Pergamon

PII: S0042-6989(97)00022-9

Vision Res., Vol. 37, No. 15, pp. 2119–2131, 1997

Printed in Great Britain
0042-6989/97 $17.00 + 0.00

Saccadic Reaction Times: a Statistical Analysis of Multimodal Distributions

STEFAN GEZECK,*‡ BURKHART FISCHER,* JENS TIMMER†

Received 17 January 1996; in revised form 6 August 1996; in final form 6 November 1996

The distributions of saccadic reaction times (SRT) often deviate from unimodal normal distributions. An excess-mass procedure was used to detect peaks in 963 data sets containing 90,927 reaction times from 170 subjects. About 55% showed one, 30% two, 12% three and 3% four peaks. According to their clustering along the reaction time scale the modes could be classified into express (90–120 msec), fast regular (135–170 msec) and slow regular (200–220 msec) modes. Among the unimodal distributions 29% had peaks in the range of the express mode and 46% had peaks in the range of the fast regular mode. Therefore, 87% of the data sets support the notion of saccadic reaction time distributions being the superposition of three modes. All experimental distributions were fitted by as many gamma distributions as determined by the excess-mass test. The significance of the multimodality for saccade generation processes is discussed. © 1997 Elsevier Science Ltd.

Saccade Express saccade Multimodality Excess mass test

INTRODUCTION

Reaction time measurements have been used widely to understand certain sensory and/or motor functions. In particular, saccadic reaction times (SRT), i.e., the time between the onset of a target stimulus and the beginning of the saccade, have been studied extensively in oculomotor research and cognitive sciences. A common procedure is to measure a number of reaction times in one experimental condition (the control) and to compare these data with those obtained in another condition (the test). Usually, mean values and standard deviations are computed from each set of data and compared statistically to evaluate whether or not they are significantly different. However, saccadic reaction times almost never form gaussian (normal) distributions. Positive skewness is often observed and makes it difficult to describe such data by calculating the mean and standard deviation, and to compare certain test values with the control values.

Furthermore SRT-distributions can exhibit quite strong deviations from normal distributions by forming distinct peaks or modes. In some cases these modes can quite easily be identified by visual inspection of the SRT-distribution, because they occur sufficiently apart from each other. These observations have been reviewed by Fischer & Weber (1993) and have been replicated recently by Nozawa *et al.* (1994). The interpretation of the reaction time data and their implications for the underlying brain functions critically depends on the existence of such modes.

A conservative notion maintains that with changing experimental conditions a unimodal distribution would be shifted to lower or higher mean values, and two or more modes appear by chance rather than being the result of a systematic stochastic feature of the system producing the data. An alternative notion assumes that one or the other mode has more or less weight in the distribution depending on the experimental condition, but remaining essentially at the same position along the reaction-time scale. In extreme cases one mode can be replaced by another mode when conditions are changed, or two modes may merge into each other.

Examples for multimodal SRT-distributions have been published by several authors in monkeys and in human subjects (Fischer & Boch, 1983; Fischer & Ramsperger, 1984; Jüttner & Wolf, 1992; Munoz & Wurtz, 1992; Sommer & Schiller, 1992; Nothdurft & Parlitz, 1993; Rohrer & Sparks, 1993; Currie *et al.*, 1993; Priori *et al.*, 1993; Matsue *et al.*, 1994; Schiller & Lee, 1994; Tam & Ono, 1994). Other authors find multimodal distributions in some subjects but not in others (Reuter-Lorenz *et al.*, 1991). Nozawa *et al.* (1994) report that 23% of their subjects produced clearly separate modes and "many others did not have clearly defined modes but could still have come from a mixture of two distributions". Still other groups report problems in finding clear bimodal

*Brain Research Unit, Institute of Biophysics, University of Freiburg, Hansastr. 9a, D 79104 Freiburg, Germany.
†Center for Data Analysis and Modelling, University of Freiburg, Freiburg, Germany.
‡To whom all correspondence should be addressed [*Fax*: +49-761 2039540; *Email*: gezeck@ruf.uni-freiburg.de].

distributions (Wenban-Smith & Findlay, 1991; Sereno & Holzman, 1993). The issue has become controversial with respect to the existence of a separate mode of express saccades in human subjects (Fischer & Weber, 1993; Kingstone & Klein, 1993).

While in the overlap task the fixation point remains present, when the target appears, it is turned off in the gap task some time (the gap duration) before target onset. Saslow was the first to use gap and overlap trials. He reported that under gap conditions saccadic reaction times are considerably shorter (about 140 msec) than under overlap conditions (about 200 msec) (Saslow, 1967). This phenomenon is called the gap effect. When the gap experiment was repeated in monkeys (Fischer & Boch, 1983) and in human subjects (Fischer & Ramsperger, 1984), the analysis of the reaction times on the basis of their distributions revealed a mode about 100 msec, which became larger, when the gap duration was increased from 0 to 200 msec. When increasing the gap duration to 400 msec, this mode became smaller and the mode at about 150 msec increased (Weber *et al.*, 1995; Weber & Fischer, 1995).

Reuter-Lorenz *et al.* (1991) attributed the gap effect to a facilitation of premotor programming in the superior colliculus. Reulen (1984) attributed the latency reduction to facilitated sensory processing. A general shortcoming of his facilitation concept is that it cannot explain the occurrence of more than one mode in the latency distribution; neither can it account for the increase of SRT when increasing the gap duration from 200 to 400 msec.

A model for saccadic reaction time, which is based on the idea that the SRT includes the time consumption for the afferent and efferent processes (for example, relays in pathway) and central processes such as computation of the movement metrics, was presented by Rogal & Fischer (1986).

The three-loop model proposed by Fischer (1987) is based on neurobiological findings. Fischer proposed that saccades are generated by three main pathways (loops) connecting the retina of the eye with the efferent eye movement generating system. Each loop is associated with a certain brain process that must be accomplished during saccade preparation, thus contributing to the reaction time. Saccades generated through the shortest loop are the express saccades (mean SRT = 100–135 msec) forming the first peak. If two or three processes have to be completed after target onset fast regular saccades (140–180 msec) or slow regular saccades (above 200 msec) are obtained again forming separate peaks. Fischer *et al.* (1995) presented a computer simulation of the 3 loop model. It reproduced complex experimental reaction time distributions very similar to the experimental data by a simple neural network.

Similar to the facilitation model (Reulen, 1984) Carpenter and Williams presented a model based on the idea that the presence of a target causes a signal in a decision unit that rises linearly until it reaches a fixed threshold (Carpenter & Williams, 1995). The proposed gaussian distribution of the slope leads to a SRT-distribution with positive skewness. Carpenter and Williams plotted the cumulative distribution of the data (overlap task) against the cumulative distribution of the model. This method is suitable for data presentation, because all data point form one straight line, if the assumptions of unimodality and the shape of the distribution are correct [see Chambers *et al.* (1983) for an introduction to quantile–quantile plots]. The results show that straight lines are only obtained in two segments with an "elbow" around 140 msec. The failure of the data points to form a single straight line indicates that the model assumption of only one mode was wrong. The conclusion of the presence of two modes, a small express mode and a large mode is compatible with the results of Fischer & Boch (1983) for monkeys and Fischer *et al.* (1993) for naïve and young persons, and the results of this study.

In this study we have used large numbers of distributions, each containing between 75 and 200 single reaction times. Saccades directed to the left and right side were analyzed separately, because many subjects were more or less asymmetric in performing the tasks. Data obtained in gap and overlap trials from naïve and trained adult normal subjects were analyzed. The excess-mass procedure (Müller & Sawitzki, 1991) was used for peak detection. This method determines the number and the position of different modes in the data. A superposition of corresponding gamma distributions with the number and position of modes taken from the excess-mass procedure was fitted to the data with a least square fitting procedure. Fit parameters were the weights, means and standard deviations of each mode.

The results strongly support the view that the hypothesis of only one mode must be rejected in 45% of cases. The number of modes and their strengths depend on the subjects and the experimental conditions. This article also provides a statistical tool for handling multimodal distributions by describing them by a small set of numbers.

METHODS

Subjects and database

Altogether, 170 normal subjects contributed to the database of this study. Ten were adults (age 30–50 yr), who in advance performed the gap task (gap = 200 msec) until stable distributions were obtained. They are considered trained subjects in the sense of Fischer *et al.* (1984). All others were naïve (untrained) with respect to the scientific purpose of the study and the saccade tasks used. Their age ranged from 20 to 62 yr. Each subject performed at least 75 trials in one session. In this way 963 data sets containing 90,927 reaction times were collected and analyzed.

