Bioelectromagnetics 19:310–317 (1996)

Acute Effects of 50 Hz, 100 μT Magnetic Field Exposure on Visual Duration Discrimination at Two Different Times of the Day

Nikolaos Kazantzis, John Podd,* and Craig Whittington

Department of Psychology, Massey University, Palmerston North, New Zealand

A two-alternative, forced-choice visual duration discrimination task was used to examine the effect of an intermittent, 50 Hz, 100 μT magnetic field on accuracy at two different times of the day. A total of 59 female and 40 male subjects with an age range of 18 to 46 years were studied under both field-exposed and sham-exposed conditions. The subject's task was to decide which of two sequentially presented light flashes had the longer duration, percentage correct being the measure of performance. The data were gathered under double-blind conditions with sham and real exposure counterbalanced. Exposure to the magnetic field produced a small improvement in accuracy but only at the most difficult level of the task, with female subjects showing a larger improvement than males. The time of day at which the study was run had no effect on performance. Despite the relatively large number of subjects used and a relaxed alpha level (P = .3), the statistical power of the test to detect the observed effect was still only 0.71. Bioelectromagnetics 19:310–317, 1998.

Key words: extremely low frequency; statistical power; human performance; visual discrimination

INTRODUCTION

A recent study [Whittington et al., 1996] of the acute effects of an intermittent, 50 Hz, 100 μT sinusoidal magnetic field (MF) on human reaction time (RT) and accuracy in a visual duration discrimination task showed that RT was faster in the presence of the MF, but only for the most difficult condition of the task (effect size = 0.20, P = .04). Accuracy, measured as percentage correct decisions (PC), was unaffected. They reported no evidence that the speeded RTs were due to a speed-accuracy trade-off. Speed-accuracy trade-offs are always a concern with the measurement of RT because speeded responses may come at the expense of accuracy, rather because of the experimental manipulation hypothesised to change RT. Similarly, accuracy might be increased simply by taking more time to respond.

The results obtained by Whittington et al. [Table 1 in Whittington et al., 1996] show that the faster RTs at the most difficult level of the task were not accompanied by a decrease in PC. However, calculating the association between RT and PC across all conditions of their study yields a correlation of −0.2, suggesting that these two performance measures might not have been completely independent. Furthermore, the negative sign indicates that RT may have improved at the expense of accuracy. It is acknowledged that this association between RT and PC is small, but it is our belief that the ELF MF effects we are attempting to verify are also small [Podd et al., 1995; Whittington et al., 1996]. Thus, even a weak association between RT and PC might have been sufficient to mask any MF effects on PC. The main purpose of the present investigation was to determine whether the MF used by Whittington et al. [1996] has an effect on accuracy when speeded responses are not an accompanying task requirement.

In addition, to remove the potentially confounding effect of RT on accuracy, we attempted to improve the sensitivity of the study in other ways. One factor that might compromise MF effects on performance is

*Correspondence to: Dr. John Podd, Department of Psychology, Private Bag 11222, Palmerston North, New Zealand. E-mail: j.v.podd@massey.ac.nz

Received for review 5 December 1996; final revision received 26 January 1998

the time of day that the study is conducted. For instance, several investigations have shown that RT is influenced by the time of day at which it is measured [Colquhoun, 1971; Johnson et al., 1981; Moore-Ede et al., 1982]. However, time-of-day effects on accuracy measures are less clear-cut. For example, Colquhoun et al. [1968] found that PC in a vigilance task varied by about 5% across two times of the day, times similar to those used in the present study (0900 and 1600 h). On the other hand, Craig [1979] found no effect for time-of-day in a similar task. To the best of our knowledge, only one human performance study has examined time-of-day in the presence of an ELF MF [Graham and Cohen, 1985], finding no effects. However, that study was carried out in the presence of an electric field, as well as a MF.

