Miller 1994). However, a mean cannot be computed from all trials when there are misses. Because we employed several soft reaction stimuli for our purposes, which led to considerable proportions of misses, the mean RT is not considered here. To avoid any possibility of confusion with mean RT or median RT or with the RTs on individual trials, the 0.561 quantile of the RT distribution will, from hereon, be referred to as the RQ.

Correcting reaction times and reaction probabilities for false alarms. Most listeners pressed the response key occasionally, and one frequently, also during catch trials. These false alarms indicate that also some of the reactions of these listeners following stimulus onset must have been false alarms (Luce 1986, p. 56), which consequently bias somewhat the estimates of thresholds derived from RTs and RPs. We corrected this bias by using a race model. Generally, race models assume that reactions in an RT task can be triggered by two or more different and possibly independent processes, and that the reaction time observed on a given trial is the finishing time of the first of these processes (e.g., Ollman and Billington 1972; Kornblum 1973; Burbeck and Luce 1982; Ulrich and Giray 1986; Meyer et al. 1988; Ulrich and Miller 1997; Miller and Ulrich 2003). In our case, we assume reactions to be evoked either by the stimuli of the experimental regime (“stimulus-controlled reactions”) or by something else (false alarms). If they are independent, the race model leads to:

$$F_{ST}(t) = \frac{F(t) - F_{FA}(t)}{1 - F_{FA}(t)} \quad (7)$$

where $F(t)$, $F_{FA}(t)$, and $F_{ST}(t)$ are the cumulative probabilities that a reaction, a false alarm, and a stimulus-controlled reaction have occurred by time t . For our RT data, we have shown elsewhere (Tiefenau et al., unpublished data) that this model provides an

excellent description and that $F_{FA}(t)$ is well described by the exponential distribution:

$$F_{FA}(t) = 1 - e^{-\lambda_{FA}t} \quad (8)$$

We obtained the constant hazard rate, λ_{FA} , by fitting Eq. (8) to the empirical distribution functions of reactions during catch trials, and then applied Eq. (7), with $F_{FA}(t)$ given by Eq. (8), to derive $F_{ST}(t)$. Finally, we derived the 0.561 quantile RT of $F_{ST}(t)$ by interpolation.

When the duration of the analysis window is used as the value for t , Eq. (7) yields a formula to derive the probability of stimulus-controlled reactions, F_{ST} , from the total probability of reactions, F , for a stimulus of level L . It has the same formal structure as Eq. (6). Nevertheless, the probability of lucky guesses in the forced-choice procedure, viz. $1/n$, is not identical to the probability of false alarms, F_{FA} , in the RT task: the former is forced and high, the latter unforced and low. Consequently, the mechanisms generating such responses are likely also different.

RESULTS

From each listener, we obtained detection threshold levels for up to 13 different stimuli as well as reaction probabilities and reaction times to 60 presentations of each of 16 different stimuli in a given session (Table 1) on a given day, and listeners completed up to 13 sessions, over periods of months or more than a year in some cases. In the following, we first analyze the intersession variation in detection thresholds, reaction probabilities, and reaction times, and examine their relationships. We then extract reaction thresholds from reaction probabilities and from reaction times to tones of different levels and durations, compare those thresholds, and finally examine the relationship between reaction and detection thresholds directly.

Intersession variation in detection thresholds, reaction probabilities, and reaction times

Detection thresholds. In this section, we show that there is substantial intersession variation in detection threshold levels in excess of intrasession variation. We can do this, because we measured, for all listeners and in combination with RT experiment A, detection threshold levels for the same 12 stimuli in each session. Figure 1 plots, for two informative listeners (L11 and L5), the difference between the detection threshold level measured in a given session and the mean detection threshold level across all sessions, $\Delta L_{T(D)}$, against session number. The differences shown are averages across two independent groups

FIG. 1. Detection thresholds are subject to intersession variation. The panels plot, for two listeners (a, L11; b, L5), the differences between the detection threshold level measured in a given session and the mean detection threshold levels across all sessions (dashed horizontal

lines), $\Delta L_{T(D)}$. Differences shown are averages across two independent groups of stimuli, viz. stimuli forming the multiple-burst (MB) duration series ($n = 6$) and those forming the onset–offset (OO) duration series ($n = 6$).

of stimuli, viz. the group forming the multiple-burst (MB) duration series ($n = 6$) and that forming the onset–offset (OO) duration series ($n = 6$). It is obvious that the values of $\Delta L_{T(D)}$ for the two groups of stimuli fluctuate from session to session in very similar ways. In both listeners, the $\Delta L_{T(D)}$ values vary from about -3 to $+5$ dB, reflecting lower- and higher-than-average thresholds, respectively. These data strongly point to the existence of substantial intersession variation in detection thresholds.

Figure 2 shows that this was indeed the case in nearly every listener. Figure 2a shows that for 20 of 22 listeners included in this analysis, the observed intersession variance of the average of the detection thresholds for all 12 stimuli, V_{mean} , is larger, and in many cases considerably so, than the one expected in the absence of systematic intersession variation, i.e., for $V_{\text{inter}} = 0$ (indicated by the dashed line). This reveals that $V_{\text{inter}} > 0$ and suggests that V_{mean} is captured by Eq. (3) rather than Eq. (2). Figure 2b provides a scatterplot of V_{inter} versus V_{intra} , calculated with Eqs. (4) and (5), respectively. They sum to yield V_{total} [Eq. (1)]. It is apparent that V_{inter} can be as large as or even exceed V_{intra} . We found no trend for the two variances to covary ($r^2 = 0.013$; $n = 22$).

We did observe a weak tendency for detection thresholds to improve with session number, consistent with findings of other researchers (e.g., Zwislocki et al. 1958; Henry et al. 2001 and references therein). Because listeners differed in their absolute thresholds, we first calculated for each listener and session the detection threshold (in dB SPL), $L_{T(D)}$, for the *reference* stimulus (Methods), and then the difference between the $L_{T(D)}$ of each individual session and the mean $L_{T(D)}$ across all sessions, $\Delta L_{T(D)}$. Across the 11 listeners who performed the RT experiments, $\Delta L_{T(D)}$ decreased by about 0.08 dB per session ($r^2 = 0.05$, $n = 169$; $p < 0.01$). This improvement may be attributable to perceptual learning in some listeners, but other factors are clearly involved as well. Listener L9 quit smoking halfway through experiment A, at which point her sensitivity improved rapidly by about 4 dB.

In any event, these analyses clearly show that there is considerable intersession variation in a listener's sensitivity. Such variation will inevitably broaden psychometric functions when computed from detection responses collected on different days relative to functions obtained from responses on a given day.

Reaction probability and Reaction time. Reaction probabilities and reaction times also exhibited substan-

FIG. 2. Detection thresholds are subject to intersession variation. (a) Observed intersession variance of average thresholds, V_{mean} [Eq. (3)], plotted against that expected if there were no intersession variance, but only intrasession variance [Eq. (2)]. Note that the observed variance considerably exceeds the one predicted with this assumption. (b) Scatterplot of the intra- versus the intersession variance of

detection thresholds [Eqs. (4) and (5)]. Note that the intersession variance can be as large as, or even larger than, the intrasession variance. The figure includes data from 11 additional listeners who participated in measurements of detection thresholds but not in RT experiments, yielding 22 listeners in total.

FIG. 3. Reaction probability (RP) (a–c) and 0.561 quantile of the empirical reaction time distribution (RQ) (b–d) of listener L11 plotted as functions of the level of the *reference* stimulus (536.64 ms duration). Each function was derived from a single session. In the top row (a, b), L is the SPL corresponding to the stimulus maximum (plateau)

amplitude. These are the raw data without any corrections applied (CP0). In the bottom row, the same data are shown after application of correction procedure 2 (CP2) for RP (c) and of 3 (CP3) for RQ (d) and as described in the text. Note the close alignment of the functions from different days.

tial intersession variation. To demonstrate this intersession variation and to analyze it quantitatively, we exploit the fact that the *reference* stimulus (536.64 ms duration) was presented at 11 and 6 different SPLs in RT experiments A and B, respectively. Figure 3a plots, for listener L11 and experiment A, the RP and Figure 3b the corresponding 0.561 quantile of the empirical RT distribution (RQ) as functions of the level (L , in dB SPL) of this stimulus. Corresponding data for listener L5 are shown in Figure 4. Data obtained in a given session are represented by the same symbol and are connected by lines, and in the following are referred to as $RP(L)$ and $RQ(L)$ functions. The panels show that on each day, RP increases and RQ decreases monotonically as SPL increases, but that RP and RQ can vary considerably from session to session at a given SPL. For example, for L11 and at 2 dB SPL, RP was close to 0 on day 8 and close to 1 on day 13. The panels also allow one to appreciate that the *shapes* of the $RP(L)$ functions and of the $RQ(L)$ functions obtained from a given listener on different days are rather similar. Thus, most of the variation appears to be attributable to simple displacements of the individual functions relative to each other along the level axis, and in the case of RQ also along the RQ axis. Consequently, it should be possible to bring these functions into close register by appropriate corrections for such displacements. Indeed, this proves to be the case, as shown next.