Data collection

All reaction time data were collected using the same

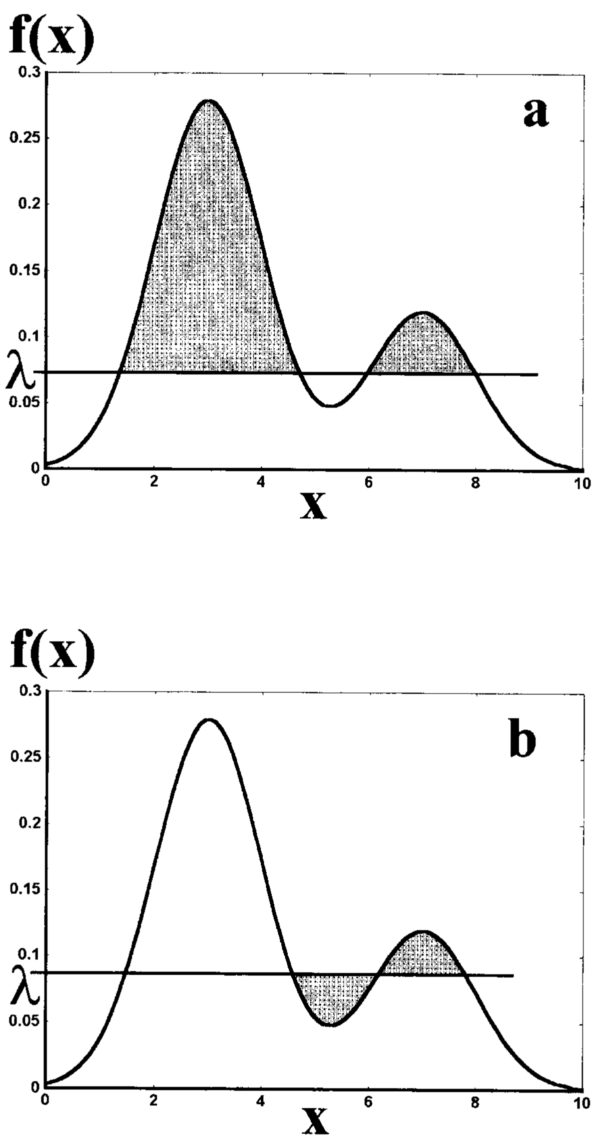

FIGURE 1. (a) The excess-mass is the sum of contributions from the λ-clusters, that is, the intervals where $f(x) \geq \lambda$. (b) The maximal excess-mass difference measures the amount of mass to be moved to convert a bimodal distribution into a unimodal one. Here the gap between the peaks is filled and the unimodal function g is created when the shadowed region above λ is moved to the left.

methods described in detail earlier (Weber *et al.*, 1995). The Iris Scalar system was used for recording the position of the left eye. Reaction times were determined off-line using a velocity criterion for detecting saccades. The latency was defined by the time the velocity reached 15% of its maximum in each detected saccade. Latency values below 80 msec were considered anticipatory by means of the occurrence of direction errors (Wenban-Smith & Findlay, 1991) and excluded from this analysis.

Visual display

All visual stimuli were generated on a RGB monitor using a high resolution graphic interface (Mirograph 510). Target onset time was synchronized to the screen (frame rate 83 Hz) and the position of the target at the screen. All saccades were made to targets presented 4 deg to the left or right in random order. The size of the white target was 0.2 deg × 0.2 deg, the red fixation point was 0.1 deg × 0.1 deg in size. Both had a luminance of 50 cd/m^2, while the green background luminance was 10 cd/m^2 bright and 20 deg × 15 deg in size. Viewing distance was 57 cm.

Types of saccade tasks

The subjects performed the gap task with a gap duration of 200 msec (unless otherwise stated) and the overlap task. The fixation period prior to target presentation was 1.2 sec. The targets remained visible for 800 msec. The intertrial interval was 1 sec. Subjects were instructed to fixate the fixation point and to look at the target when it appeared. They were not encouraged to respond "as fast as possible". No feedback was given.

STATISTICAL METHODS

Suggest someone draws random numbers $r_1..r_A$ from a bimodal distribution f. His estimate $\hat{f}$ of this distribution may by chance exhibit three or more modes. If the number of modes of f is unknown, a statistical method is necessary to decide how many modes of $\hat{f}$ are significant.

The excess-mass test

The excess-mass test of Müller and Sawitzki was used for testing the distributions for multimodality (Müller & Sawitzki, 1991). The idea is that a mode is present where an excess of probability mass is concentrated. The excess-mass $E(\lambda)$ measures the amount of probability mass exceeding level λ. Here the reaction times $r_1..r_A$ are considered as realizations of random variables of an unknown distribution f and used to form the estimate $\hat{f}$.

An example is shown in Fig. 1(a). There are two intervals where f exceeds the level λ. Müller and Sawitzki call these intervals $\{x\colon f(x) \geq \lambda\}$ λ-clusters. As λ increases, the λ-clusters concentrate on local maxima of f, and the amount of excess probability $E(\lambda)$ becomes smaller. If λ decreases, the intervals become larger until they contact and fuse to one large interval.

If the density $f(x)$ has m λ-clusters $C_1 \ldots C_m$, the excess $E(\lambda)$ can be expressed as a sum over the different contributions:

$$E_m(\lambda) = \sum_{j=1}^{m} \int_{C_j} (f(x) - \lambda)dx.$$

The index m indicates the modality of the λ-clusters. Müller and Sawitzki choose $\Delta_n = \max_\lambda E_{n+1}(\lambda) - E_n(\lambda)$ as a test statistic of the excess-mass test for n-modality. It measures the minimal amount of probability mass that has to be moved in order to convert the $(n+1)$modal distribution into an n-modal distribution. As shown in Fig. 1(b) for a bimodal distribution f, the quantity Δ_1 is half the total variation distance between f and the closest unimodal distribution g:

$$\Delta_1 = \max_\lambda\{E_2(\lambda) - E_1(\lambda)\} = \frac{1}{2}\inf_{g|g\ unimodal} \| f - g \| .$$

As a further example an analogous interpretation of Δ_2 yields $\Delta_2 = \frac{1}{2}\inf_{g|g\ bimodal} f - g$.

Unfortunately, the distribution of the test statistic Δ is unknown, and depends heavily on differential properties of the underlying unimodal distribution.

Considering the λ-clusters it is possible to gain information about the position of the peak. Here the mean μ^{EM} of the probability mass is used as peak position:

$$\mu_j^{EM} = \frac{\int_{C_j} x\, f(x)dx}{\int_{C_j} f(x)dx}.$$

Estimating the number of modes

The distribution of the test-statistic Δ is simulated with a smooth bootstrap procedure. Assume the hypothesis is that the distribution consists of n modes. The idea is the same as in Efron & Tibshirani (1991), section 16.5. Contrary to the normal bootstrap procedure, where the observed data were resampled, the random numbers were drawn from the smooth n-modal estimate of the population. Here, the n-modal estimator is constructed with the n largest λ-clusters.

A simulations study not presented here reveals that the results obtained for Δ do not vary when incrementing the number of values for λ above 35, opposite to the case when using less than 35 different values. Therefore the λ-clusters were computed for 35 different values of λ.

Assume the measurements are $r_1 \ldots r_A$ and the question is whether the underlying distribution f has n modes. According to Efron, an n modal estimator of f is formed from the observed values $r_1 \ldots r_A$. Data sets $s_1 \ldots s_A$ consisting of realizations of random variables of this distribution $\hat{f}_n$ are generated, each has the same size A as the observed data set $r_1 \ldots r_A$. Then the value Δ_n^{sim} is built in the same way for the simulated data $s_1 \ldots s_A$ as the value Δ_n for the real data $r_1 \ldots r_A$. A smaller value $\Delta_n^{sim} < \Delta_n$ contradicts the assumption of n-modality in favor of the $(n+1)$-modality. This method of reaching significance was proposed first by Silverman (1981).

After simulating 100 data sets $s_1 \ldots s_A$ the fraction of simulations p_n is considered, where $\Delta_n^{sim} < \Delta_n$. If this fraction p_n exceeds a given significance level, the hypothesis of n-modality is rejected in favor of the alternative of $(n+1)$-modality.

In order to investigate the reported p-values of the proposed testing procedure, a simulation study was conducted. One hundred random numbers are drawn of a normal distributed random variable. Then unimodality is tested using the test described above. This procedure is repeated 1000 times. The 1000 p-values are plotted in a quantile–quantile plot. If the reported p-values are to be accurate (or conservative), the quantiles of these p-values from truly unimodal distributions should lie on (or above) the line $x = y$. The results can be seen in Fig. 2. The quantiles of the p-values are all well above the line $x = y$ indicating that the probability of obtaininga p-value of α is less than α.