Usually, any potential variations in performance due to time-of-day are controlled for by running all subjects at the same time each day. However, because of circadian fluctuations in hormonal and neurophysiologic processes [Moore-Ede et al., 1982; Arendt et al., 1989], it is possible that humans are more or less susceptible to MF effects at different times of the day. Our interest in time-of-day was motivated primarily out of a concern to maximise the sensitivity of our subjects to the applied MF, thus, in effect, increasing statistical power. Therefore, half of our subjects were tested in the early morning and half in the late afternoon. In addition, we took into account whether subjects reported working best in the morning or in the evening. Subject preference for morning or evening activity was evaluated using the Morningness Questionnaire [Smith et al., 1989]. The scores on the questionnaire were entered into the main analysis as a covariate.

A factor that might influence the conclusions drawn from any experimental investigation is insufficient statistical power. Statistical power can be thought of as the probability of detecting an experimental effect. As statistical power decreases, the likelihood of obtaining a null result (concluding that there is no experimental effect) increases [Cohen, 1988]. When statistical power is low, the interpretation of a null result is decidedly problematic; it might mean that there is no experimental effect, the usual conclusion drawn. But another possibility (rarely considered) is that the investigation was not powerful enough to detect the effect. Based on previous findings [Whittington et al., 1996], we had good reason to assume that any MF effect on PC would be small.

Whittington and Podd [1996] have shown that for small effect sizes, running a study with the typical number of subjects (about 20 per condition) and using the conventional significance level of .05, yields a statistical power of less than 10%. In other words, the chance of *failing* to detect a small but real MF effect (a type II error) is about 90%, a totally unacceptable level of risk. It seems absurd to run a study that is almost certain to fail to detect the hypothesised effect, even when it is a real effect. Therefore, we increased statistical power by increasing the number of subjects ($n = 99$) and relaxing the alpha level from the conventional $P = .05$ to $P = .30$. With these values, we estimated that our statistical tests would have a power of about 0.70 and a type II error probability of 0.30 ($1 -$ power) [Cohen, 1988]. Thus, in the present study, the probability of a type II error (erroneously concluding there is no MF effect) was roughly equal to the probability of a type I error (erroneously concluding a real MF effect exists).

Our rationale for these adjustments was twofold. First, the typical type II error probability of around 0.90 for small effect sizes is far too high. Under such circumstances, it is very unlikely that a small, but nevertheless real, MF effect will be detected, or replicated. Second, in exploratory research, in which one is trying to establish the existence of MF effects on human performance, however small those effects may be, type II errors are just as serious as type I errors. In fact, type II errors may be the more serious because obtaining a null result has the nasty habit of forestalling further research into a promising lead. A better approach is to ''set the net wide,'' capturing those findings that *may* indicate MF effects. The critical follow-up procedure is then to subject these individual findings to a meta-analysis that can demonstrate the existence of an effect (or otherwise) with a high degree of certainty (Schmidt, 1996).

In summary, the present study examined the effect of an ELF MF on human accuracy in a visual duration discrimination task identical to that used by Whittington et al. [1996]. Performance data were collected at two different times of the day to check for any temporal variation in MF sensitivity. From our earlier research [Whittington et al., 1996], we knew that the effect size associated with a PC change in the presence of a MF was likely to be very small. Therefore, special attention was paid to improving statistical power and to reducing the normally excessively high type II error rate associated with small effect sizes in human MF studies.

MATERIALS AND METHODS

Subjects

A total of 59 women and 40 men, between 18 and 46 years old, participated. All were students or staff at Massey University who volunteered for the experiment. They were screened using a slightly modified form of the health survey questionnaire used by Cook et al. [1992]. This questionnaire determined whether a subject was free of chronic health problems, cardiovascular problems, had had no recent illnesses, and so on. Any subject answering in the affirmative to one or more of these questions was thanked for volunteering but did not complete the experimental trials. The study protocol, including the health survey questionnaire and standard informed consent procedures, was reviewed and approved by the Massey University Human Ethics Committee.