We explored three correction procedures (termed CP1, CP2, and CP3) and quantified their effect on a *measure of distance* between the functions of different

days. This measure was derived by first obtaining the differences between all possible pairs of functions at 100 supporting points for each pair. These points were equally spaced along the level axis over the range from the lowest to the highest stimulus levels shared by both functions of a pair. We then averaged the squared differences for each pair of functions and divided the sum of these averages by the number of pairs of functions.

The first procedure (CP1) shifted each $RP(L)$ function and $RQ(L)$ function along the SPL axis by the amount necessary to compensate for the intersession differences in a listener's sensitivity. The $RP(L - \Delta L_{T(D)})$ and $RQ(L - \Delta L_{T(D)})$ functions of different days resulting from this correction procedure (not shown) were indeed in closer register than the corresponding $RP(L)$ and $RQ(L)$ functions (Fig. 5).

The second correction procedure (CP2) shifted each $RP(L)$ and $RQ(L)$ function obtained on a given day along the level axis by the amount required to minimize the *measure of distance* without altering the mean SPL across all days. For RP, the correction produced by this procedure was essentially optimal. As shown in Figures 3c and 4c, the $RP(L - \Delta L_{(RP)})$ functions of different days are in very close register. As expected, the measures of distance between the functions resulting from CP2 were considerably smaller than those between the corresponding $RP(L - \Delta L_{T(D)})$ and $RQ(L - \Delta L_{T(D)})$ functions (Fig. 5).

For RQ, the correction produced by CP2 was not yet optimal. We therefore applied a third correction procedure (CP3) that minimized the *measure of*

FIG. 4. Data as those in Figure 3, but for listener L5. Same format as in Figure 3.

distance by not only allowing each $RQ(L)$ function to be shifted along the level axis, by $\Delta L_{(RQ)}$, but also along the RQ axis, by ΔRQ , again maintaining the mean SPL and the mean RQ across all days. Figures 3d and 4d show that now the $(RQ - \Delta RQ)$ versus $(L - \Delta L_{(RQ)})$ functions of different days resulting from this correction are also in very close register. As expected, the *measures of distance* between these functions are smaller than with CP2 (Fig. 5). Notably, across all subjects, the reduction in the *measure of distance* achieved for RQ with CP3 is roughly similar to that achieved for RP with CP2 (interquartile ranges: 8- to 20-fold for RP; 11- to 23-fold for RQ) (Fig. 5).

For reasons of space, we restrict all following analyses to the displacement estimates derived from those correction procedures that yielded the optimal alignment of $RP(L)$ and $RQ(L)$ functions from different days, viz. CP2 for RP and CP3 for RQ.

Relationships between displacement estimates of $RP(L)$ and $RQ(L)$ functions. The relative displacements of the $RP(L)$ and the $RQ(L)$ functions along the level axis estimated with these procedures viz., $\Delta L_{(RP)}$ and $\Delta L_{(RQ)}$, were closely correlated and essentially identical (Fig. 6a). We performed the following analyses to examine whether the differences between $\Delta L_{(RP)}$ and $\Delta L_{(RQ)}$ might be systematic. First, a straight line was fit to the data. We used a perpendicular regression for this purpose, because both parameters are subject to error and a linear regression, which only minimizes the vertical distance of the data points from the regression line, cannot be used. The perpendicular regression, which minimizes the sum of the squares of the perpendicular distance of each data point from the regression line, is ideally suited here as both axes have the same units. It yielded a regression line with an intercept of 0 and a slope of 1.05 ($r^2 = 0.818$).

FIG. 5. Medians (symbols) and interquartile ranges (error bars), across all listeners and experiments, of the *measures of distance* resulting from the different correction procedures and normalized with respect to those for the original $RP(L)$ and $RQ(L)$ functions (i.e., with respect to CP0).

FIG. 6. (a) Scatterplot of the relative displacements of the RP(L) functions along the level axis estimated with CP2, ΔL_{RP} , against the relative displacements of the RQ(L) functions along the level axis

estimated with CP3, ΔL_{RQ} . Note their close correlation and lack of systematic differences. (b) Scatterplot of ΔRQ and ΔL_{RQ} . Note their independence.

Second, we examined the distribution of the differences between ΔL_{RP} and ΔL_{RQ} . Their mean was 0 dB and the SD was 0.60 dB. A Kolmogoroff-Smirnoff test revealed that the null hypothesis of a normal distribution of the differences around zero could not be rejected. That test yielded a value of $D = 0.0339$, even below the critical value of $D_{crit} = 0.0567$ for the liberal criterion of $\alpha = 0.2$, and with the Lilliefors modification of that test (Sachs 1997, p. 428). These analyses support the notion that there are no systematic differences between ΔL_{RP} and ΔL_{RQ} .

There was no correlation between ΔL_{RQ} and ΔRQ (Fig. 6b). This supports the view that RQ(L) functions obtained from a given listener on different days can be displaced relative to one another along the level *and* along the RQ axis, and that ΔL_{RQ} and ΔRQ are caused by independent factors. Nevertheless, in 3 of the 11 listeners in experiment A, there was some trend for ΔL_{RQ} and ΔRQ to covary (L31: $r^2 = 0.806$, $n = 6$; L14: $r^2 = 0.447$, $n = 13$; L5: $r^2 = 0.267$, $n = 13$). In L31, the measures also changed systematically with session number, suggesting that procedural and/or perceptual learning may underlie the covariation. In the other two listeners, the covariation of ΔL_{RQ} and ΔRQ may be attributable to other factors, such as intersession differences in general fitness.

Relationships between displacements of RP(L) and RQ(L) functions along the level axis and differences in detection thresholds. The relative displacements of RP(L) functions and RQ(L) functions along the level axis, i.e., ΔL_{RP} and ΔL_{RQ} , were closely correlated with, and essentially identical to, the intersession variation in detection thresholds, ΔL_{TD} , of the listeners. Figure 7a provides a scatterplot of ΔL_{RP} against ΔL_{TD} . Here, the perpendicular regression yielded a straight line with an intercept of 0 and a slope of 1.02 ($r^2 = 0.725$). The mean difference between ΔL_{RP} and ΔL_{TD} was 0 dB and the SD was 0.75 dB. A Kolmogoroff-Smirnoff test yielded $D = 0.0382$, which is below D_{crit} . When ΔL_{RQ} , instead of ΔL_{RP} , is compared with ΔL_{TD} (Fig. 7b), the results are similar (intercept of 0 dB,

slope of 0.96; $r^2 = 0.692$; mean difference between ΔL_{RQ} and ΔL_{TD} was 0 dB and SD was 0.78 dB; $D = 0.0578$).

These intimate and likely causal relationships between intersession differences in detection threshold levels and in the positions of RP(L) and RQ(L) functions along the level axis are also reflected in Figure 7c. This panel plots ΔL_{RP} , ΔL_{RQ} , and ΔL_{TD} averaged across listeners and the two experiments as a function of session number. The panel emphasizes the common fluctuations of these measures from session to session. Of course, due to averaging these fluctuations are smaller than those seen in individual listeners (cf. Fig. 1).

In summary, these analyses strongly suggest that the true relative displacements of RP(L) and RQ(L) functions along the level axis are identical, and also identical with the differences in detection thresholds. The small differences between the estimates of these displacements and differences, viz. ΔL_{RP} , ΔL_{RQ} , and ΔL_{TD} , are likely attributable to unsystematic errors.