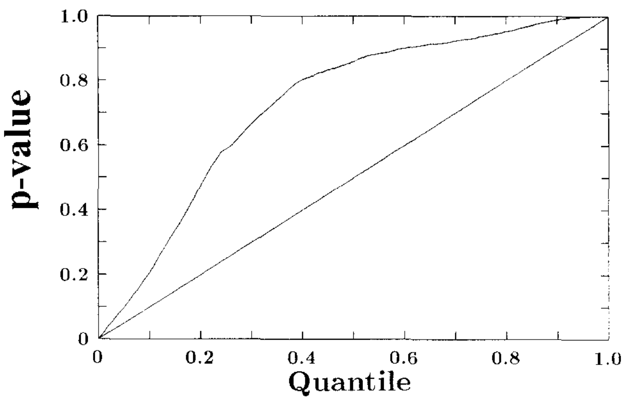

FIGURE 2. Quantile–quantile plot of 1000 p-values generated from a sample of 100 values of a normal distributed random variable. The reported p-values are conservative, because the quantiles of these p-values from truly unimodal distributions lie above the line $x = y$.

Because of the very conservative p-values, the test-procedure described above may detect only one mode in a three-modal distribution. Therefore, very often significance levels higher than 5% were chosen in order to detect the true number of modes (Silverman, 1981; Matthews, 1983; Wong, 1985).

Here we apply the idea of the excess-mass test in order to estimate the number of modes. When estimating the number of modes the procedure is the same compared with testing the number of modes: the number of modes is incremented until the p-value exceeds a threshold level p_{thr}. However, the threshold level p_{thr} is fixed in order to reproduce the true number of modes via a simulation study, when estimating the number of modes. The result of this simulation study not presented here is that the probability of correct identification becomes maximal, if the threshold level p_{thr} is set to $p_{thr} = 30\%$. If a larger threshold level $p_{thr} > 30\%$ is chosen, the method detects three or more peaks in a sample of a bimodal distribution, and if a smaller threshold level $p_{thr} < 30\%$ is chosen, it very often detects only one peak in a sample of a bimodal distribution. Estimating the number of modes of a data set $r_1 \ldots r_A$ is carried out by sequentially computing p_n for $n = 1,2,3, \ldots$ until $p_n < p_{thr}$.

A simulation study was performed to systematically investigate the quality of the estimation. Figure 3 shows the probability of detecting the correct number of peaks at the correct position with an error of no more than 10 msec, detected in a superposition of two identical gaussians (standard deviation σ of 10 msec each) with increasing difference between their mean values. While the superposition of two gaussian functions is unimodal for distances below 2σ, it becomes bimodal for larger distances. From top to bottom the number of data points simulated is increased from 100, 200, 400 to 1000. Quite clearly, the procedure reveals the existence of two correct peaks with a probability of more than 50% only when they are separated by more than 35 msec, i.e., more than three times their standard deviation.

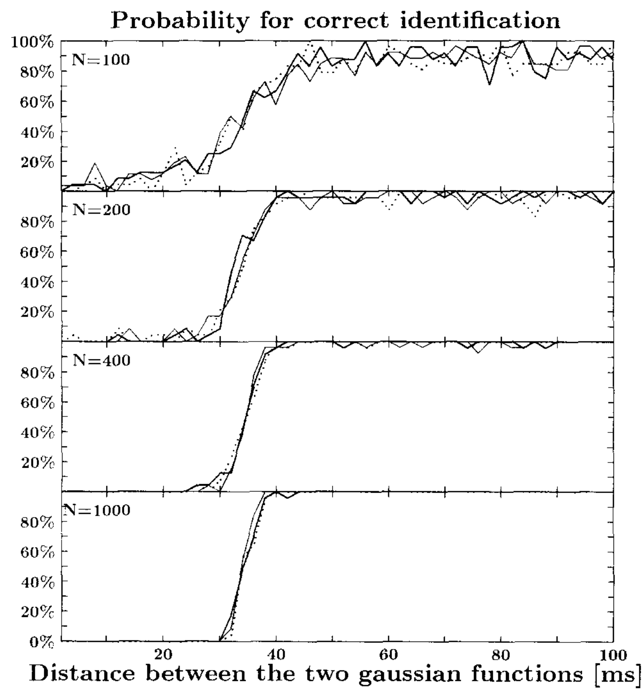

FIGURE 3. The probability of detecting two peaks at the correct positions is plotted as a function of the distance of the mean values of two gaussian functions. Each gaussian had a standard deviation of 10 msec. The relative weight of the two functions was 5:5 (thick lines), 7:3 (thin lines), and 9:1 (dashed lines). The number of simulated data points is varied from 100 to 1000 as indicated in each plot. For 100 simulated data points this probability is greater than 80% only when the difference exceeds 4 standard deviations.

We analyzed the wrong decisions and found, that in most cases the number of peaks is underestimated and that only in a very few cases the excess-mass estimator detects three peaks in these bimodal samples. In summary, the analysis of Fig. 3 shows that the proposed procedure is rather conservative, detecting separate peaks only in very clear cases. It can accept one peak even when the distribution certainly does not look unimodal.

Fitting the data by gamma distributions

The Levenberg–Marquardt fit-procedure (Press *et al.*, 1986) was used to fit the data by a superposition of n peaks each having the form of a gamma distribution:

$$g(t; \rho, a, t_0) = \frac{\rho(\rho(t - t_0))^{a-1} e^{-\rho(t-t_0)}}{\Gamma(a)}.$$

Its mean is $\mu = (a/\rho) + t_0$, its standard deviation is $\sigma = \sqrt{a}/\rho$ and its skewness is $s = 2/\sqrt{\rho}$. The parameters ρ, a and t_0 can be calculated from the values of μ, σ and s:

$$a = \frac{4}{s^2};\ \rho = \frac{2}{\sigma s};\ t_0 = \frac{\mu s - 2\sigma}{s}$$

We present only the values of μ, σ and s for the fits, with the equation above it is possible to evaluate the values of ρ, a and t_0. Notice that the mean $\mu = a/\rho$ is greater than the maximum $a - 1/\rho$ of the distribution, because the distribution has a positive skewness. In a n-modal mixture distribution, $n - 1$ additional parameters occur that describe the mixing probabilities.

The number n of gamma functions that were fitted to the data is the number of peaks detected by the excess-mass estimator. The mean μ_i, $i = 1\ ..\ n$ may vary only 10 msec from the value μ_i^{EM} identified by the procedure. The weight and the values σ_i and s_i of each mode i are free parameters of the fitting procedure.

The data $r_1\ ..\ r_A$ are grouped into bins $b_1\ ..\ b_B$ with binwidth 5 msec; B is the largest reaction time divided by the binwidth. We performed a least-square minimization. Let g_i be the fraction of the superposition of the gamma functions $g(t;\rho,a)$ lying in the i-th bin. Then the error of the model is:

$$\chi^2 = \sum_{i=1}^{B} \frac{(g_i - b_i)^2}{g_i}.$$

A large error χ^2 indicates that the superposition of gamma distributions does not fit the data well. Notice that the squared error of the fit g_i is the fit g_i itself, because the number of reaction times is multinomial distributed and can be approximated with a gaussian distribution, if there are more than five events expected in the bin i. The gaussian distributed variable with mean g_i has an standard deviation $\sqrt{g_i}$. Therefore, neighboring bins were put together until $g_i \geq 5$ for all i.

The error χ^2 is χ^2-distributed with the difference between the number of bins and the number of fitted parameters as degrees of freedom. A detailed description of the χ^2-test is given in Sachs (1982) and Honerkamp (1994). Considering the χ^2 distributions we calculate the probability P, that a larger error occurs by chance, if the data $r_1\ ..\ r_A$ are realizations of a random variable of the fitted function. Because the fit is bad, if P is rather small and because the fit is good if P has a large value, the probability P indicates the goodness of fit.

Assume a superposition of n gamma distributions and a superposition of $(n + 1)$ gamma distributions are fitted to a sample of a n-modal distribution. Of course, the error χ_{n+1}^2 is smaller than the error χ_n^2. Since the number of fitted parameters is considered when computing the probability P, fitting $(n + 1)$ gamma-distributions may result in a smaller probability P_{n+1}:

$$\chi_{n+1}^2 \leq \chi_n \text{ but } P_{n+1} \leq P_n!$$

Using more-gamma distributions, the probability for sufficient number of peaks may decrease although the error χ^2 decreases!

Figure 4 shows an example of real data. The distribution of reaction times is estimated using a gaussian kernel with bandwidth 3 msec (thick line). Multiple peaks can be seen between 100 and 200 msec. These data were selected because they most clearly show the problem. The excess-mass estimator detected only two peaks, for the other peaks the excess-mass was too small to be statistically significant. The result of this fit is shown by the continuous line in Fig. 4. The dotted lines indicate the components of this particular analysis.