Experimental Design

A double-blind procedure [Podd et al., 1995] was used in which subjects participated in a single session of approximately 40 min. All subjects completed the visual duration discrimination task at all three levels of difficulty under both real and sham exposure conditions. That is, difficulty level and exposure were within-subjects factors. The between-subjects factor was time of day, with subjects being randomly assigned to the morning (0900 h) or afternoon (1600 h) condition. To control for practice effects and any other carryover effects, real and sham exposure conditions were counterbalanced across the two blocks of experimental trials.

Task and Measures

The experimental task was a two-alternative, forced-choice, visual duration discrimination task with three levels of difficulty. On each trial, a red light-emitting diode (LED) always produced a standard duration stimulus of 50 ms. This stimulus was paired with one of three possible comparison stimuli of 125 ms (easy), 100 ms (intermediate), and 65 ms (hard) durations. This task was identical to that used by Whittington et al. [1996], in which (under sham conditions) the hard task yielded a PC of 61.9%, the intermediate task 83.6%, and the easy task 89.6%.

Two subjective, pencil-and-paper tests were administered after participants had completed the experimental trials. The first was the Field Status Questionnaire [FSQ; Cook et al., 1992], used to evaluate the effectiveness of the double-blind procedure. The main question asked of both the subjects and the experimenter was ''In your judgement, was the field on or off in the first block of trials?'' The second test was the Morningness Questionnaire (MQ) [Smith et al., 1989]. The MQ, used to assess preference for morning or evening activity, is a 13-item scale that has satisfactory internal consistency (Cronbach's alpha = 0.83). It also correlates reasonably well (average, $r = 0.87$) with scores on other Morningness scales [Horne and Ostberg, 1976]. MQ scores (maximal possible range of 10 to 56) were used to classify subjects as evening ($\leq$22), intermediate (23–43), or morning ($\geq$44) types.

Exposure Facility

The exposure facility and associated apparatus were the same as those used by Whittington et al. [1996]. However, there was a major change in regard to the method of exposure to the MF. In the present investigation, we modified some parts of the facility to allow testing of up to four subjects simultaneously. One method of increasing statistical power is to increase the number of subjects. Whittington et al. [1996] found an acceptable level of power (80%) for a moderate effect size using 100 subjects in a within-subjects design. Running such a large number of people through the exposure facility singly is quite time consuming. Therefore, we redesigned the interior of the facility to generate simultaneously four independent, homogeneous MFs using four identical sets of Helmholtz-configured coils. Each coil was wound onto a wooden form using 120 turns of 1.5-mm resin-coated copper wire. The radius of the coils (0.2 m) allowed for an intercoil distance (0.2 m) large enough to accommodate a subject's head.

The four pairs of coils were suspended from a square framework of PVC tubing resting in slots on top of four triangular-shaped, wooden cubicles arranged in the form of a 1.9 m × 1.9 m square. Each coil set was suspended at right angles to adjacent pairs with a 1.15 m separation between adjacent coils in adjacent sets (see Fig. 1). The cubicle walls were high enough to allow ample height adjustment of the coils, and to obscure visual contact between subjects whether seated or standing. Each seat had an adjustable head-rest allowing for accurate positioning of the head between the coils. Once the subject was seated, the LED was positioned at eye level directly in front of the subject at a distance of 0.8 m. The whole exposure apparatus was centred in an unshielded 5.5 m × 4.5 m room.

All measurements of the applied and ambient AC MFs were made using a fourth-generation Hall-effect probe in conjunction with a gaussmeter (model 9200; F.W. Bell, Orlando, FL). The ambient field in the vicinity of the exposure facility was less than 0.6 μT, the sensitivity limit of the gaussmeter. For the experimental trials, the coils were connected in series, and the MF was intermittent (1 s on and 1 s off). The MF was gated on using a zero axis crossing switch. The amplifier was set to generate a 100 μT field at each of the coil sets. This MF strength was verified by produc-