Displacements of RQ(L) functions along the RQ axis. Figure 7d shows a plot of ΔRQ , the displacements of the RQ(L) functions along the RQ axis from their average, against the session number. We interpret ΔRQ to reflect session-to-session differences in the speed with which the responses were executed, i.e., on the motor side of the reaction process. In three listeners, the largest ΔRQ was obtained for the first session of experiment A. It is conceivable that in these listeners the smaller ΔRQ in subsequent sessions reflect an improvement in the speed of response execution as a result of procedural learning (e.g., Bonnet 2000). However, the listener with the largest drop of ΔRQ from sessions 1 to 2, L3, already had extensive training on this task (thousands of trials). It is thus more likely that the drop is caused by other factors. Overall, there was no indication that ΔRQ would decrease systematically with increasing session number. Rather, the values appeared to fluctuate irregularly around zero. The fluctuations

FIG. 7. Relationships between displacement estimates of RP(L) and RQ(L) functions and detection threshold levels. Scatterplots of (a) $\Delta L_{(RP)}$ and (b) $\Delta L_{(RQ)}$ against the intersession differences in detection threshold level, $\Delta L_{T(D)}$, for the same stimulus. Note the close and presumably causal relationships. (c) Means of $\Delta L_{(RP)}$, $\Delta L_{(RQ)}$, and

$\Delta L_{T(D)}$ across listeners and experiments plotted as a function of session number. Note the close correlation of the three measures. (d) Plot of ΔRQ (small crosses, dashed lines) and their mean across listeners and experiments (larger filled circles). ΔRQ varies little and unsystematically from session to session.

were generally small; when lumped together across listeners and experiments, the SD of ΔRQ was 14 ms (and the mean 0, of course). A Kolmogoroff–Smirnov test revealed that the null hypothesis of a normal distribution of ΔRQ around zero could not be rejected ($D = 0.0482$, below $D_{crit} = 0.0567$).

Analyses of possible effects of False Alarms. The analyses reported so far were based on RP and RQ not corrected for false alarms. It is conceivable that intersession differences in false alarms might have biased the displacement estimates. However, this was not the case. First, we found that for all listeners but one the observed numbers of false alarms in different sessions of an experiment were compatible with binomial distributions of constant probability (binomial statistics were used because the number of false alarms on a given catch trial could either be 1 or 0). The only exception was listener L1. She had completed 12 sessions of experiment A and 12 of B before completing another four sessions of experiment A, about a year after the initial ones. In these later sessions, her probability for producing false alarms had dropped more than 10-fold from the previous value. Therefore, we treated the last four sessions as a separate (repeat) experiment. Second, when we corrected RP and RQ for false alarms [with Eqs. (7) and (8)] and estimated the displacements of the resulting $RP_{(-FA)}(L)$ and $RQ_{(-FA)}(L)$ functions as described above, we found them to be very similar to those of the corresponding $RP(L)$ and $RQ(L)$ functions. For RP, the mean difference in displace-

ment estimates along the level axis was 0 dB and the SD 0.24 dB. For RQ, the mean difference was 0 dB and the SD 0.36 dB. The mean difference in displacements along the RQ axis was 0 ms and the SD 3 ms.

Extraction of reaction thresholds from reaction probabilities

In this section, we use RP to extract reaction thresholds for stimuli of different durations. We first focus on the RP to the *reference* stimuli (536.64 ms duration) that formed the level series in each experiment (A-1 to A-11; B-1 to B-6), because these data were most extensive and generally formed complete psychometric functions (e.g., Figs. 3c and 4c). For each listener and experiment, we used all of the $RP(L - \Delta L_{(RP)})$ functions to extract a single reaction threshold level for the *reference* stimulus. Because these functions from different days are in close register (see Figs. 3c and 4c and text above), they are treated from hereon as a single psychometric function. An additional example (L4) is shown in Figure 8 (filled circles). Each such data set was fitted with a normal distribution function with μ (in dB SPL) and σ (in dB) as free parameters. For simplicity, we refrain here from including the generally low probability of lapses (see Treutwein and Strasburger 1999) as another free parameter. The normal distribution function provided excellent (least-squares) fits to the data (continuous lines through filled

FIG. 8. Psychometric functions and reaction thresholds derived from response probabilities. (a) RP before (RP; filled circles) and after correction of false alarms ($RP_{(-FA)}$; open circles) of listener L4 to *reference* stimuli A1–A11 plotted as a function of $(L - \Delta L_{(RP)})$. Reaction probabilities during catch trials, F_{FA} , are plotted at -14 dB SPL. The inset shows a blowup of the bottom left-hand section of the functions. The continuous and dashed lines show the fits of a normal distribution function to RP and $RP_{(-FA)}$, respectively. From these fits, reaction threshold levels, $L_{T(RP)}$ and $L_{T(RP; -FA)}$, respectively, were derived (dotted lines in inset).

circles in Fig. 8). From the fits, we derived the reaction threshold, $L_{T(RP)}$, defined as the value of $(L - \Delta L_{(RP)})$ (in dB SPL) at which RP reached 0.561. For reasons of space, we restrict the following analyses to RPs corrected for false alarms. As shown for listener L4 in Figure 8, the probabilities corrected for false alarms, $RP_{(-FA)}$, are slightly lower than their uncorrected counterparts; the data points also lie at slightly different values along the abscissa, because the values of $\Delta L_{(RP)}$ obtained from $RP_{(-FA)}(L)$ functions could differ from those derived from $RP(L)$ functions if $F_{FA} > 0$ (see above). The psychometric functions fitted to $RP_{(-FA)}$ were also somewhat steeper (σ was smaller) than those fitted to RP. And finally, the reaction threshold level derived from the corrected probabilities, $L_{T(RP; -FA)}$, was always higher than the one derived from the uncorrected probabilities, $L_{T(RP)}$ (inset in Fig. 8). The difference between the two measures increased as F_{FA} increased, to a maximum of 2.2 dB for an F_{FA} of 0.43 obtained from L5 in experiment B.

In each experiment, we also presented stimuli with durations other than 536.64 ms, although in general at one or two SPLs only (cf. Table 1). Nevertheless, as a result of the correction for the intersession differences in sensitivity (described above), we obtained values of $RP_{(-FA)}$ at several values of $(L - \Delta L_{(RP)})$. Figure 9a shows such data for listener L11 and for stimuli A-12 to A-14, in addition to that for stimuli A-1 to A-11. The data for the shorter-duration stimuli were of course less extensive and covered a psychometric function only incompletely. Nevertheless, the

evoked $RP_{(-FA)}$ were generally of sufficient spread to allow a normal distribution function to be fitted to them. Only the $RP_{(-FA)}$ to the multiple-burst stimuli (A-15 and A-16) were generally so low that we could not fit them reliably. Therefore, they were discarded from further analyses and are not shown in Figure 9a. We tested two fitting variants, one in which μ and σ were free parameters, and another in which only μ was free whereas σ was fixed to the value obtained from the fit to the extensive data for the *reference* stimuli. The reaction thresholds obtained from the two fitting variants were virtually identical in most cases, but in some cases, we had little confidence in the fits when both parameters were free. Therefore, we report only the results from the variant with σ fixed.

Figure 9b plots the reaction thresholds, $L_{T(RP; -FA)}$, as a function of stimulus duration for all listeners and both experiments. In each case, $L_{T(RP; -FA)}$ decreases systematically as stimulus duration increases.

Extraction of reaction thresholds from reaction times

Here, we explore the possibility that a reaction threshold can also be obtained from RT and not only from RP. The idea stems from the fact (shown above) that (1) RTs, just as RPs, are exquisitely sensitive to the variation in a listener's detection threshold, and (2) reaction thresholds are a function of stimulus amplitude and duration. One component of RT, a minimum reaction time, needs to be considered before we can embark on the extraction of a reaction threshold from RT.