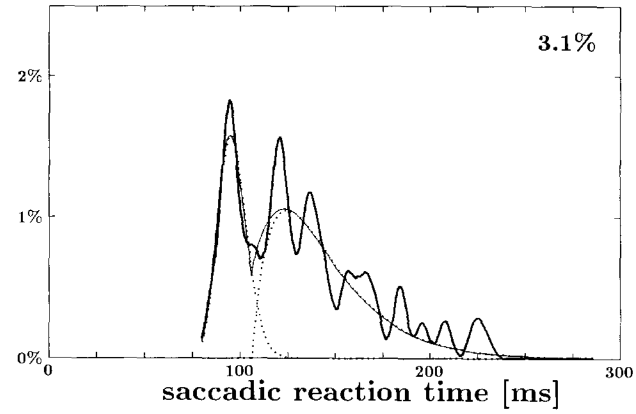

FIGURE 4. Example of data (thick line). A problematic case was selected to show that not all apparent peaks are statistically significant from the excess-mass estimator. The thin line is the superposition of the two components (dashed lined) found by the excess-mass estimator and fitted by two gamma distributions (see text). The weight of the first gamma distribution is 0.35, their parameters are $\mu = 96$ msec, $\sigma = 7.5$ msec and $s = 0.35$. The other gamma distribution has the weight 0.65 and the parameters $\mu = 145$ msec, $\sigma = 28$ msec and $s = 1.47$.

RESULTS

Distribution of peaks

The excess-mass estimator was applied to analyze the data obtained in gap and in overlap tasks. The data from trained subjects and naïve subjects are treated separately.

Figure 5 shows the results from the trained subjects in the gap and overlap tasks. were was data available from gap durations 0, 100, 200, 300 and 400 msec from 10 trained subjects. Because the distribution of the number of peaks for the 181 experiments with gap duration 200 msec are very similar to the other, we present the results for all gap durations. The differences in the distribution of the position of the peaks are described below.

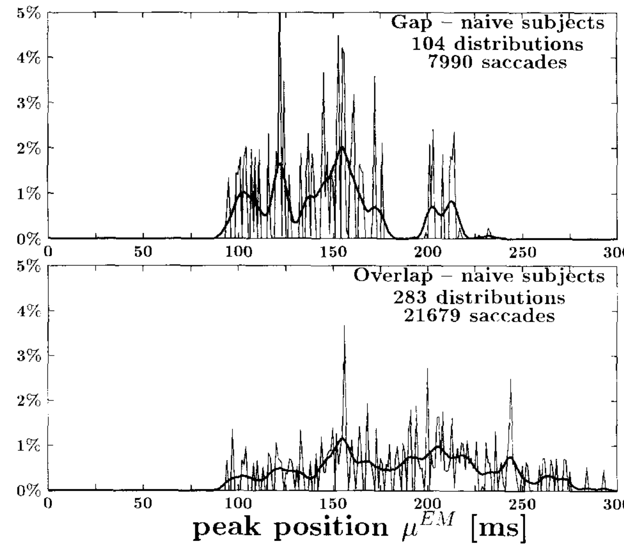

FIGURE 6. Same as Fig. 5 but data from the naïve subjects.

Within the 435 distributions analyzed, 692 peaks were found in the gap data. The percentage of peaks is plotted against the time of their position along the horizontal. The smooth curve was obtained using a gaussian kernel with a bandwidth of 3 msec. It estimates the probability that a distribution exhibits a peak at the position μ^{EM}. (For an overview of kernel estimators see Silverman, 1986).

The upper part depicts the result for gap data, the lower for overlap data. There are two regions where peaks are

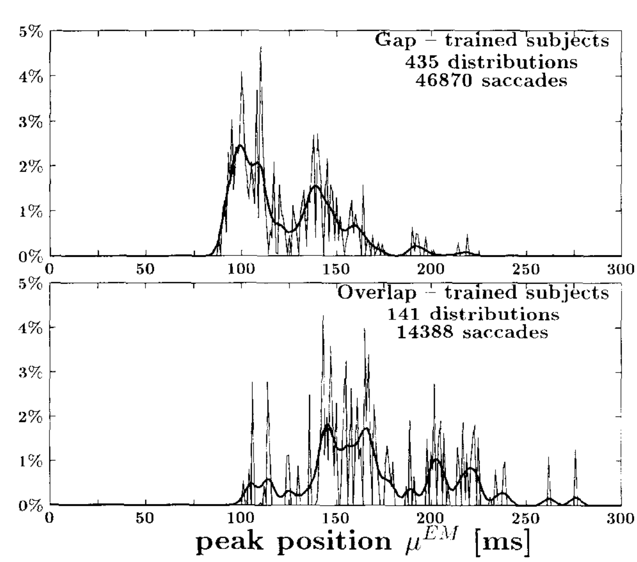

FIGURE 5. The relative frequency of detected peaks is plotted against the corresponding position along the reaction time axis. The gap and overlap data from trained subjects are analyzed. The thick lines result when smoothing the data (thin lines) with a gaussian kernel of bandwidth 5 msec.

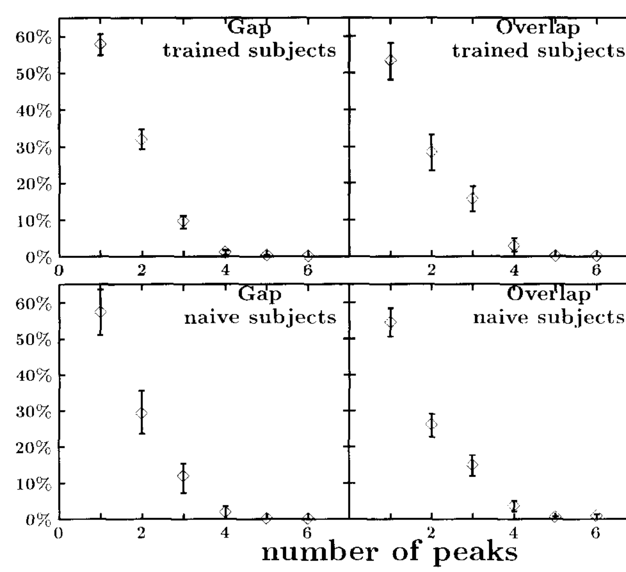

FIGURE 7. The distribution of the number of detected peaks for the four sets of data. In 50–60% of cases only one peak was detected in every set. The frequency of detecting more than three peaks is less than 4%. Note that when only one peak was detected, this did not imply that it was at the same position as the others (see text).

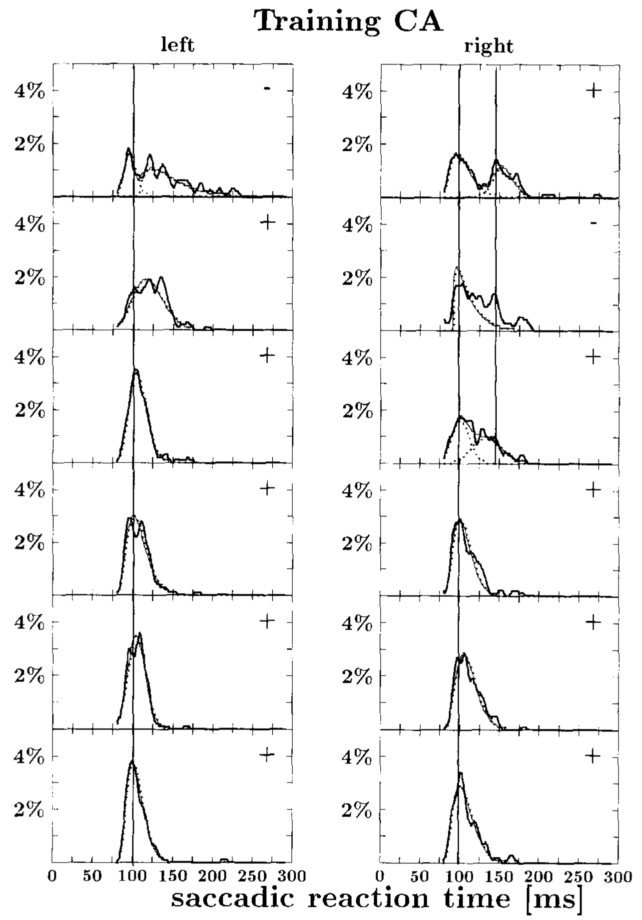

FIGURE 8. SRT-distributions obtained from a single subject (CA) in consecutive training sessions (200 saccades randomly to the right and left). The thick lines represent the experimental data, the dotted lines show the components as determined by the excess-mass estimator and fitted by the Levenberg–Marquardt procedure. "+" indicates a good fit ($P \geq 0.05$) and "−" a bad fit ($P < 0.05$). The vertical lines mark the average position of the first (and second) peaks that are present in several sessions. Note that an express mode was detected together with a second mode in the very first session. At the end of training, only the express mode can be seen.