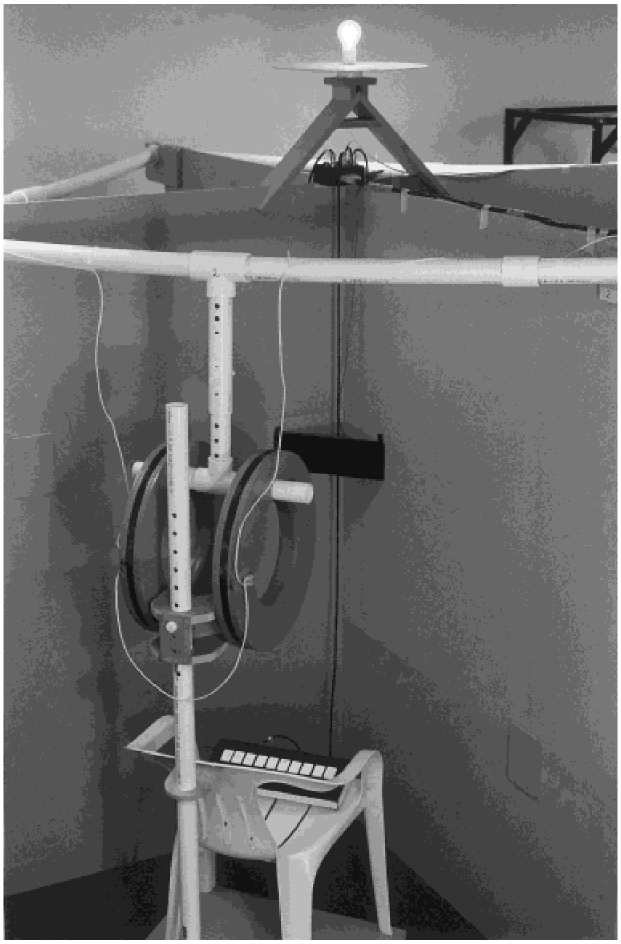

Fig. 1. Photograph of the exposure facility. The facility has four triangularly shaped booths arranged in a square with each coil set positioned at right angles to adjacent sets. The headrest and coils are independently height adjustable.

ing a continuous field simultaneously at all four coil pairs, and then measuring the field with the gaussmeter probe centred between the coils of each pair. The apparatus produced no detectable interference among the coil sets, and no perceptible vibration, sound, or thermal radiation, even after the coils had been continuously energised for several hours.

Field Exposure

The applied field was an intermittent, 50 Hz, 100 μT (root mean square) MF to which subjects were exposed for 7.9 min during the course of the experimental trials. The apparatus used to generate the MF was interfaced with a computer (model 310; Hewlett Packard, Santa Clara, CA), which controlled the trial sequencing and data acquisition. Field exposure was assigned randomly by the computer to either the first or second block of experimental trials. The computer also monitored continuously the output of the amplifier during the experiment. At the beginning of each daily session, the MF flux density was checked at each of the four coil sets.

Trial Sequence

The subject's task on each trial was to decide which of two consecutive LED light flashes had the longer duration. Each trial began with a 100-ms warning tone followed 400 ms later by the first stimulus; the second stimulus occurred after a 500-ms interstimulus interval. A stimulus of standard duration (50 ms) occurred in either the first or second interval, accompanied by one of three comparison stimuli of 65-, 100-, or 125-ms duration. The order of presentation of the standard and comparison stimuli was determined randomly, with the restriction that the standard stimulus should not appear in the same interval for more than three consecutive trials. The response interval was set at a relatively long 2000 ms to ensure that subjects were under no time pressure to respond. Subjects indicated their choice using a two-key response pad, the left-hand button being used when the longer duration stimulus appeared in the first interval and the right-hand button when it appeared in the second interval.

Procedure

Once all subjects were present, they read an information sheet, signed a consent form, and completed the health questionnaire. These tasks were completed in the laboratory to allow adaptation to the dim lighting. After these preliminary procedures, subjects were individually positioned in the exposure apparatus.