Minimum reaction time. RT is generally thought to include a minimum time needed to execute the proper response once the stimulus has been detected. This time, RT_{min} , has been termed the “irreducible minimum” (Woodworth and Schlosberg 1954), “residual latency” (Luce and Green 1972), or “nonsensory factor” (Scharf 1978). Thus, in order to obtain the stimulus-dependent component of RT, RT_{min} should be subtracted from the RT on every trial. However, RT_{min} (or its distribution) is unknown and cannot be directly measured. A common way to estimate RT_{min} from RT data is to fit Piéron's law (Piéron 1920) to the functions relating RT to stimulus amplitude, P , or intensity:

$$RT - RT_{min} = \beta \cdot (P)^{-\alpha} \quad (9)$$

Piéron's law is an empirical law whose physiological or psychological foundations are as yet unclear. But, to a first approximation, it provides a reasonable fit to the dependence of RT on stimulus intensity in several sensory systems and in general a better fit than the alternative functions that have been sug-

FIG. 9. Reaction thresholds for stimuli of different durations. (a) Reaction probabilities after correction for false alarms ($RP_{(-FA)}$) of listener L11 to stimuli of different duration (see key) plotted as a function of $(L - \Delta L_{(RP)})$. (b) Reaction thresholds, $L_{T(RP; -FA)}$ (specified as the dB SPL corresponding to the stimulus maximum amplitude), plotted as a function of stimulus duration for all listeners and experiments.

gested (see Pins and Bonnet 1996, 2000; Bonnet et al. 1999; Bonnet 2000; McKeefry et al. 2003; Stafford and Gurney 2004).

Because our measure of location of the RT distribution was RQ (Methods), fits of RQ data with Eq. (9) will return values for RQ_{min} , an estimate of the minimum time after onset of a stimulus by which reactions can have occurred on a fraction of 0.561 of all trials. In analogy to the way in which we derived thresholds from RPs, we used the RQs corrected for false alarms and for intersession differences, i.e., $(RQ_{(-FA)} - \Delta RQ)$, as the measure of RT, and the SPL corrected for intersession displacements along the level axis, i.e., $(L - \Delta L_{(RQ)})$, as the measure of level. The latter values were converted into pressure units (Pascal) before fitting Eq. (9). We restricted the fits to data from the *reference* stimuli that formed the level series in each experiment. To determine a single value of RQ_{min} for each listener and experiment, we fitted all of the $(RQ_{(-FA)} - \Delta RQ)$ versus $(L - \Delta L_{(RQ)})$ functions in one go, because such functions

from different days are in close register (see Figs. 3d and 4d and text above).

The values of RQ_{min} that were obtained from such fits ranged from about 240 to 410 ms in different listeners. This range was large relative to the within-listener variation of RQ_{min} between experiments (Figs. 10 and 11d). The latter did not exceed 15 ms and thus was similar to the intersession variation of RQ, viz. ΔRQ (cf. Fig. 7d).

The extraction of threshold from reactions times and the measure of threshold level. The left side of Eq. (9), in our case $(RQ_{(-FA)} - \Delta RQ - RQ_{min})$, can be viewed as the duration of an initial portion of a stimulus, with a given amplitude and envelope, that would have sufficed to evoke reactions on a fraction of 0.561 of all trials. That initial portion of the stimulus can thus be considered as a threshold quantity. The stimulus rise time can cover a considerable fraction, or even all, of that initial portion when, with increasing stimulus level, $(RQ_{(-FA)} - \Delta RQ - RQ_{min})$ approaches or even falls below the rise time. Consequently, the maximum stimulus amplitude, the one prevalent during the plateau, may not be the best measure from which to derive a specification of threshold level. We have previously shown (Heil and Neubauer 2001, 2003; Neubauer and Heil 2004) that the mean amplitude during the interval from stimulus onset to the end of the initial portion that may suffice to evoke the response provides a better measure for the type of stimuli used here. Therefore, we calculated the SPL based on the mean amplitude during each initial portion (specified as SPL_m). This measure of threshold will be referred to as $L_{T(RQ; -FA)}$.

Figure 10 plots these thresholds as a function of $(RQ_{(-FA)} - \Delta RQ - RQ_{min})$ (open circles). Eight representative examples from six listeners are shown, with data from both experiments shown for two listeners in Figures 10e–h. In each case, $L_{T(RQ; -FA)}$ decreases with increasing $(RQ_{(-FA)} - \Delta RQ - RQ_{min})$ over the range shown. To emphasize the effects of the subtraction of RQ_{min} , the $(RQ_{(-FA)} - \Delta RQ)$ versus stimulus duration functions are also shown in each panel (crosses). The subtraction of RQ_{min} causes the data points to move to the left and slightly downward, as indicated for a single pair of corresponding crosses and open circles in Figure 10a (pointer and line). The distance that data points have moved to the left in the plot increases as $(RQ_{(-FA)} - \Delta RQ)$ decreases, because the abscissa is logarithmically scaled.

We also obtained estimates of $L_{T(RQ; -FA)}$ from the stimuli with durations other than 536.64 ms, by subtracting the same RQ_{min} as that used for the *reference* stimuli. The resulting threshold estimates are also plotted in Figure 10 (see key) and are virtually identical with those obtained from the *reference* stimuli (open circles).

FIG. 10. Comparison of reaction thresholds obtained from reaction times with those obtained from reaction probabilities and with detection thresholds. Each panel (a–h) shows data from a different listener and experiment (A or B). The abscissas represent stimulus duration or time-to-threshold given by $(RQ_{(-FA)} - \Delta RQ - RQ_{min})$. The ordinates represent level corresponding to the mean stimulus amplitude (referred to as dB SPL_m), derived either from the entire stimulus duration or from the time-to-threshold. Filled squares represent reaction thresholds obtained from reaction probabilities, $L_{T(RP, -FA)}$; open squares detection threshold obtained from 3I-3AFC experiments, $L_{T(D)}$. The crosses represent the functions relating SPL_m of the

stimuli of the level series (536.64 ms duration) to $(RQ_{(-FA)} - \Delta RQ)$, i.e., RQ_{min} is set to zero. All other symbols represent reaction thresholds estimated from $RQ_{(-FA)}$, $L_{T(RQ, -FA)}$, to stimuli of different durations after subtraction of an $RQ_{min} > 0$. The value used for each data set is specified in each panel and was derived by a fit of Piéron's law to data from the reference stimuli (see text for further explanations). The line in (a) marked by the "Ed Rubel" pointer connects a corresponding pair of crosses and open circles to visualize the approximate trajectory of the cross as RQ_{min} increases from 0 to the value used. Note the close match of reaction thresholds obtained from reaction times and reaction probabilities and their distance to detection thresholds.

Comparison of reaction thresholds obtained from RQ with those obtained from RP

The decrease in $L_{T(RQ, -FA)}$ with increasing $(RQ_{(-FA)} - \Delta RQ - RQ_{min})$ is reminiscent of that of $L_{T(RP, -FA)}$ with stimulus duration (cf. Fig. 9b). To enable a more

direct comparison, we also calculated $L_{T(RP, -FA)}$ in units of dB SPL_m and plotted the data in Figure 10 (filled squares). The two threshold estimates are essentially identical for listeners L2 (Fig. 10a), L9 (Fig. 10b), and L11 (Fig. 10c). For the other listeners, slight discrepancies between the two threshold esti-

FIG. 11. Comparison of reaction thresholds with detection thresholds. (a) Plots of reaction thresholds obtained from reaction probabilities, $L_{T(RP;-FA)}$, and of detection thresholds, $L_{T(D)}$, as a function of stimulus duration for three listeners. SPL refers to the maximum stimulus amplitude. (b) Plot of the difference between reaction and detection threshold, $(L_{T(RP;-FA)} - L_{T(D)})$, for the reference stimulus (536.64 ms duration) as a function of the probability of false alarms in the simple RT experiment. Data from the four listeners who contributed more than one data point are shown with different

symbols and are connected by dashed lines. Note the trade-off between $(L_{T(RP;-FA)} - L_{T(D)})$ and accuracy. (c) Plot of $(L_{T(RP;-FA)} - L_{T(D)})$ relative to that for the reference stimulus (dashed lines) as a function of stimulus duration. The thick line represents a running average over 5 points. Note the slight increase as stimulus duration decreases. (d) Plot of RQ_{min} , derived from fits of Pieron's law, as a function of the probability of false alarms in the simple RT experiment. Other conventions as in (b). Note the trade-off between RQ_{min} and accuracy across listeners but the lack of such a trade-off within listeners.

mates remain (Fig. 10d–h). The discrepancies in these cases are still negligible when stimulus duration and $(RQ_{(-FA)} - \Delta RQ - RQ_{min})$ are large, but can increase to nearly 4 dB at the shortest duration where comparisons are possible (Fig. 10f, h). In each of those cases, $L_{T(RQ;-FA)}$ is larger than $L_{T(RP;-FA)}$, and this difference is reproducible (cf. Fig. 10e with f and g with h). It is conceivable that the small discrepancies between the two threshold estimates ensue because the subtraction of a constant RQ_{min} to estimate $L_{T(RQ;-FA)}$ is too simple a model, as RQ_{min} is likely to vary from trial to trial. The discrepancies also give hint to limitations in Pieron's law. If that law provided an optimal fit, then the open symbols in Figure 10 should fall on a straight line with slope α (and if the ordinate represented the stimulus maximum amplitude in dB). However, the data points often showed a slight upward curvature before curving downward for very short values of $(RQ_{(-FA)} - \Delta RQ - RQ_{min})$ (not shown in Fig. 10). And the shortest RQs measured were always shorter than the RQ_{min} estimated from a fit of Pieron's law, and were thus unexplained by that law.