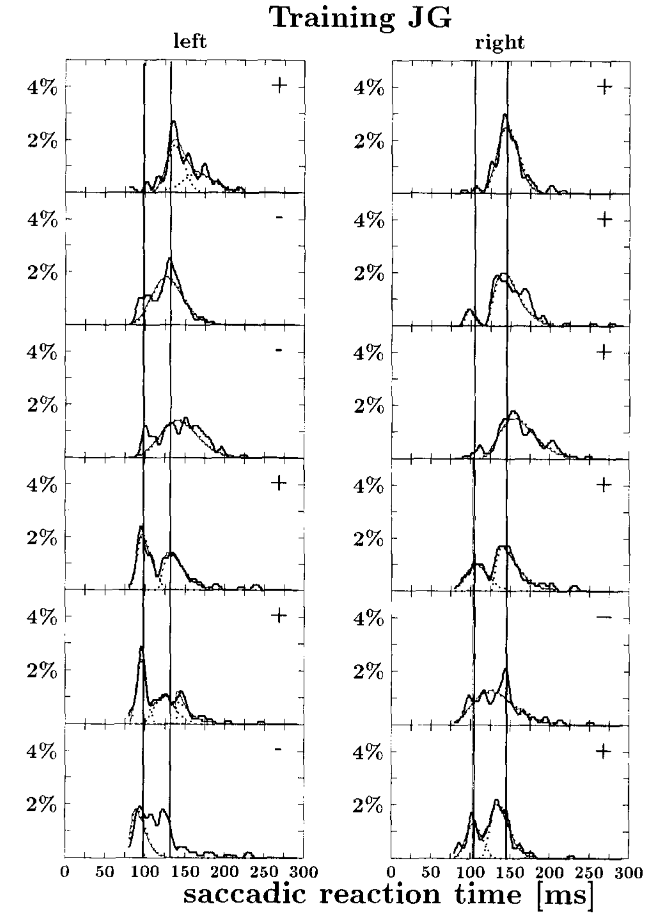

FIGURE 9. Same format as Fig. 8 but for subject JG. An express mode together with a second mode was not detected before the second session. Note that in a number of cases, either one or the other mode remained undetected even though it can be seen and identified by comparing the distributions with others from the same subject.

found with low probability in both cases: between 120 and 135 msec, and between 170 and 190 msec. Regions with high peak density are 90–120 msec (express mode) and 135–170 msec (fast regular mode) in the gap task. The data from the overlap are not so clear. An increased incidence of peaks around 135–170 msec (fast regular mode) and 200–220 msec (slow regular mode) can be seen, which may correspond to the regions obtained from the gap data. Notice that in the overlap data there are also modes around 100 msec.

The results from the naïve subjects are shown in Fig. 6. In this case the gap duration was always 200 msec. As for the trained subjects, detection of peaks around 100 msec is more common on gap than on overlap trials, but there are also peaks in the region of 125 msec. A second region of high peak density is found just above 150 msec and above 200 msec, most clearly in gap and much less clear in overlap trials. In the overlap task there is a large number of peaks between 230 and 300 msec. In this region we found no peak in the gap task and only a very small number of peaks in the overlap task for trained subjects.

Number of peaks

The incidence of multimodality in these data was accessed by plotting the percentage of distributions as a function of the number of peaks found (Fig. 7). Regardless of where the peaks were detected along the time scale, only one peak was detected in 55% of all distributions analyzed so far. Three peaks or more were detected rarely, and almost never in gap trials. This result does not differ greatly between subject groups and tasks. The number of peaks varies only slightly between gap and overlap tasks, because in the gap task an express mode often occurs and a slow regular mode rarely occurs, while in the overlap task a slow regular mode is frequently observed while an express mode is rarely observed.

Analyzing the peak position of the unimodal distributions for trained subjects in the different gap tasks reveals that the mode about 150 msec is most frequently present for the gap 0 msec paradigm (29 of 39 unimodal

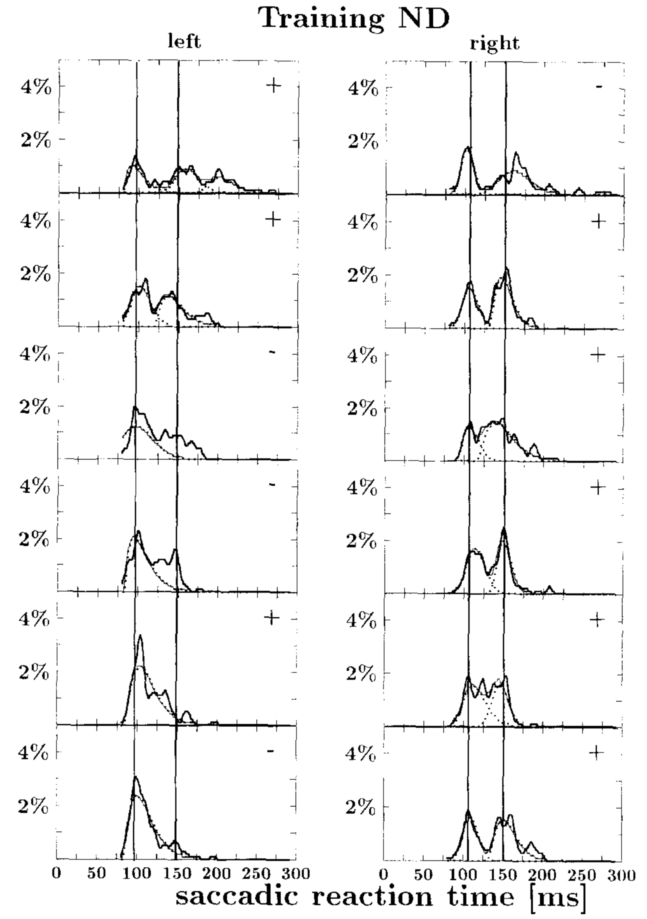

FIGURE 10. Same format as Fig. 8 but for subject ND. As for subject CA an express mode was identified together with a second mode (rightward saccades) and a third mode (leftward saccades) in the first session.

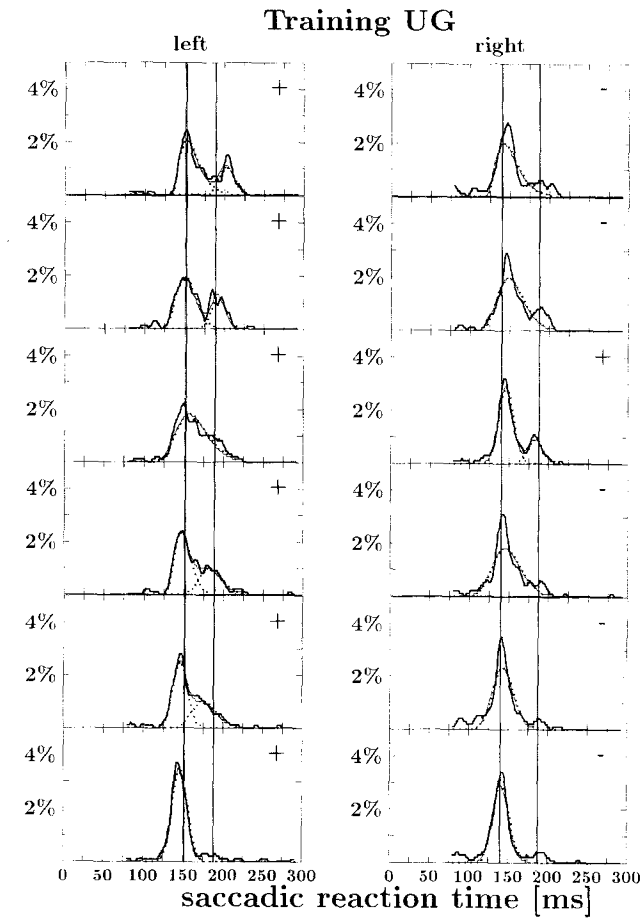

FIGURE 11. Same format as Fig. 8 but for subject UG. This subject started without an express mode and never produced one. Instead, two additional modes were detected: one around 150 msec and the second around 185 msec, predominantly for leftward saccades. Note that this mode is present in all distributions for the rightward saccades, but was detected only in one case. This demonstrates how conservative the excess-mass estimator is.

distributions have a peak position between 135 and 170 msec), while the mode about 100 msec is most frequently present in the gap 200 msec paradigm (78 of 109 unimodal distributions have a peak position between 90 and 120 msec).

Among the unimodal distributions 29% had peaks in the range of the express mode and 46% had peaks in the range of the fast regular mode. Therefore 87% (45% + 75% × 55%) of the data sets support the notion of saccadic reaction time distributions being the superposition of three modes.