Twenty-five practice trials were run to familiarise subjects with the sequence and timing of the trials. After an approximate 30-s break, the first of two blocks of 150 trials were run. There was a 2- to 3-min break before the second block of trials began. Each block contained approximately 50 trials at each difficulty level, presented in random order. Immediately after the second block of trials ended, the MQ was given. Then the subjects completed the FSQ before being debriefed concerning the experimental design.

Statistical Analysis

Overall performance effects of both the between-groups and within-groups factors were analysed using multivariate analysis of variance (MANOVA) [SPSS Inc., 1992].

The possibility of multivariate outliers was assessed using Mahalanobis' distance [see SPSS Inc., 1992]; none were found. Multivariate normality was

TABLE 1. Percentage of Subjects Making Correct and Incorrect Judgements About the Presence of the Magnetic Field Across Trial Blocks*

Decision outcome	% of subjects
Correct block 1	31.3
Incorrect block 1	35.3
Correct block 2	18.2
Incorrect block 2	15.2

*The expected value for the percentage of subjects falling in each of the four groups is 25. Thus, the response 'block 1,' irrespective of whether it was correct or incorrect, has an expected value of 50%, the same as for block 2. In fact, the 'block 1' response (correct plus incorrect) was much higher (66.6%), and the block 2 response considerably lower (33.4%).

TABLE 2. Mean Percentage of Correct Decisions for Males and Females During Magnetic Field and Sham Exposure as a Function of Task Difficulty*

	Task difficulty		
Exposure condition	Hard	Intermediate	Easy
Males[a]			
Magnetic field	60.0 (8.0)	78.1 (13.1)	85.2 (11.9)
Sham	59.6 (8.1)	78.2 (12.5)	85.0 (12.1)
Females[b]			
Magnetic field	60.9 (7.1)	77.5 (11.1)	84.6 (11.6)
Sham	58.6 (7.7)	79.0 (11.4)	84.3 (10.8)

*S.D.s are shown in parentheses.
[a]n = 40.
[b]n = 59.

assumed because of the large sample size ($n = 99$). Assumptions of multivariate and univariate homogeneity were regarded as valid after nonsignificant outcomes for Box's M test and the Bartlett-Box F test [Tabachnick and Fidell, 1989].

Following the methods of Whittington et al. [1996], an experiment-wise α of .3 was used for all field-related tests; the conventional $\alpha = .05$ was used for all other tests. Effect size (ES) values are reported along with the value of the test statistic and its *P* value. ES, the degree to which the phenomenon under investigation is present in the population, is equivalent to Cohen's [1988] *f*, calculated from η^2, the latter being taken as equivalent to Pillai's trace for multivariate tests [Hager and Moller, 1986]. For univariate tests of significance $\eta^2 = [F \times (\text{df effect})]/[F \times (\text{df effect}) + \text{df error}]$. Cohen's ES was then determined as $f = [\eta^2/(1 - \eta^2)]^{1/2}$.

Post hoc calculations of statistical power (SP) were made using the above statistical package and GPOWER, a program for calculating power [Erdfelder et al., 1996].

RESULTS

FSQ Analysis

It will be recalled that at the end of the experimental session, both the subjects and the experimenter were asked to guess which trial block (1 or 2) had been the real MF condition. Table 1 shows that 49.5% of subjects guessed correctly. The experimenter guessed correctly exactly 50% of the time. Quite clearly, neither the subjects nor the experimenter were able to judge when the MF was applied at better than chance levels. However, irrespective of the correctness of these judgements, subjects were more likely to choose the *first* block of trials as the real MF condition. Trial block 1 was chosen by 66.6% of the subjects whereas only 33.4% choose block 2 (see Table 1). This bias in favour of the first trial block was statistically significant $\chi^2 = 11.51$, $df = 3$, $P = .009$).