Clearly, this issue needs further attention in future studies.

In summary, by and large, very similar estimates of reaction thresholds are obtained from reaction probabilities and reaction times.

Comparison of detection and reaction thresholds

In this final section, we compare the reaction and the detection thresholds directly. We restrict this analysis to reaction thresholds obtained from reaction probabilities, $L_{T(RP;-FA)}$, because their determinations do not require to fit RQ_{min} and so are not subject to the limitations associated with this parameter. The open squares in each panel of Figure 10 show the detection thresholds, $L_{T(D)}$ (also in units of dB SPL_m), for single-burst stimuli of different durations obtained from each of the six listeners, allowing an easy comparison with the reaction thresholds. In each case, the two threshold functions are roughly similar and parallel, but the function for the reaction threshold always lies above that for the detection

threshold obtained with the adaptive 3I-3AFC procedure. Figure 11a plots data from another three listeners.

We next calculated the difference (in dB) between the two threshold estimates, first at the stimulus duration of 536.64 ms, where the determination of $L_{T(RP; -FA)}$ was most reliable. Figure 11b plots this difference, $(L_{T(RP; -FA)} - L_{T(D)})$, as a function of F_{FA} , the probability of a listener to produce a false alarm during a catch trial in the RT task. Two observations on those data are remarkable. First, $(L_{T(RP; -FA)} - L_{T(D)})$ is always positive. This means that a higher SPL is required to yield a probability of 0.561 to react to a stimulus than to perform at that level in the forced choice procedure. Second, $(L_{T(RP; -FA)} - L_{T(D)})$ decreases as F_{FA} increases. This covariation is also seen in three (L1, L2, L5) of the four individual listeners who took part in both experiments. F_{FA} of L4 did not change much between experiments. In Figure 11b, data from each of those listeners are represented by the same symbol and connected by dashed lines. The F_{FA} of L1 in the four repeat sessions of experiment A (see above) had dropped more than 10-fold from its previous value. This drop was paralleled by an increase in $(L_{T(RP; -FA)} - L_{T(D)})$ from about 1.4 to 3.6 dB (left and right filled triangles in Fig. 11b).

Had we used $L_{T(RP)}$, i.e., the reaction threshold estimated from reaction probabilities uncorrected for false alarms, instead of $L_{T(RP; -FA)}$, we would have obtained slightly smaller threshold differences and a steeper decline of those differences with increasing F_{FA} than those shown in Figure 11b. In addition, the difference between this reaction and the detection threshold would have been negative in the case of the largest F_{FA} observed, a rather implausible difference.

Finally, we examined whether $(L_{T(RP; -FA)} - L_{T(D)})$ varied with stimulus duration. Although the functions relating $L_{T(RP; -FA)}$ and $L_{T(D)}$ to stimulus duration appear roughly parallel (Figs. 10 and 11a), they tended to diverge slightly as stimulus duration decreased. To illustrate this point more clearly, we calculated the amount by which $(L_{T(RP; -FA)} - L_{T(D)})$ for a stimulus of a given duration differs from that for the *reference* stimulus of 536.64 ms duration. Figure 11c plots this difference as a function of stimulus duration. Although the data are noisy, they indicate that over the range of durations studied $(L_{T(RP; -FA)} - L_{T(D)})$ increases ($r^2 = 0.216$; $n = 78$; $p < 0.001$) by about 1 dB with a 10-fold decrease in stimulus duration.

DISCUSSION

We have shown here that auditory thresholds (for tones in quiet) can be extracted from the reaction

probabilities and reaction times of listeners in a simple RT paradigm, and that such reaction thresholds are somewhat higher than the detection thresholds measured from the same listeners with an adaptive forced-choice procedure under otherwise nearly identical conditions.

Intersession variability

In a first step, we showed that RP and RT, even to clearly audible stimuli, are intimately linked to detection thresholds by exploiting the intersession variation in RP, RQ, and detection thresholds (Figs. 1, 2, 3, 4, 5 and 6), and revealing their strong covariance (Fig. 7). To demonstrate this covariance, it was essential to probe the auditory system with a sufficient percentage of near-threshold stimuli to cover the steep portions of the psychometric, $RP(L)$, functions and of the $RQ(L)$ functions. The intersession variations and their magnitudes of RP, RQ, and detection thresholds are interesting in their own right, and raise questions regarding the factors that cause such variation, particularly that of detection thresholds. Variations in absolute thresholds have been previously observed in studies concerned with test-retest reliability of pure-tone thresholds (see, e.g., Henry et al. 2001 and references therein). It is unlikely that differences in headphone placements on different days are a major cause of this variation. First, the output of the special circumaural earphone used is largely independent of placement, as long as it is placed correctly (own measurements with an artificial ear yielded an estimate of variability of about 0.1 dB). Second, potential placement effects would tend to average out because listeners put the headphones on and took them off at least three times in the course of a session. In any event, the intersession variations also make clear that the common practice of simply pooling RT and RP data collected on different days without properly accounting for intersession variation in a listener's sensitivity will cause psychometric and RT versus "intensity" functions to appear flatter than they would be in the absence of such variation. Thus, if their exact shapes are of interest, e.g., for modeling, such pooling should be avoided.

Extraction of reaction thresholds from RP and RQ

A major purpose of our study was to compare reaction and detection thresholds. This comparison required the use of comparable measures. We achieved this by determining the level of a stimulus, of given envelope and duration, which led to an RP of 0.561, the same quantile upon which the adaptive forced choice procedure and stepping rule used by us converges after correcting for lucky guesses. When

specified in dB SPL, both thresholds decrease in an orderly fashion with increasing stimulus duration (Figs. 9, 10, 11), in line with the literature on temporal integration (see references in Heil and Neubauer 2003). To extract a threshold from RT, we determined RQ, the 0.561 quantile of the empirical RT distribution for every stimulus. We then followed the reasoning and the approach developed to extract the thresholds of single (auditory) neurons from their first-spike latencies (Heil and Neubauer 2001, 2003; Heil 2004). The approach is based on the rationale that the response (the first spike in the case of a neuron; the reaction in the case of a listener in an RT experiment) following the onset of a stimulus is evoked when the stimulus reaches the threshold, of the neuron or the listener, respectively, but occurs with some delay thereafter. In the case of a neuron, the delay is mainly a transmission delay that can be considered constant for a given neuron and for a given stimulus frequency (see Heil and Neubauer 2001, 2003; Heil 2004 for discussion). In the case of a listener in an RT experiment, the delay, $RQ_{\min}$, comprises transmission delays in both the sensory and the motor pathways. In addition, the required motor response is under conscious control and, thus, the time of its initiation and the speed with which it is executed will be subject to more variability than a passive transmission delay in a sensory system of an anesthetized animal. Thus, the assumption of a constant $RQ_{\min}$ for a given listener during a given session, as done here, can only be a first approach. Nevertheless, even this simple approach showed that the thresholds derived from RQ are very similar to those derived from RP (Fig. 10). A better model may reveal that they are identical.