Individual data from subjects under training

While the analysis has so far concentrated on data from either trained or naïve subjects, we also applied the peak detection method and the fit procedure to the data from subjects under training. Figures 8–11 show the data and the fitted curves for leftward and rightward saccades separately from six training sessions obtained from four different subjects. The top two panels were obtained when the subjects were still naïve (first session). The number of peaks fitted to the data is the number of peaks estimated by the excess-mass method. We have not presented all parameters of the gamma distributions, because the distributions of one figure contain more than 50 parameters and they are not necessary to detect the agreement of peak occurrence and position of the same subject in different sessions. The goodness of fit is indicated in each plot using the significance level 0.05. A "+" indicates a good fit ($P \geq 0.05$) and a "−" a bad fit ($P < 0.05$).

For subject CA (Fig. 8) two modes were encountered in only three sessions. Interestingly, two were obtained in the very first session. In the second session, two modes may still be present but did not reach significance. In the third session only the data from rightward saccades exhibited two modes. The training ends with unimodal distributions for either side peaking at 100 msec.

Subject JG (Fig. 9) started with a bimodal distribution for leftward saccades with the first mode below 150 msec and the second at about 170 msec. Rightward saccades exhibited only one mode at about the same position as the first mode to the left. In the second session an early mode can be seen on both sides but significance was reached only for the right side. The training ended with a bimodal distribution to the right side with modes at 100 and

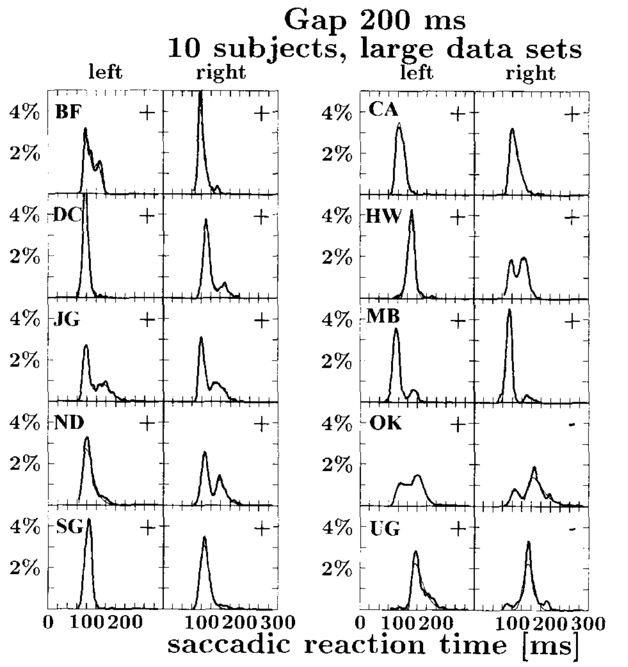

FIGURE 12. SRT-distributions from data files containing more than 280 observations. The leftward and rightward saccades from each of the 10 subjects are shown separately. The thick lines represent the experimental data, the thin lines show the fit. "+" indicates a good fit ($P \geq 0.05$) and "−" a bad fit ($P \geq 0.05$).

130 msec, while for the left side the second mode being detected in previous sessions failed to reach significance in the final session even though these distributions were not statistically different in the Kolmogorov–Smirnov test. The example of subject JG shows how conservative the excess-mass estimator works, rejecting peaks even when they can be seen at positions where they have been found in other distributions from the same subject.

Subject ND (Fig. 10) produced bimodal distributions to the right side throughout the training with the modes remaining at about the same positions, approx. 100 and 150 msec. To the left an early mode around 100 msec was detected throughout the training, but the second mode is not always identified. This subject started with a trimodal distribution for leftward saccades.

Subject UG (Fig. 11) never produced an express mode. However, bimodality was found in five cases. Note that the right mode just below 200 msec was detected by the excess-mass estimator only once, even though it can be seen in all distributions. Only on the left side is the bimodal distribution of fast and slow regular saccades turned into a unimodal fast regular peak throughout the time course of training.

These data suggest that all distributions are produced bi- or trimodally, but in some cases the weight of one mode is too small and/or its separation from the neighboring mode is too small to allow for clear separation using the excess-mass method. The shape of a mode of a SRT-distribution may change during training, but the peak position remains at about the same position.

Analysis of large data sets from individual subjects

The statistical quality of data derived from stochastic processes can be improved by increasing the number of observations, given that the process is stationary. Stationarity of saccadic reaction time data can be obtained by training the subjects. Therefore, in this section, we consider distributions from trained subjects, when they performed different gap tasks after having been trained in the gap 200 msec task until they reached stable results. Single sets of data collected in one experimental session were put together only after they had been tested to be not statistically different by the Kolmogorov–Smirnov test (significance level 0.05). Ten subjects performed the gap 200 msec task several times (3–8) after their training was completed, such that a minimum of 280 observations could be collected for each side of random target presentation. The distributions were then constructed separately for each side using only the correct responses, (i.e., saccades going in the direction of the target). This way the 20 sets of data presented in Fig. 12 were obtained.

The analysis described above was applied to each of the distributions. The number of fitted gamma distributions was estimated by the excess-mass procedure. The resultant fits are shown by the thin lines in the figure. Very often the thin line is hidden behind the thick line, indicating a perfect fit. The goodness of fit is shown in each plot using the significance level 0.05, the "+" indicates a good fit ($P \geq 0.05$) and a "−" a bad fit ($P < 0.05$). The quality of fits was better than the quality of the fits of small samples.

Inspection of these distributions shows: (i) all but three distributions (HW-left, UG-left and right) show a first mode at about 100 msec. (The express mode of the subject HW could, however, be elicited with the left target being presented at 8 instead of 4 deg). (ii) The maxima of the first mode are almost, but not exactly, at the same position. Pooling the data across sides and subjects could have destroyed the separation of peaks. (iii) Bimodality is clearly present in 10 sets, unimodality in the other ones. (iv) In the cases of only one peak it is located at about the same positions where the first or the second peak in bimodal distributions can be seen.

These data again suggest that all distributions are produced bi- or trimodally but in some cases the weight of one mode is too small and/or its separation from the neighboring mode is too small to allow for clear separation using the excess-mass method.

To support this view, we analyzed larger data files from three subjects using different gap durations between 0 and 400 msec. Figure 13 shows the data and the fits. Leftward and rightward saccades were required in these sessions in random order. The subject OK produced bi- or trimodal distributions for gap durations between 100 and 400 msec, where the express peak occurs. For gap duration 0 msec there was no indication for an express peak and only one mode at about 150 msec can be seen similar to subjects BF and DC. At a gap of 100 msec this mode became rather weak and the express mode can be

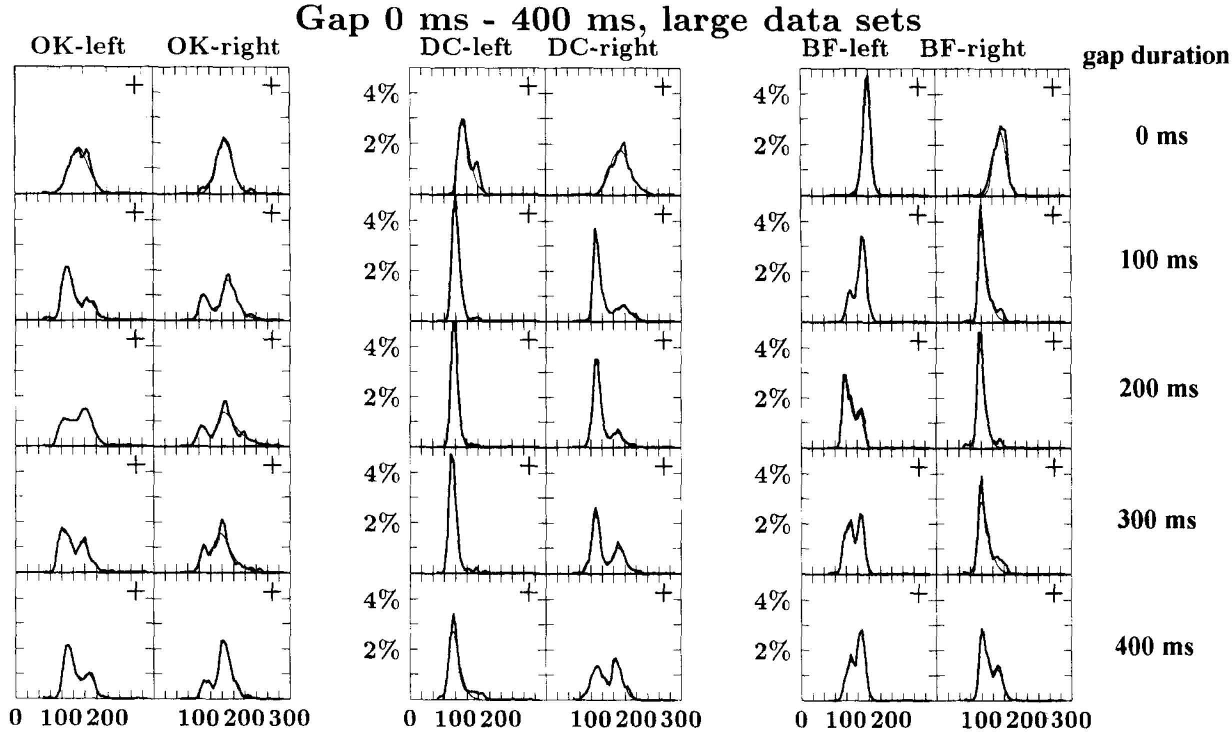

FIGURE 13. SRT-distributions for subjects OK, DC and BF containing more than 320 observations each. From top to bottom the gap duration was varied between zero and 400 msec as indicated on the right. Thick lines represent the experimental data, thin lines show the fit. "+" indicates a good fit ($P \geq 0.05$).

seen. For gap durations longer than 200 msec the express mode becomes weaker again in favor of the second mode. Note that within a few milliseconds the modes tend to remain at the same positions.