MQ Analysis

Only 5% of subjects scored as morning types, 13% as evening types, and the remaining 82% as intermediate types. The main purpose of the MQ data was to find out whether subject perceptions of when they worked best (a.m. or p.m.) influenced the main analysis for time-of-day effects. However, including the MQ scores in the analysis as a covariate produced no evidence for an adjusted between-groups effect for time of day. The small number of morning and evening types precluded any further analysis of the MQ data.

PC Analysis

There are few data available on gender differences in the presence of ELF MFs. Therefore, as a first step we examined the present data for any differences between male and female subjects. Table 2 presents the mean PC values as a function of gender and task difficulty. None of the six comparisons possible between males and females yielded a difference of more than 1%. However, there was a complex three-way interaction involving MF, gender, and time-of-day ($F_{1,91} = 2.40$, $P = .13$, ES = 0.16, SP = 0.70). Further analysis revealed that much of the variance associated with this interaction was created by the MF having a greater effect on females than on males. Subsequent *t* tests showed this difference to be significant only at the most difficult level of the task ($t_{58} = 1.85$, $P = .07$, ES = 0.24, SP = 0.77) (see Table 2). Further research is planned to find out if this gender effect will stand up to replication using both the present task and other performance measures.

TABLE 3. Mean Percentage of Correct Decisions for All Subjects During MF and Sham Exposure for Each Level of Task Difficulty at Two Different Times of Day*

	Task difficulty		
Exposure condition	Hard	Intermediate	Easy
Morning condition			
Magnetic field	60.6 (7.8)	76.9 (11.2)	84.2 (11.1)
Sham	58.7 (7.7)	77.8 (11.2)	83.4 (11.4)
Evening condition			
Magnetic field	60.6 (7.2)	78.5 (12.6)	85.5 (12.3)
Sham	59.3 (8.0)	79.5 (12.4)	85.7 (11.2)

*n = 99. S.D.s are shown in parentheses.

The data were collapsed over gender. Table 3 shows the results as a function of task difficulty and time of day. Averaging over exposure condition and task difficulty, subjects in the evening trials were slightly more accurate (74.8%) than those in the morning (73.3%), but this main effect was not significant ($F < 1$).

Finally, the data were collapsed over gender and time-of-day and are presented as a function of exposure and task difficulty (Table 4). As expected, there was a strong main effect for task difficulty ($F_{2,90} = 435.7$, $P < .001$, ES = 3.11, SP ≈ 1.00).

There was no main effect due to field exposure ($F < 1$), but there was a significant interaction between exposure and task difficulty ($F_{2,97} = 2.50$, $P = .09$, ES = 0.16, SP = 0.83). Tests of the simple interaction effects showed that subjects were more accurate at the most difficult level of the task in the presence of the MF ($t_{98} = 1.58$, $P = .12$, ES = 0.16, SP = 0.71). There were no such differences at the intermediate and easy levels of the task.

DISCUSSION

The main finding from the present study was that acute exposure to a 100 μT, 50 Hz intermittent MF produced a small (1.6%) increase in PC at the most difficult level of a visual duration discrimination task. (As Table 2 shows, most of the increase was due to the female subjects.) With the significance level adjusted upward to .3 to reduce the high probability of committing a type II error, this increase was statistically significant ($P = .12$; ES = 0.16), a result consistent with that obtained by Whittington et al. [1996]. They found that RT under MF exposure was speeded by 2.0% (14 ms) in the most difficult condition of exactly the same visual duration discrimination task ($P = .04$; ES = 0.20).

It is worth noting that the Whittington et al. [1996] result represents a ''reliable'' effect using the conventional $P < .05$ cut-off, whereas the present finding would be considered as support for the null hypothesis (i.e., no MF effect) using the same cut-off. Yet the percentage change in the effect size differs by only 0.40% across the two studies. It seems to us that the similarity between the results of the two studies should take precedence over a rather arbitrary ''significance'' level. In particular, the ES must be evaluated alongside the statistical significance of the test. It is recommended that future studies report ES values, or at least the information necessary to calculate them. Only then can an informed judgement be made about the typical ESs in bioelectromagnetic research. Once these ESs are known, then experiments can be planned to ensure sufficient statistical power to detect an effect, should one exist. Statistical power can be increased by increasing subject numbers, increasing the significance level, improving experimental control, or by a combination of these factors.