Comparison of detection and reaction thresholds and the amount of evidence

Our analyses have shown that for each listener, the reaction threshold derived from reaction probabilities and after correction for false alarms, $L_{T(RP; -FA)}$, is higher than the detection threshold, $L_{T(D)}$, obtained from the 3I-3AFC procedure (Figs. 10 and 11a, b, c). This would also be true had we used the estimates of reaction threshold derived from $RQ_{(-FA)}$ instead of $RP_{(-FA)}$, because these estimates were essentially identical (Fig. 10). However, when we derived the reaction threshold from RPs uncorrected for false alarms, we obtained the reverse order in the listener with the highest probability of producing false alarms. This emphasizes the need to correct RPs (and RTs) for false alarms, especially when the probability for false alarms is high, before meaningful conclusions can be drawn.

Already early models of simple RT include a decision stage (e.g., McGill 1961; Grice 1968; Kohfeld

1969; Murray 1970; Luce and Green 1972; Sanford 1972). As listeners may have some resistance to react, they may need to accumulate more evidence to reach the decision criterion to press the key in an unforced design than necessary for detection in a forced-choice design. Therefore, reaction thresholds can be expected to be higher than, or possibly equal to, detection thresholds. The positive difference between the reaction and the detection threshold, $(L_{T(RP; -FA)} - L_{T(D)}) > 0$, can be viewed as the additional amount of evidence that a listener requires on top of that necessary for stimulus detection in the forced-choice situation before he or she will react, by pressing the response key, in the RT paradigm. This conclusion was also drawn by Pfingst et al. (1975b). This amount is expressible in decibels and when done appears to increase slightly as stimulus duration decreases (Fig. 11c). The finding that $(L_{T(RP; -FA)} - L_{T(D)})$ decreases from about 5 to 1 dB (for the *reference* stimuli of 536.64 ms duration) as the listeners' probabilities of false alarms, F_{FA} , increase from near zero to about 0.4 (Fig. 11b) indicates an inverse relationship (across listeners) between the required amount of evidence and response accuracy, because false alarms can be considered inaccurate reactions. Because this relationship is derived from RP corrected for the effects of false alarms, it indicates that a large F_{FA} is the consequence, and not the origin, of a small amount of evidence, and *vice versa*. Our data are consistent with the notion that such a relationship might also be seen in individual listeners: when a listener's F_{FA} had changed notably between experiments, this was paralleled by a change in $(L_{T(RP; -FA)} - L_{T(D)})$ in much the same way as that seen across listeners (Fig. 11b).

This relationship could, in some sense, also be viewed as a speed-accuracy trade-off. An increase in the required amount of evidence by a certain number of decibels, with its concomitant decrease in F_{FA} , is equivalent to, and would manifest itself in, a relative displacement of the $RQ_{(-FA)}(L)$ function to the right by the same number of decibels. Thus, with everything else unchanged, the listener would produce fewer false alarms, when the required amount of evidence is high than when it is low, but would take longer to respond to a stimulus of a given SPL. The delay would be more pronounced as the SPL becomes lower, consistent with the observed stimulus-intensity-dependent effects of speed versus accuracy emphasis on RT (e.g., Murray 1970; Sanford 1972; Miller and Ulrich 2003). On the other hand, however, and provided the stimuli cover the steep portion of the psychometric function, the proportion of misses will also increase when the required amount of evidence increases. This stresses the need to precisely define the term accuracy before dealing with this type of trade-off.

We also observed a negative relationship between $RQ_{\min}$ and F_{FA} across listeners (Fig. 11d), but, remarkably, the $RQ_{\min}$ of a given listener changed little between experiments despite changes in F_{FA} . This intralister stability of $RQ_{\min}$ contrasts with the clear inverse relationship of $(L_{T(RP; -FA)} - L_{T(D)})$ with F_{FA} in the same listeners (cf. Fig. 11b) and with the inverse relationships of both $(L_{T(RP; -FA)} - L_{T(D)})$ and $RQ_{\min}$ with F_{FA} across listeners (Fig. 11b, d). Currently, we have no simple explanation for this phenomenon. In this context, it is interesting that Seitz and Rakerd (1997) observed that an individual's overall speed is largely modality-independent.

Simple RT and loudness

Simple RT to auditory stimuli is generally thought to be a measure of the loudness of the test stimuli. The idea was originally proposed by Chocholle (1940), who studied RT as a function of sound frequency and level. He found that at each frequency RT decreased as level increased, a result that confirmed even earlier observations (Wundt 1874; Cattell 1886; Piéron 1920), and that equally loud tones of different frequencies resulted in equal RTs. A number of subsequent investigations obtained results that supported Chocholle's conclusion of a close inverse relationship between loudness and RT (e.g., McGill 1961; Pfingst et al. 1975a; Marshall and Brandt 1980; Humes and Ahlstrom 1984; Wagner et al. 2004; Florentine et al. 2005; see also Scharf 1978). To our knowledge, only Kohfeld and colleagues (Santee and Kohfeld 1977; Kohfeld et al. 1981) have suggested that the relationship between loudness and RT may not be as close and argued that the detection of a signal and the estimation of its loudness require different perceptual operations. However, the validity of their results was questioned (Buus et al. 1982; Luce 1986, p. 70), because their equal-loudness contours differed substantially from those in most other studies. The general consensus, therefore, seems to be that there is a close relationship between loudness and RT. Consequently, it has been proposed that RT may serve as an indirect estimate or surrogate for loudness and that RT measurements are a useful tool to investigate loudness perception (e.g., Chocholle 1940; Stebbins 1966; Reason 1972; Moody 1973, 1979; Pfingst et al. 1975a; Dooling et al. 1975; Marshall and Brandt 1980; Buus et al. 1982; Humes and Ahlstrom 1984; Seitz and Rakerd 1997; Leibold and Werner 2002; Ariei and Marks 2003; Wagner et al. 2004; Florentine et al. 2005; Little and May 2005; see also Scharf 1978; Luce 1986). Snodgrass (1975, p. 37ff) has even suggested that RT may be used more generally as an index of the magnitude of a listener's sensation.

Our data shed some light onto the link between simple RT and loudness and the constraints under which the two measures may be correlated. Our data provide good evidence that simple RTs, once corrected for a plausible irreducible minimum reaction time, reflect the listener's reaction threshold. Note that the reaction thresholds extracted from RQ are very similar to, if not identical with, those extracted from RP, and that they can maintain an approximately constant and small difference to the listener's detection thresholds, despite considerable changes in stimulus SPLs and associated changes in loudness (Figs. 10 and 11). Hence, simple RTs reflect threshold more widely than hitherto assumed in the literature, where a link between RT and threshold seems to have been suspected only to the extent that iso-RT versus frequency functions for very long RTs yielded contours in close proximity to, and of similar shape as, an animal's audiogram (e.g., Pfingst et al. 1975a; Dooling et al. 1975). It is conceivable, and even plausible given the listener's task in a simple RT paradigm, namely, to react upon detection of the stimulus, that simple RT may actually reflect thresholds rather generally. How is it then that simple RT can correlate with loudness? We suggest here that the link is via a phenomenon called "temporal summation of loudness," which refers to the observation that the loudness of a tone or noise stimulus increases as stimulus duration increases, with SPL unchanged (e.g., Munson 1947; Niese 1959; Port 1963; Reichardt and Niese 1965, 1970; Ekman et al. 1966; Scharf 1978; Poulsen 1981; Kumagai et al. 1984; Sone et al. 1986; Namba 1987; Takeshima et al., 1988; Ogura et al. 1991; Florentine et al. 1996; Buus et al. 1997; Verhey and Kollmeier 2002; Sanpetrino and Zwislocki 2004). In combination with these observations, our data suggest the following reasoning. When different stimuli lead to identical RT *and* when they are also perceived as equally loud by a listener, then the temporal summation of loudness for the different stimuli must have been identical or at least very similar. If this reasoning is correct, then an experimenter who wants to use RT as a measure of loudness must ensure identical temporal summation of loudness for the test stimuli. This might require the stimuli to have identical temporal amplitude envelopes and durations, although it seems to be the case that they can differ in their overall amplitudes and, of course, differ in spectra. It is easy to imagine stimuli that are detected at the same time after onset but which are perceived as different with respect to loudness, e.g., stimuli whose amplitude envelopes differ markedly beyond the common point in time after onset at which they are detected. In this context, several studies have observed that RT can be independent of stimulus duration (Gregg and Brogden

1950a,b; Stebbins 1966 and as also shown here; see Fig. 10), or even increase with increasing stimulus duration. This is noteworthy because, in light of temporal summation of loudness, loudness would be expected to increase as stimulus duration increases. Clearly, therefore, the conditions under which simple RT and loudness can be expected to be closely correlated are constrained.