Altogether, we had 70 data sets from 10 subjects, each containing more than 280 data points. To overcome the problem of small data sets with correspondingly high errors we applied the excess-mass procedure to these data sets. The distributions shown in Figs. 12 and 13 are part of this set.

As a result, in the 70 distributions one peak was found in 57% of all cases, two peaks in 34%, and three peaks in 9%. These numbers closely resemble the results obtained from the data sets containing only 75–100 entries (see Methods section). Again among the unimodal distributions the positions of the modes were not uniformly distributed but preferentially around 100 and 150 msec.

DISCUSSION

The present study has shown for the first time that statistical analysis reveals the existence of several modes in saccadic reaction distributions. In about 45% of all distributions analyzed more than one mode was detected and in about 15% three modes were encountered. The procedure for detecting these modes appeared as rather conservative and in many cases where the distributions clearly deviated from unimodality the second mode remained non-significant due to the strong criterion used.

The fact that 55% of the analyzed distributions exhibit only one mode cannot be used as evidence for the conclusion that only half of the subjects produce multimodality. The detailed analysis of the position of the peaks in the unimodal distributions rather support the notion that the unimodal distributions are extreme cases where only one of several modes dominates the distribution. This can be seen in Figs. 8–11, showing the development of the peaks during the training sessions. A consideration of the peak position of the unimodal distributions leads to the conclusion that 87% of the data sets support the notion of saccadic reaction time distributions being the superposition of three modes.

The results support the point of view that a SRT-distribution is a superposition of three underlying distributions (basis functions) that occur with different weights in different experiments. Different subjects have different sets of basis functions, but each subject has a basis function in the express range, one in the fast regular range and one in the slow regular range.

While in this article the number of gamma distributions fitted to the data is determined by the excess-mass estimator, an alternative method would have been to start with $n = 1$ gamma distributions and increase n until the probability P_n, indicating the goodness of fit, exceeds a critical value. It is even possible to test whether the improvement in the error χ^2 is worth the added number of parameters [see Chapter 5.8 of Flury & Riedwyl (1988) for details]. Since a bad fit results, if the assumptions about the number of modes *or* the distribution of the modes is inappropriate, the results for the distribution of the number of modes are more contestable when using a multimodality test, where no assumptions are necessary about the shape of the distribution of the data.

The fit procedure applied to the data using the previously detected number and positions of the peaks gave satisfactory results in 64% of cases. Poor fits were mostly due to a failure in identifying deviations from unimodality as different modes. We fitted gamma distributions to the data. Using different functions may have resulted in similar fit qualities because the limited number of observations will not allow, in all cases, minimization of the error more for one function than for another.

Another approach is to analyze approx. 10 different distributions of the same subject at the same time. The idea is that a mode is present where neighboring bins scaled commonly. Contrary to the excess-mass procedure it is possible with this test to detect different modes, even if their superposition is unimodal (Gezeck & Timmer, 1995). The preliminary results have shown that in most subjects three basis functions are detected in accordance with the present study.

The experimental data and the statistical analysis presented above suggest that the human (and probably also the monkey) saccade system reacts in different modes. They act in serial, each one adding its time consumption to the reaction time. Each mode then can add 20–50 msec, an order of magnitude that can be accounted for by neural summation time. Which modes determine a given SRT-distribution depend on the state of activity of the optomotor system (probably at the brain stem level) at the time of target appearance, and it is this state of activity which is determined by the task. This notion circumvents the necessity of selecting one or the other mode in a mutually exclusive way.

We searched for models of saccade generation which can explain multimodal distributions of the reaction times. The facilitation model of Reulen (1984) can explain the gap effect as far as the decrease of the mean value of the reaction time is concerned, but it cannot produce multimodal distributions. Similarly, Carpenter's model relies on unimodal distributions from the start (Carpenter & Williams, 1995). Analysis of his data, all obtained from overlap trials, shows that the model predictions are not fulfilled: instead of a single straight line the data form two sections with an "elbow" at about 140 msec. This indicates that at least two modes can be identified in the data and that the second mode begins at about 140 msec. This is in close agreement with the present analysis because the transition between the express and the fast regular mode occurs between 130 and 140 msec. According to Ruhnau and Haase synchronized oscillations in the visual cortex produce multimodal distributions (Ruhnau & Haase, 1993; Kirschfeld *et al.*, 1996).

The early version of the three-loop-model assumes serial processing of three central stages but requires the additional assumption of a probabilistic variable which determines the chances of the optomotor system being in one or the other state at the time of target presentation (Rogal & Fischer, 1986). More recently, a two-stage serial model was presented which produces bimodal distributions and also takes into account the sensory factors such as stimulus intensity and visual-auditory integration in saccade generation (Nozawa *et al.*, 1994). When the processes were implemented as assemblages of interacting neurons, the probabilistic aspects of being in one or the other stage are provided "automatically" in the form of the stochastic nature of impulse trains in the sensory channels representing the effect of stimulus onset and fixation point offset (Fischer *et al.*, 1995).

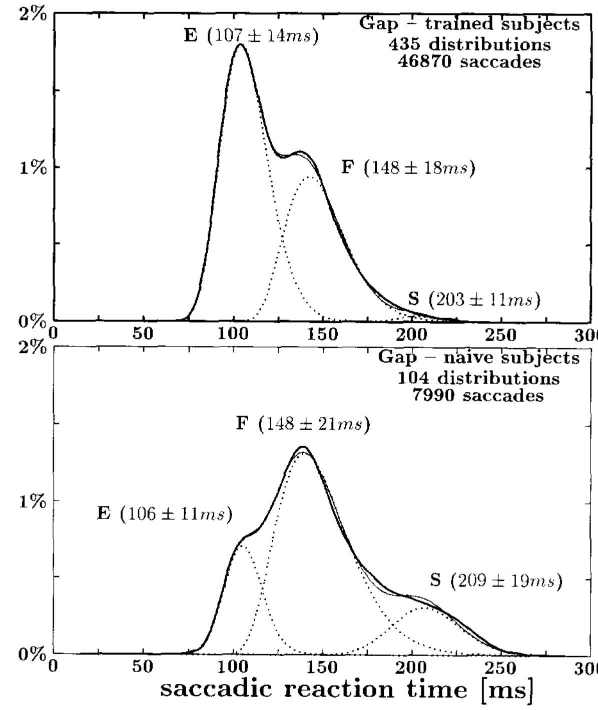

FIGURE 14. Superposition of all fitted curves of the gap data sets from the trained (upper part) and the naïve subjects (lower part). The express peak is marked "E", the fast- (slow-) regular peak is marked "F" ("S"). The mean and standard deviation of the modes are given in parentheses.

A number of authors have published SRT-distributions that are bimodal beyond any doubt and without the need for statistical approval. Examples can be found for normal human subjects in Jüttner & Wolf (1992), Fischer *et al.* (1993), Weber *et al.* (1995), Nozawa *et al.* (1994), Currie *et al.* (1993), Cavegn (1993), Reuter-Lorenz *et al.* (1995), Walther-Müller & Znoj (1994), for patients in Matsue *et al.* (1994), Braun *et al.* (1992), and for monkeys in Fischer & Boch (1983), Munoz & Wurtz (1992), Sommer & Schiller (1992) and Schiller & Lee (1994). These results support the point of view that SRT-distributions are generally the result of a superposition of a finite number of modes. From this perspective a unimodal distribution is a distribution, where only one mode is present and it may be instructive to investigate which mode is active. Extreme cases can be seen in the express saccade makers. These are subjects who produce almost exclusively express saccades even in the overlap task (Biscaldi *et al.*, 1996). Only the express modes seem to be activated because these subjects may not have enough control over their fixation system and therefore any stimulus activates more or less directly the saccade-generating system at a relatively low level leading to a saccade immediately. Other subjects may have extremely good control over their fixation system thereby inhibiting the saccade system such that the express mode has no chance to determine the reaction time alone.