TABLE 4. Mean Percentage of Correct Decisions for All Subjects During Magnetic Field and Sham Exposure as a Function of Task Difficulty*

	Task difficulty		
Exposure condition	Hard	Intermediate	Easy
Magnetic field	60.6 (7.4)	77.7 (11.9)	84.9 (11.7)
Sham	59.0 (7.8)	78.7 (11.8)	84.6 (11.3)

*n = 99. S.D.s are shown in parentheses.

Taken together, the results of the present study and those of Whittington et al. [1996] suggest that ESs associated with the effects of weak ELF MFs on at least some aspects of human performance are likely to be small. The question arises as to how small is *too* small. Should we be bothering with the results of studies that produce ESs in the range, say, 0.05–0.20? There are at least two answers to this question. First, at the present time there simply are not enough data to decide. If most ESs in MF research with biological systems turn out to be of this order of magnitude, then most likely they should be reported. At present, studies producing small ESs must not be downgraded as less important than those yielding larger effects.

The second answer is that the research literature on the effects of MFs on biological systems, especially as it relates to the behavioural level of analysis, has little or no established theoretical underpinnings. Because we do not understand the mechanism by which weak, ELF MFs affect behaviour, it is impossible to predict the size of any obtained effects. Thus, at this point in time, there is insufficient empirical and theoretical evidence to help us decide when an ES is too small to be concerned about.

Nevertheless, if it turns out that ESs are typically small (say, < 0.20), then to detect and reproduce them, individual studies will need to have sufficient statistical power. However, as Schmidt [1996] points out, to detect a small effect may require sample sizes of ''often 1000 or more'' with $P = .05$ and statistical power at the recommended 0.8 level [Cohen, 1988]. The present study illustrates the problem. The critical test of the simple interaction effect at the most difficult level of the task had a power of only 0.71, despite relaxing the significance level to 0.3 and running 99 subjects. Running large numbers of subjects is time consuming and can be beyond the resources of some laboratories.

In the view of Schmidt [1996], results from individual studies having insufficient power on their own can be combined using the technique of meta-analysis [Hunter and Schmidt, 1990]. By using this technique, overall ESs can be estimated and confidence intervals can be established for them. In this way, a reasonable estimate of the population ESs can be determined. And, in the final analysis, that is the estimate we are after, not whether individual studies produce a ''significant'' result. As Schmidt [1996] ably demonstrates, counting up the number of significant findings (at $P < .05$) without also examining ESs may produce a misleading, even incorrect, conclusion.

Although time of day did not seem to affect performance either in or out of the MF, one interesting result did emerge from the data obtained from the FSQ. Overall, subjects performed at chance levels in judging whether the MF was applied in the first or second sessions. However, there was a strong preference for judging that the MF was applied in the first session. One possible explanation for this effect is that the subjects came to the study with a degree of anticipatory anxiety. This anxiety would have subsided by the time the second block of 150 trials got underway. Subjects may then have attributed their feelings of anxiety to the anticipated effects of the MF. One way to test this hypothesis would be to run the task over, say, four sessions, assessing anxiety level at the beginning and end of each session. Anxiety levels should be considerably lower in session 4 compared with session 1, as familiarity with the task increased. One would then predict that subject choice as to whether the MF field occurred in the first or second trial block of a session would equalise. Whatever the explanation, this result makes it clear that wherever possible, one should not rely on subjects' perceptions or interpretations of events. These can easily be coloured by the expectations generated by the conditions of the study.

In summary, then, the present findings suggest that weak ELF MFs may have a small effect on visual duration discrimination. More data are required before the results of the present study and similar studies can be combined in a meta-analysis. One of the concerns we share with Schmidt [1996] is that the failure of a result to reach the .05 significance level (usually with no regard for the associated ES) effectively assigns the result to the scrap heap. In fact, the data from such studies will almost certainly contain valuable information that could contribute to a more general analysis of the true population ES.