ACKNOWLEDGMENTS

The study was supported by grants of the Deutsche Forschungsgemeinschaft (He1721/4-2, He1721/7-1, and SFB-TR 31 A6) and the GEERS Stiftung to P.H. Dr. Michael Brosch and two anonymous reviewers provided helpful comments on earlier versions of the manuscript.

REFERENCES

- ABEL SM, RAJAN DK, GIGUERE C. Stimulus parameters in detection and reaction time. *Scand. Audiol.* 19:229–235, 1990.
- ARIEH Y, MARKS LE. Recalibrating the auditory system: A speed-accuracy analysis of intensity perception. *J. Exp. Psychol. Hum. Percept. Perform.* 29:523–536, 2003.
- BONNET C. Foreword: Reaction time psychophysics. *Proc. 16th Ann. Meet. Int. Soc. Psychophys.* 1–4, 2000.
- BONNET C, ZAMORA MC, BURATTI F, GUIRAO M. Group and individual gustatory reaction times and Piéron's law. *Physiol. Behav.* 66:549–558, 1999.
- BURBECK SI, LUCE RD. Evidence from auditory simple reaction times for both change and level detectors. *Percept. Psychophys.* 32:117–133, 1982.
- BUUS S, GREENBAUM H, SCHARF B. Measurements of equal loudness and reaction times. *J. Acoust. Soc. Am.* 72(Suppl 1):594, 1982.
- BUUS S, FLORENTINE M, POULSEN T. Temporal integration of loudness, loudness discrimination, and the form of the loudness function. *J. Acoust. Soc. Am.* 101:669–680, 1997.
- CATTELL J MCK. The influence of the intensity of the stimulus on the length of the reaction time. *Brain* 8:512–515, 1886.
- CHOCOLLE R. Variation des temps de réaction auditif en fonction de l'intensité à diverses fréquences. *Ann. Psychol.* 41:65–124, 1940.
- CLOCK EDDINS A, SALVI RR, SAUNDERS SS, POWERS NL. Neural correlates of temporal integration in the cochlear nucleus of the chinchilla. *Hear. Res.* 71:37–50, 1993.
- CLOCK EDDINS A, SALVI RR, WANG J, POWERS NL. Threshold-duration functions of chinchilla auditory nerve fibers. *Hear. Res.* 119:135–141, 1998.
- DAI H, WRIGHT BA. Predicting the detectability of tones with unexpected duration. *J. Acoust. Soc. Am.* 105:2043–2046, 1999.
- DAU T, PISCHER D, KOHLRAUSCH A. A quantitative model of the "effective" signal processing in the auditory system. I. Model structure. *J. Acoust. Soc. Am.* 99:3615–3622, 1996.
- DOOLING RJ, ZOLOTH SR, BAYLIS JR. Auditory sensitivity, equal loudness, temporal resolving power, and vocalizations in the house finch (*Carpodacus mexicanus*). *J. Comp. Physiol. Psychol.* 92:867–876, 1975.
- EDDINS DA, GREEN DM. Temporal integration and temporal resolution. In: Moore BCJ (ed) *Hearing*. San Diego, Academic Press, pp. 207–242, 1995.
- EKMAN G, BERGLUND B, BERGLUND U. Loudness as a function of duration of auditory stimulation. *Scand. J. Psychol.* 7:201–208, 1966.
- FLORENTINE M, BUUS S, POULSEN TJ. Temporal integration of loudness as a function of level. *J. Acoust. Soc. Am.* 99:1633–1644, 1996.
- FLORENTINE M, BUUS S, ROSENBERG M. Reaction-time data support the existence of softness imperception in cochlear hearing loss. In: Pressnitzer D, de Cheveigne A, McAdams S, Collet L (eds) *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*. Springer-Verlag, New York, pp. 30–39, 2005.
- GERKEN GM, SANDLIN D. Auditory reaction time and absolute threshold in cat. *J. Acoust. Soc. Am.* 61:602–607, 1977.
- GERKEN GM, BHAT VKH, HUTCHISON-CLUTTER M. Auditory temporal integration and the power function model. *J. Acoust. Soc. Am.* 88:767–778, 1990.
- GESCHIEDER GA. *Psychophysics: The Fundamentals*. Lawrence Erlbaum Associates, Mahwah, NJ, 1997.
- GREGG LW, BROGDEN JW. The relation between duration and reaction time difference to fixed duration and response terminated stimuli. *J. Comp. Physiol. Psychol.* 43:329–337, 1950a.
- GREGG LW, BROGDEN JW. The relation between reaction time and the duration of the auditory stimulus. *J. Comp. Physiol. Psychol.* 43:389–395, 1950b.
- GRICE GR. Stimulus intensity and response evocation. *Psychol. Rev.* 75:359–373, 1968.
- HEIL P. First-spike latency of auditory neurons revisited. *Curr. Opin. Neurobiol.* 14:461–467, 2004.
- HEIL P, NEUBAUER H. Temporal integration of sound pressure determines thresholds of auditory-nerve fibers. *J. Neurosci.* 21:7404–7415, 2001.
- HEIL P, NEUBAUER H. A unifying basis of auditory thresholds based on temporal summation. *Proc. Natl. Acad. Sci. U.S.A.* 100:6151–6156, 2003.
- HEIL P, NEUBAUER H. New insights into absolute thresholds of normal and hearing-impaired ears. *Audiol. Akust.* 43:188–195, 2004.
- HENRY JA, FLICK CL, GILBERT A, ELLINGSON RM, FAUSTI SA. Reliability of hearing thresholds: Computer-automated testing with ER-4B Canal Phone earphones. *J. Rehabil. Res. Dev.* 38:567–581, 2001.
- HUMES LE, AHLSTROM JB. Relation between reaction time and loudness. *J. Speech Hear. Res.* 27:306–310, 1984.
- JOHN ID. A statistical decision theory of simple reaction time. *Aust. J. Psychol.* 19:27–34, 1967.
- KLEIN SA. Measuring, estimating, and understanding the psychometric function: A commentary. *Percept. Psychophys.* 63:1421–1455, 2001.
- KOHFELD DL. Effects of intensity of auditory and visual ready signals on simple reaction time. *J. Exp. Psychol.* 82:88–95, 1969.
- KOHFELD DL, SANTEE JL, WALLACE ND. Loudness and reaction time. I. *Percept. Psychophys.* 29:535–549, 1981.
- KORNBLUM S. Simple reaction time as a race between signal detection and time estimation. *Percept. Psychophys.* 13:108–112, 1973.
- KRISHNA BS. A unified mechanism for spontaneous-rate and first-spike timing in the auditory nerve. *J. Comp. Neurosci.* 13:71–91, 2002.
- KUMAGAI M, SUZUKI Y, SONE T. Comparison of loudness of impact sounds with and without steady duration (a study on the loudness of impact sound. V). *J. Acoust. Soc. Jpn. (E)* 5:31–36, 1984.
- LEIBOLD IJ, WERNER LA. Relationship between intensity and reaction time in normal-hearing infants and adults. *Ear Hear.* 23:92–97, 2002.
- LEVITT H. Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49:467–477, 1971.
- LITTLE N, MAY BJ. Equal loudness contours for the domestic cat. *Abstr. 28th Meeting Assoc. Res. Otolaryngol.* 725, 2005.
- LUCE RD. Response Times. Their Role in Inferring Elementary Mental Operations. Oxford University Press, New York, 1986.
- LUCE RD, GREEN DM. A neural timing theory for response times and the psychophysics of intensity. *Psychol. Rev.* 79:14–57, 1972.
- MARSHALL L, BRANDT JF. The relationship between loudness and reaction time in normal hearing listeners. *Acta Otolaryngol.* 90:244–249, 1980.
- MCGILL WJ. Loudness and reaction time: A guided tour of the listener's private world. *Acta Psychol.* 19:193–199, 1961.