In conclusion, the experimental data and their statistical analysis suggest that the assumption that discrete modes and their superpositions form the distribution of saccadic reaction times appears to be more attractive than the assumption of only one mode. To gain an impression of an average SRT-distribution of trained and naïve subjects in gap condition, Fig. 14 presents the collapsed data. Quite clearly these data (heavy lines) will not be recognized as bi- or trimodal, neither do they form normal distributions. When fitted by three gamma distributions (dotted lines) almost perfect fits (thin lines) are obtained. Note that the three modes have peaks at about the same positions as naïve and trained subjects.

REFERENCES

Biscaldi, M., Fischer, B. & Stuhr, V. (1996). Human express-saccade makers are impaired at suppressing visually-evoked saccades. *Journal of Neurophysiology, 76*, 199–214.

Braun, D., Weber, H., Mergner, T. & Schulte-Mönting, J. (1992). Saccadic reaction times in patients with frontal and parietal lesions. *Brain, 115*, 1359–1386.

Carpenter, R. & Williams, M. (1995). Neural computation of log likelihood in control of saccadic eye movements. *Nature, 377*, 59–62.

Cavegn, D. (1993). Attention engagement and the generation of express saccades, Proceedings of Seventh European Conference on Eye Movements, (Abstracts), 21.

Chambers, J., Cleveland, W., Kleiner, B. & Tukey, P. (1983). *Graphical methods for data analysis*. Belmont: The Wadsworth Statistics.

Currie, J., Joyce, S., Maruff, P., Ramsden, B., Mc Arthgur-Jackson, C. & Malone, V. (1993). Selective impairment of express saccade generation in patients with schizophrenia. *Experimental Brain Research, 97*, 343–348.

Efron, B. & Tibshirani, R. (1991). *An introduction to the bootstrap*. London: Chapman and Hall.

Fischer, B. (1987). The preparation of visually guided saccades. *Review of Physiology and Biochemical Pharmacology, 106*, 1–35.

Fischer, B. & Boch, R. (1983). Saccadic eye movements after extremely short reaction times in the monkey. *Brain Research, 260*, 21–26.

Fischer, B., Boch, R. & Ramsperger, E. (1984). Express-saccades of the monkey: effect of daily practice on probability of occurrence and reaction time. *Experimental Brain Research, 55*, 232–242.

Fischer, B., Gezeck, S. & Huber, W. (1995). The three-loop-model: a neural network for the generation of saccadic reaction times. *Biological Cybernetics, 72*, 185–196.

Fischer, B. & Ramsperger, E. (1984). Human express saccades: extremely short reaction times of goal directed eye movements. *Experimental Brain Research, 57*, 191–195.

Fischer, B. & Weber, H. (1993). Vision and visual attention. *Behavioral and Brain Sciences, 16*, 553–610.

Fischer, B., Weber, H., Biscaldi, M., Aiple, F., Otto, P. & Stuhr, V. (1993). Separate populations of visually guided saccades in humans: reaction times and amplitudes. *Experimental Brain Research, 92*, 528–541.

Flury, B. & Riedwyl, H. (1988). *Multivariate statistics*. London: Chapman and Hall.

Gezeck, S. & Timmer, J. (1995). Detecting multimodality in saccadic reaction time distributions. Proceedings of Eighth European Conference on Eye Movements (Abstracts), pp. 66–67.

Honerkamp, J. (1994). *Stochastic dynamical systems*. Weinheim: VCH.

Jüttner, M. & Wolf, M. (1992). Occurrence of human express saccades depends on stimulus uncertainty and stimulus sequence. *Experimental Brain Research, 89*, 678–681.

Kingstone, A. & Klein, R. (1993). What are human express saccades? *Perception and Psychophysics, 54*, 260–273.

Kirschfeld, F., Feiler, R. & Wolf-Oberhollenzer, F. (1996). Cortical oscillations and the origin of express saccades. *Proceedings of the Royal Society of London, 263*, 459–468.

Matsue, Y., Osakabe, K., Saito, H., Goto, Y., Ueno, T. & Matsuoka, H. (1994). Smooth pursuit eye movements and express saccades in schizophrenic patients. *Schizophrenia Research, 12*, 121–130.

Matthews, M. (1983). On Silverman's test for the number of modes in a univariate density function. B.A. Honor thesis, Harvard University, Dept. of Statistics.

Müller, D. & Sawitzki, G. (1991). Excess mass estimates and tests for multimodality. *Journal of the American Statistical Association, 86*, 738–746.

Munoz, D. & Wurtz, R. (1992). Role of the rostral superior colliculus in active visual fixation and execution of express saccades. *Journal of Neurophysiology, 67*, 1000–1002.

Nothdurft, H. & Parlitz, D. (1993). Absence of express saccades to texture or motion defined targets. *Vision Research, 33*, 1367–1383.

Nozawa, G., Reuter-Lorenz, P. & Hughes, H. (1994). Parallel and serial processes in the human oculomotor system: bimodal integration and express saccades. *Biological Cybernetics, 72*, 19–34.

Press, W., Flannery, B., Teukolsky, S. & Vetterling, W. (1986). *Numerical recipes*. Cambridge: Cambridge University Press.

Priori, A., Dertolasi, L., Rothwell, J., Day, B. & Marsden, C. (1993). Some saccadic eye movements can be delayed by transcranial magnetic stimulation of the cerebral cortex in man. *Brain, 116*, 355–367.

Reulen, J. (1984). Latency of visually evoked saccadic eye movements. 1. Saccadic latency and the facilitation model. *Biological Cybernetics, 50*, 251–262.

Reuter-Lorenz, P., Hughes, H. & Fendrich, R. (1991). The reduction of saccadic latency by prior offset of the fixation point: an analysis of the gap effect. *Perception and Psychophysics, 49*, 167–175.

Reuter-Lorenz, P., Oonk, H., Barnes, L. & Hughes, H. (1995). Effects of warning signals and fixation point offsets on the latencies of pro-versus anti-saccades: implications for an interpretation of the gap effect. *Experimental Brain Research, 103*, 287–293.

Rogal, L. & Fischer, B. (1986). Eye–hand coordination: a model for computing reaction times in a visual reach task. *Biological Cybernetics, 55*, 263–273.

Rohrer, W. & Sparks, D. (1993). Express saccades: the effects of spatial and temporal uncertainty. *Vision Research, 33*, 2447–2460.

Ruhnau, E. & Haase, V. (1993). Parallel distributed processing and integration by oscillations. *Behavioral and Brain Sciences, 16*, 587–588.

Sachs, L. (1982). *Applied statistics*. Berlin: Springer.

Saslow, M. (1967). Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *Journal of the Optical Society of America, 57*, 1022–1029.

Schiller, P. & Lee, K. (1994). The effects of lateral geniculate nucleus, area V4, and middle temporal (MT) lesions on visually guided eye movements. *Vision Neuroscience, 11*, 229–241.

Sereno, A. & Holzman, P. (1993). Express saccades and smooth pursuit eye movement function in schizophrenic, affective disorder, and normal subjects. *Journal of Cognitive Neuroscience, 5*, 303–316.

Silverman, B. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society, Series B, 43*, 97–99.

Silverman, B. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.

Sommer, M. & Schiller, P. (1992). Express saccades elicited during natural fixations of visual search. *Society of Neuroscience Abstracts, 18*, 1398.

Tam, W. & Ono, H. (1994). Fixation disengagement and eye-movement latency. *Perception and Psychophysics, 56*, 251–260.

Walther-Müller, P. & Znoj, H. (1994). Saccadic reaction times revisited: the rabbit behind the tree. *Swiss Journal of Psychology, 53*, 151–165.

Weber, H., Biscaldi, M. & Fischer, B. (1995). Intertrial effects of randomization on saccadic reaction times in human observers. *Vision Research, 35*, 2615–2642.

Weber, H. & Fischer, B. (1995). Gap duration and location of attention focus modulate the occurrence of left/right asymmetries in the saccadic reaction times of human subjects. *Vision Research, 35*, 987–998.

Wenban-Smith, M. & Findlay, J. (1991). Express saccades: is there a separate population in humans? *Experimental Brain Research, 87*, 218–222.

Wong, M. (1985). A bootstrap testing procedure for investigating the number of subpopulations. *Journal of Statistical Computatation & Simulation, 22*, 99–112.

Acknowledgements—The help of Dr Heike Weber and Kerstin Dörflinger in data collection and Dipl. Phys. Klaus Hartnegg and Dr Franz Aiple in providing computer programs for stimulus presentation, data collection, and data evaluation is gratefully acknowledged. The critical comments of Dr Heike Weber and Dipl. Phys. Volker Stuhr have been very helpful. This work was supported by the Deutsche Forschungsgemeinschaft (Fi227/10-1;DFG).