CONCLUSION

The present results are consistent with the view that weak ELF MFs have small effects on human performance. Our results, and those of Whittington et al. [1996], strongly suggest that these small ESs can be detected, or reproduced, only by running studies with much greater statistical power than has been typical in studies of MF effects on human performance to date. More research is required to replicate these findings and to extend them to other human performance tasks involving, for example, memory and attentional factors. We have further investigations under way in an effort to find human performance tasks that are more sensitive to MF effects that will better stand up to replication attempts in our own and others' laboratories.

REFERENCES

Arendt J, Minors DS, Waterhouse JM (eds) (1989): ''Biological Rhythms in Clinical Practice.'' London: Wright.

Cohen J (1988): ''Statistical Power Analyses for the Behavioral Sciences, 2nd Ed.'' New York: Academic Press.

Colquhoun WP (1971): Circadian variations in mental efficiency. In Colquhoun WP (ed): ''Biological Rhythms and Human Performance.'' London: Academic Press.

Colquhoun WP, Blake MJF, Edwards RS (1968): Experimental studies of shift work: I. A comparison of ''rotating'' and ''stabilized'' 4-h shift systems. Ergonomics 11:437–453.

Cook MR, Graham C, Cohen HD, Gerkovich MM (1992): A replication study of human exposure to 60-Hz fields: Effects on neurobehavioral measures. Bioelectromagnetics 13:261–285.

Craig A (1979): Discrimination, temperature, and time of day. Hum Factors 21:61–68.

Erdfelder E, Faul F, Buchner A (1996): GPOWER: A general power analysis program. Behv Res Meth Instrum Comp 28:1–11.

Graham C, Cohen HD (1985): ''Influence of 60-Hz Fields on Human Behavior Physiology Biochemistry. Contractors' final report to New York State Power Lines Project (Contract No. 218203).'' New York: Wadsworth Center for Laboratories and Research.

Hager W, Moller H (1986): Tables and procedures for the determination of power and sample size in univariate and multivariate analyses of variance and regression. Biometr J 28:647–663.

Horne J, Ostberg O (1976): A self-assessment questionnaire to determine morningness and eveningness in human circadian rhythms. Int J Chronobiol 4:97–110.

Hunter JE, Schmidt FL (1990): ''Methods of Meta-analysis: Correcting Error and Bias in Research Findings.'' Newbury Park, CA: Sage.

Johnson LC, Tepas DI, Colquhoun WP, Colligan MJ (eds) (1981): ''Biological Rhythms, Sleep and Shift Work.'' New York: Spectrum.

Moore-Ede MC, Sulzman FM, Fuller CA (1982): ''The Clocks That Time Us.'' Cambridge, MA: Harvard University Press.

Podd JV, Whittington CJ, Barnes G, Page W, Rapley B (1995): Do ELF magnetic fields affect human reaction time? Bioelectromagnetics 16:317–323.

Schmidt FL (1996): Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. Psychol Meth 1:115–129.

Smith CS, Reilly C, Midkiff K (1989): Evaluation of three circadian rhythm questionnaires with suggestions for an improved measure of morningness. J Appl Psychol 74:728–738.

SPSS Inc. (1992): ''SPSS/PC+ Advanced Statistics, ver. 5.0.'' Chicago: SPSS, Inc.

Tabachnick BG, Fidell LS (1989): ''Using Multivariate Statistics, 2nd Ed.'' New York: Harper Row.

Whittington CJ, Podd JV (1996): Human performance and physiology: A statistical power analysis of low-intensity ELF electromagnetic field research. Bioelectromagnetics 17:274–278.

Whittington CJ, Podd JV, Rapley BR (1996): Acute effects of 50 Hz magnetic field exposure on human visual task and cardiovascular performance. Bioelectromagnetics 17:131–137.