- McKEEFY DJ, PARRY NRA, MURRAY IJ. Simple reaction times in color space: The influence of chromaticity, contrast, and cone opponency. *Invest. Ophthalmol. Vis. Sci.* 44:2267–2276, 2003.
- MEDDIS R. Auditory-nerve first-spike latency and auditory absolute threshold: A computer model. *J. Acoust. Soc. Am.* 119:406–417, 2006.
- MEYER DE, IRWIN DE, OSMAN AM, KOUNIOS J. The dynamics of cognition and action: Mental processes inferred from speed–accuracy decomposition. *Psychol. Rev.* 95:183–237, 1988.
- MILLER J, ULRICH R. Simple reaction time and statistical facilitation: A parallel grains model. *Cogn. Psychol.* 46:101–151, 2003.
- MOODY DB. Behavioral studies of noise-induced hearing loss in primates: Loudness recruitment. *Adv. Oto-Rhino-Laryngol.* 20:82–101, 1973.
- MOODY DB. Partial masking in the monkey: Effect of masker bandwidth and masker band spacing. *J. Acoust. Soc. Am.* 66:107–114, 1979.
- MOORE BCJ, OXENHAM AJ. Psychoacoustic consequences of compression in the peripheral auditory system. *Psychol. Rev.* 105:108–124, 1998.
- MUNSON W. The growth of auditory sensation. *J. Acoust. Soc. Am.* 191:584–591, 1947.
- MURRAY HG. Stimulus intensity and reaction time: Evaluation of a decision-theory model. *J. Exp. Psychol.* 84:383–391, 1970.
- NAMBA S. On the psychological measurement of loudness, noisiness and annoyance: A review. *J. Sound Vib.* 116:491–507, 1987.
- NEUBAUER H, HEIL P. Towards a unifying basis of auditory thresholds: The effects of hearing loss on temporal integration reconsidered. *J. Assoc. Res. Otolaryngol.* 5:436–458, 2004.
- NIESE H. Die Trägheit der Lautstärke in Abhängigkeit vom Schallpegel. *Hochfrequenztech. Elektroakust.* 68:143–152, 1959.
- OGURA Y, SUZUKI Y, SONE T. A temporal integration model for loudness perception of repeated impulsive sounds. *J. Acoust. Soc. Jpn. (E)* 12:1–11, 1991.
- OLLMAN RT, BILLINGTON MJ. The deadline model for simple reaction times. *Cogn. Psychol.* 3:311–336, 1972.
- PINGST BE, HIENZ R, KIMM J, MILLER J. Reaction-time procedure for measurement of hearing. I. Suprathreshold functions. *J. Acoust. Soc. Am.* 57:421–430, 1975a.
- PINGST BE, HIENZ R, MILLER J. Reaction-time procedure for measurement of hearing. II. Threshold functions. *J. Acoust. Soc. Am.* 57:431–436, 1975b.
- PIÉRON H. Nouvelles recherches sur l'analyse du temps de latence sensorielle en fonction des intensités excitatrices. *Ann. Psychol.* 22:58–142, 1920.
- PINS D, BONNET C. On the relation between stimulus intensity and processing time: Piéron's law and choice reaction time. *Percept. Psychophys.* 58:390–400, 1996.
- PINS D, BONNET C. The Piéron function in the threshold region. *Percept. Psychophys.* 62:127–136, 2000.
- PORT E. Über die Lautstärke einzelner kurzer Schallimpulse. *Acustica* 13:212–223, 1963.
- POULSEN T. Loudness of tones in a free field. *J. Acoust. Soc. Am.* 69:1786–1790, 1981.
- REASON JT. Some correlates of the loudness function. *J. Sound Vib.* 20:305–309, 1972.
- RECANZONE GH, BECKERMAN NS. Effects of intensity and location on sound location discrimination in macaque monkeys. *Hear. Res.* 198:116–124, 2004.
- REICHARDT DT, NIESE H. Die Addition der Schallerregungen in den einzelnen Frequenzgruppen bei impulsiven Geräuschen. *Acustica* 16:295–304, 1965.
- REICHARDT DT, NIESE H. Choice of sound duration and silent intervals for test and comparison signals in the subjective measurement of loudness level. *J. Acoust. Soc. Am.* 47:1083–1090, 1970.
- SACHS L. *Angewandte Statistik*. Springer-Verlag, Berlin, 1997.
- SANFORD AJ. Criterion effects in simple reaction time: Results with stimulus intensity and duration manipulations. *J. Exp. Psychol.* 95:370–374, 1972.
- SANPETRINO NM, ZWISLOCKI JJ. Loudness level and loudness as functions of tone duration and intensity. *Abstr. Assoc. Res. Otolaryngol.* 127, 2004.
- SANTEE JL, KOHFELD DL. Auditory reaction time as a function of stimulus intensity, frequency, and rise time. *Bull. Psychon. Soc.* 10:393–396, 1977.
- SCHARF B. Loudness. In: Carterette EC and Friedman MP (eds) *Handbook of Perception. IV: Hearing*. Academic Press, New York, pp. 197–242, 1978.
- SEITZ PF, RAKERD B. Auditory stimulus intensity and reaction time in listeners with long-standing sensorineural hearing loss. *Ear Hear.* 18:502–512, 1997.
- SNODGRASS JG. Psychophysics. In: Scharf B and Reynolds GS (eds) *Experimental Sensory Psychology*. Scott, Foresman & Co., Glenview, IL, pp. 17–67, 1975.
- SONE T, SUZUKI Y, KUMAGAI M, TAKAHASHI T. Loudness of a single burst of impact sound: Results of round robin test in Japan (I). *J. Acoust. Soc. Jpn. (E)* 7:173–182, 1986.
- STAFFORD T, GURNEY KN. The role of response mechanisms in determining reaction time performance: Piéron's law revisited. *Psychon. Bull. Rev.* 11:975–987, 2004.
- STEBBINS WC. Auditory reaction time and the derivation of equal loudness contours for the monkey. *J. Exp. Anal. Behav.* 9:135–142, 1966.
- SU TI, RECANZONE GH. Differential effect of near-threshold stimulus intensities on sound localization performance in azimuth and elevation in normal human subjects. *J. Assoc. Res. Otolaryngol.* 2:246–256, 2001.
- TAKESHIMA H, SUZUKI Y, KONO S, SONE T. Growth of the loudness of tone burst with duration up to 10 seconds. *J. Acoust. Soc. Jpn. (E)* 9:295–300, 1988.
- TREUTWEIN B. Adaptive psychophysical procedures. *Vis. Res.* 35:2503–2522, 1995.
- TREUTWEIN B, STRASBURGER H. Fitting the psychometric function. *Percept. Psychophys.* 61:87–106, 1999.
- ULRICH R, GIRAY M. Separate-activation models with variable base times: Testability and checking of cross-channel dependency. *Percept. Psychophys.* 39:248–254, 1986.
- ULRICH R, MILLER J. Effects of truncation on reaction time analysis. *J. Exp. Psychol. General* 123:34–80, 1994.
- ULRICH R, MILLER J. Tests of race models for reaction time in experiments with asynchronous redundant signals. *J. Math Psychol.* 41:367–381, 1997.
- VERHEY JL, KOLLMEIER B. Spectral loudness summation as a function of duration. *J. Acoust. Soc. Am.* 111:1349–1358, 2002.
- VIEMEISTER NF, WAKEFIELD GH. Temporal integration and multiple looks. *J. Acoust. Soc. Am.* 90:858–865, 1991.
- VIEMEISTER NF, SHIVAPUJA BG, RECIO A. Physiological correlates of temporal integration. In: Cazals Y, Horner K and Demany L (eds) *Auditory Physiology and Perception*. Pergamon Press, New York, pp. 323–330, 1992.
- WAGNER E, FLORENTINE M, BUUS S, MCCORMACK J. Spectral loudness summation and simple reaction time. *J. Acoust. Soc. Am.* 116:1681–1686, 2004.
- WOODWORTH RS, SCHLOSBERG H. *Experimental Psychology*. Holt, New York, 1954.
- WUNDT W. *Grundzüge der physiologischen Psychologie*. Engelmann, Leipzig, 1874.
- ZWISLOCKI JJ, RELKIN EM. On a psychophysical transformed-rule up and down method converging on a 75% level of correct responses. *Proc. Natl. Acad. Sci. U.S.A.* 98:4811–4814, 2001.
- ZWISLOCKI J, MAIRE F, FELDMAN AS, RUBIN H. On the effect of practice and motivation on the threshold of audibility. *J. Acoust. Soc. Am.* 30:254–262, 1958.